This International Student Edition is for use outside of the U.S.

EVERGREEN RELEASE

AI IN BUSINESS
Creating Value Responsibly

First Release

Self-Print! Directly from your eBook



Patrick Johanns
James Chaffee
Jackie Rees Ulmer

Mc Graw Hill

This International Student Edition is for use outside of the U.S.

EVERGREEN RELEASE

AI IN BUSINESS
Creating Value Responsibly

First Release

Self-Print! Directly from your eBook



Patrick Johanns
James Chaffee
Jackie Rees Ulmer

Mc Graw Hill

Table of Contents

1. Frontmatter
    1. Cover Page
    2. Title Page
    3. Copyright Page
    4. Dedications
    5. About the Authors
        1. Patrick (Pat) Johanns
        2. James Chaffee
        3. Jackie Rees Ulmer
    6. A Letter to New Adopters
    7. Guided Tour
        1. Learning Objectives
        2. Focus on Ethics and Sustainability
        3. Visual Diagrams
        4. Summary of Key Points
        5. Quick Wins
        6. Case Studies
    8. Connect
    9. Acknowledgments
    10. Brief Contents
    11. Detailed Table of Contents
2. Chapter 1: Overview of AI in Business
    1. Chapter 1: Overview of AI in Business
    2. Introduction
    3. Types of Artificial Intelligence
        1. Machine Learning
        2. Neural Networks
        3. Generative AI
    4. Other Applications of AI in Business
        1. The Use of Artificial Intelligence in Virtual Reality
        2. Computer Vision and Its Role in Business Applications
        3. Speech Recognition and Its Business Impact
    5. Data Science and AI
    6. Ethical Considerations in AI
        1. Data Ownership, Privacy, and Consent
        2. Environmental Sustainability
        3. Economic Inequality and AI Access
        4. Bias and Discrimination
        5. Manipulation and Deception
        6. Intellectual Property and AI Creativity
        7. Transparency and Accountability
    7. Summary of Key Points and Quick Wins
        1. QUICK WINS
    8. Key Terms
    9. Discussion Questions
    10. Exercises
    11. Case Study 1.1: The Portrait of Edmond De Belamy—Intellectual Property and AI Creativity
        1. U.S. and French Copyright Laws
        2. Discussion Questions
        3. Case References
    12. Case Study 1.2: Coded Bias
        1. Discussion Questions
        2. Case References
    13. Case Study 1.3: Closing the Loop—How H&M and HKRITA Use AI to Revolutionize Textile Recycling
        1. What Happens after Sorting?
        2. Insights
        3. AI’s Growing Presence in Physical, Traditionally Manual Industries
        4. The Potential of AI for Sustainability and Circular Economies

5. The Importance of Data-Rich Environments
6. AI in Long-Term Environmental Goals
7. Discussion Questions
8. Case References

14. Endnotes

3. Chapter 2: The Rise of AI: Innovation and Impact
    1. Chapter 2: The Rise of AI: Innovation and Impact
    2. Introduction
    3. Foundations of AI (Pre–1950s)
        1. Artistic Foundations
        2. Scientific Foundations: Early Thinkers and Theories
    4. The Dawn of AI (1950–1970)
        1. The Turing Test: Can Machines Think?
        2. The Dartmouth Conference
        3. Early AI Programs
        4. Emerging Computer Technologies
        5. Impact on Business Practices
    5. The Age of Experimentation (1970–1985)
        1. Expert Systems
        2. Advancements in Neural Networks
        3. The Microprocessor Revolution and Technological Foundations
        4. Impact of AI on Business from 1970–1985
    6. The AI Winter and Resurgence (1985–2000)
        1. The Resurgence: Seeds of Renewal
        2. Technological Foundations
        3. AI and Robotics in Business (1985–2000)
    7. The Age of Big Data and Machine Learning (2000–2020)
        1. Technological Foundations
        2. Business Applications
    8. The Generative AI Revolution (2020–Present)
        1. Scientific and Technological Advancements
        2. AI and the Internet of Things: A Seamless, Intelligent World
        3. Ethical, Safety, and Privacy Concerns in AI
    9. AI's Future: Opportunities and Challenges
    10. Summary of Key Points and Quick Wins
        1. QUICK WINS
    11. Key Terms
    12. Case Study 2.1: AI and COVID-19
        1. AI's Role in the COVID-19 Response
        2. The Pandemic's Impact on AI Adoption
        3. Discussion Questions
        4. Case References
    13. Case Study 2.2: Project Innereye: Expanding AI in Medicine
        1. Discussion Questions
        2. Case References
    14. Case Study 2.3: Pioneering Women's Sports Analytics with AI
        1. AI's Journey in Sports Analytics
        2. A Historical First for Women's Sports
        3. Impact on the Game
        4. Discussion Questions
        5. Case References
    15. Endnotes
4. Chapter 3: Data Management for AI
    1. Chapter 3: Data Management for AI
    2. Introduction
    3. Data for AI
        1. Data Structure and Types
        2. Sources of Data for AI
    4. Data Security
        1. Data Security Challenges
        2. AI Data Security Principles
    5. Data Quality and Integrity
        1. High Quality Data: Characteristics and Methods

    2. Data Cleaning
    3. Building a Data-Ready Organization
6. Data Regulations
    1. Broad Governmental Policies
    2. Industry-Specific Data and AI Regulations
7. Big Data and Data Management
    1. Larger Data Centers
    2. Distributed Cloud-Based Storage Systems
    3. Sustainability in Storage
8. Emerging Technologies in AI Data Management
    1. Blockchain
    2. AI-Powered Cybersecurity
    3. Edge Computing and Storage
    4. Federated Learning
9. Ethical Considerations in AI Data Collection and Use
10. Summary of Key Points and Quick Wins
    1. QUICK WINS
11. Key Terms
12. Discussion Questions
13. Exercises
14. Case Study 3.1: Amazon—AI in the Hiring Process
    1. The Start of the Journey
    2. Clues of Bias
    3. Repercussions and Repairs
    4. Discussion Questions
    5. Case References
15. Case Study 3.2: Darktrace—AI in Network Security
    1. Benefits and Challenges
    2. Broader Implications
    3. Discussion Questions
    4. Case References
16. Case Study 3.3: Tradelens—Blockchain as a Solution, But Why Did It Fail?
    1. Blockchain as a Solution
    2. An Unexpected Failure
    3. Lessons from Failure
    4. Conclusion
    5. Discussion Questions
    6. Case References
17. Endnotes

5. Chapter 4: Machine Learning
    1. Chapter 4: Machine Learning
    2. Introduction
    3. How Machine Learning Models Learn
        1. Supervised Learning
        2. Unsupervised Learning
        3. Reinforcement Learning
    4. Machine Learning in Business Functions
        1. Data Management and Preprocessing
        2. Prediction and Forecasting
        3. Classification and Risk Assessment
        4. Customer Segmentation and Personalization
        5. Process Optimization and Automation
        6. Comparing Standard Machine Learning Algorithms
    5. Ethical Risks of Machine Learning in Business
        1. Algorithmic Bias in Machine Learning
        2. Strategies for Reducing Bias in Machine Learning Systems
        3. Other Concerns
    6. Summary of Key Points and Quick Wins
        1. QUICK WINS
    7. Key Terms
    8. Discussion Questions
    9. Exercises
    10. Case Study 4.1: Netflix Recommendation System

        1. Netflix’s History
        2. The Heart of Netflix: Its Recommender Grid
        3. The Data Driving the System
        4. Challenges and Ethical Considerations
        5. Business Impact
        6. Discussion Questions
        7. Case References
    11. Case Study 4.2: Flight Delays at JFK
        1. Implementation
        2. Discussion Questions
        3. Case References
    12. Case Study 4.3: UPS Vehicle Routing
        1. Challenges for Logistics Companies
        2. UPS and ORION: AI for Efficiency
        3. The Left-Turn Problem
        4. Real-Time Optimization and Cost Savings
        5. Conclusion
        6. Discussion Questions
        7. Case References
    13. Endnotes
6. Chapter 4S: Machine Learning Technical Supplement
    1. Chapter 4S: Machine Learning Technical Supplement
    2. Principal Component Analysis (PCA)
        1. How PCA Works
        2. Business Applications of PCA
        3. Advantages of PCA
        4. Limitations of PCA
    3. K-Means Clustering
        1. How k-Means Clustering Works
        2. Business Applications of k-Means Clustering
        3. Advantages of k-Means Clustering
        4. Limitations of k-Means Clustering
    4. K-Nearest Neighbors (K-NN)
        1. How k-NN Works
        2. Business Applications of k-NN
        3. Advantages of k-NN
        4. Limitations of k-NN
    5. Linear Regression
        1. How Linear Regression Works
        2. Business Applications of Linear Regression
        3. Advantages of Linear Regression
        4. Limitation of Linear Regression
    6. Decision Trees
        1. How Decision Trees Work
        2. Business Applications of Decision Trees
        3. Advantages of Decision Trees
        4. Limitations of Decision Trees
    7. Random Forests
        1. How Random Forests Work
        2. Business Applications of Random Forests
        3. Advantages of Random Forests
        4. Limitations of Random Forests
    8. Logistic Regression
        1. How Logistic Regression Works
        2. Business Applications of Logistic Regression
        3. Advantages of Logistic Regression
        4. Limitations of Logistic Regression
    9. Support Vector Machines (SVMs)
        1. How SVM Works
        2. Business Applications of SVMs
        3. Advantages of SVMs
        4. Limitations of SVMs
    10. Collaborative Filtering

        1. How Collaborative Filtering Works
        2. Business Applications of Collaborative Filtering
        3. Advantages of Collaborative Filtering
        4. Limitations of Collaborative Filtering
    11. Time Series Analysis
        1. How Time Series Analysis Works
        2. Business Applications of Time Series Analysis
        3. Advantages of Time Series Analysis
        4. Limitations of Time Series Analysis
    12. Metrics for the Quality of Algorithms
        1. Confusion Matrix
        2. Metrics
        3. Summary
        4. Algorithm Selection Questions
        5. Metrics and Evaluation Questions
7. Chapter 5: Neural Networks and Deep Learning
    1. Chapter 5: Neural Networks and Deep Learning
    2. Introduction
    3. Clarifying Neural Networks and Deep Learning
    4. The Growing Importance of Neural Networks
        1. Business Applications of Neural Networks
        2. Why Businesses Are Adopting Neural Networks
        3. Real-World Impacts
    5. Deep Learning
        1. Deep Learning's Role in Business
        2. Deep Learning in Natural Language Processing
    6. NLP Applications in Business
        1. Limitations of Neural Networks and Deep Learning
    7. Emerging Capabilities in Deep Learning
        1. Foundation Models
        2. Multimodal Learning
        3. Few-Shot and Zero-Shot Learning
        4. Energy-Efficient Deep Learning (Green AI)
        5. Deep Learning for Personalized Adaptive Systems
        6. Deep Learning and Virtual Reality
        7. Growth of Neural Network Models
    8. Summary of Key Points and Quick Wins
        1. QUICK WINS
    9. Key Terms
    10. Discussion Questions
    11. Exercises
    12. Case Study 5.1: IBM Watson Openscale—Explainable AI
        1. IBM Watson and OpenScale
        2. Challenges and Tradeoffs
        3. Conclusion
        4. Discussion Questions
        5. Case References
    13. Case Study 5.2: Tesla Autopilot—The Promise and Peril of Deep Learning–Driven Autonomy
        1. The Market for Autonomous Vehicles
        2. Deep Learning and Tesla's Autonomy Strategy
        3. Challenges in Technology and Regulation
        4. Legal Difficulties
        5. Conclusion
        6. Discussion Questions
        7. Case References
    14. Case Study 5.3: Facial Recognition Technology and the Detroit Lions
        1. Technology Overview
        2. Applications and Business Value
        3. Discussion Questions
        4. Case Reference
    15. Endnotes
8. Chapter 5S: Neural Networks and Deep Learning Concepts Technical Supplement
    1. Chapter 5S: Neural Networks and Deep Learning Concepts Technical Supplement

2. Traditional Machine Learning vs. Deep Learning
3. How Neural Networks Work
    1. Neural Network Architecture
    2. Activation Functions
    3. Forward Propagation and Backpropagation
    4. Training with Stochastic Gradient Descent
    5. Office Building Analogy
    6. Overfitting
4. Deep Learning Model Types
    1. Convolutional Neural Networks: Specialists in Visual Data
    2. Recurrent Neural Networks: Specialists in Sequences and Time Series
5. Transformers and Attention Mechanisms
    1. What Are Transformers?
    2. How Attention Mechanisms Work
    3. Transformers in Modern AI Applications
    4. Conclusion
6. Endnotes

9. Chapter 6: Generative AI in Business: Transforming Industries with Innovative Technology
    1. Chapter 6: Generative AI in Business: Transforming Industries with Innovative Technology
    2. Introduction
    3. What Is Generative AI?
        1. Predictive vs. Generative AI
        2. Types of Generative AI
    4. How Generative AI Works
        1. Training Generative Models
        2. Text Generation: Step by Step
        3. Image Generation
        4. Multimodal and Video Generation
    5. Business Applications of Generative AI
        1. Text-Based Applications
        2. Custom GPTs
        3. Chatbots and Customer Engagement
        4. Generative Visual Content and Virtual Try-On
        5. Product Design
        6. Generative AI in Contract Review
        7. Training and Onboarding
        8. Multimodal and Interactive Experiences
        9. Specialized Use Cases across Industries
    6. Risks and Limitations of Generative AI
        1. Hallucinations and Factual Inaccuracy
        2. Bias and Fairness
        3. Intellectual Property and Legal Concerns
        4. Transparency and Explainability
        5. Ethical Use and Governance
        6. Work Role Changes
    7. Business Models in Generative AI
    8. Summary of Key Points and Quick Wins
        1. QUICK WINS
    9. Key Terms
    10. Discussion Questions
    11. Exercises
    12. Case Study 6.1: When AI Makes False Claims—The OpenAI Libel Controversy
        1. Discussion Questions
        2. Case References
    13. Case Study 6.2: AI Voice Cloning
        1. How It Works
        2. Uses—and Abuses
        3. When Voice Cloning Helps
        4. Grief and the Gray Areas
        5. Legal and Ethical Challenges
        6. Discussion Questions
        7. Case References
    14. Case Study 6.3: Revolutionizing Soccer Tactics with TacticAI

        1. The Role of TacticAI
        2. How It Works
        3. Business and Competitive Value
        4. Potential Risks and Limitations
        5. The Future of Sports AI
        6. Discussion Questions
        7. Case References
    15. Endnotes
10. Chapter 6S: Generative AI Technical Supplement
    1. Chapter 6S: Generative AI Technical Supplement
    2. Key Technical Components
        1. Tokenization and Embeddings
        2. Transformer Architectures and Autoregressive Generation
    3. Core Generative Model Types
        1. Transformers
        2. Diffusion Models
        3. Generative Adversarial Networks
        4. Variational Autoencoders
        5. Recurrent Neural Networks
    4. How Generative AI Balances Creativity and Coherence
        1. Sampling Strategies in Generative AI
        2. Generative AI Isn’t Done Evolving
    5. Endnotes
11. Chapter 7: Strategic AI Implementation: People, Process, and Purpose
    1. Chapter 7: Strategic AI Implementation: People, Process, and Purpose
    2. Introduction
    3. Assessing Organizational Readiness for AI
        1. Data Readiness
        2. Integrating Data across the Organization
        3. Technological Readiness
        4. Business Process Readiness
        5. Workforce Readiness
        6. Cultural and Leadership Readiness
    4. Identifying Business Functions Ready for AI Integration
        1. Customer Service
        2. Marketing
        3. Operations and Supply Chain Management
        4. Other Departmental Applications
        5. Aligning AI Projects with Business Objectives
    5. Implementing AI Solutions
        1. Selecting the Right AI Tools
        2. Selecting the Right AI Vendors
    6. Monitoring and Measuring AI Effectiveness
        1. Innovation and Creativity Metrics
        2. Operational Efficiency Metrics
        3. Customer Experience Metrics
        4. Business Outcome Metrics
        5. Balanced Scorecard Approach
    7. Basic Ethical, Legal, and Regulatory Considerations in AI Implementation
        1. Curate Diverse and Ethical Training Data Biases
        2. Ensure Explainability
        3. Promote Transparency
        4. Continually Assess Risks and Benefits
        5. Ensure Accountability
        6. Plan for Workforce Impact
        7. Legal and Regulatory Considerations
        8. A Final Note
    8. Special Considerations for Generative AI
        1. Ethical Foundations for Generative AI
        2. Reducing Hallucinations and Misinformation
        3. Managing Workforce Disruptions
        4. Safeguarding User Consent and Control
        5. Navigating Copyright and Intellectual Property Risks

        6. Legal Accountability in Generative AI Systems
    9. Summary of Key Points and Quick Wins
        1. QUICK WINS
    10. Key Terms
    11. Discussion Questions
    12. Exercises
    13. Case Study 7.1: Mastercard Fraud Detection
        1. AI as a Solution
        2. Implementation Decision
        3. The Launch of Decision Intelligence
        4. Outcomes
        5. AI Expansion beyond Fraud
        6. Takeaways
        7. Discussion Questions
        8. Case References
    14. Case Study 7.2: Pfizer-BioNTech COVID-19 Vaccine
        1. A Compressed Timeline—Enabled by AI
        2. AI Contributions across the Development Pipeline
        3. Why It Worked
        4. Ethical and Operational Considerations
        5. Conclusion
        6. Discussion Questions
        7. Case References
    15. Case Study 7.3: The New York Times and Google Cloud AI
        1. Digitization: A Human-Driven Process
        2. AI Vendor Choice
        3. From Preservation to Personalization
        4. Discussion Questions
        5. Case References
    16. Endnotes
12. Chapter 8: The Art of Prompt Engineering with FOCUS
    1. Chapter 8: The Art of Prompt Engineering with FOCUS
    2. Introduction
    3. The Focus Framework for Effective Prompts
        1. General Principles of Prompt Crafting
        2. Hard Prompts: Benchmarks for AI
    4. Iterative Refinement
        1. Prompt Refinement Example
    5. Understanding AI Models
        1. Closed-Domain Chatbots
        2. Open-Domain Chatbots
        3. How Chatbots Explain Themselves
    6. Advanced Prompting Techniques
        1. Task Decomposition
        2. Additional Techniques and Strategies
        3. Communicating Effectively with Business AI Tools
    7. Personas in Prompt Engineering
        1. Choosing the Right Persona
        2. Using Personas Effectively
        3. Potential Risks of Persona-Driven Prompts
    8. Strategies for Verifying AI Responses
    9. Ethical Considerations In Prompt Engineering
        1. Ethical Prompt Crafting
        2. Verification
        3. Ethical Use of Generative AI Output
        4. Intellectual Property and Creativity
        5. Intellectual Property and AI
        6. Ethical Prompting Checklist
        7. The End Is Just the Beginning
    10. Summary of Key Points and Quick Wins
        1. QUICK WINS
    11. Key Terms
    12. Discussion Questions

13. Exercises
14. Case Study 8.1: 2023 Writers Guild of America Strike
    1. Writers' Concerns about AI
    2. The Studios' View
    3. Legal and Ethical Implications
    4. Resolution and Aftermath
    5. Discussion Questions
    6. Case References
15. Case Study 8.2: Generative-AI Art in Board Games
    1. Jenna's Decision: Legends Unbound
    2. Business Opportunities
    3. Ethical Challenges
    4. Conclusion
    5. Discussion Questions
    6. Case References
16. Case Study 8.3: Hallucinating the Law
    1. What Went Wrong?
    2. Could a Persona Have Helped?
    3. Open vs. Closed Chatbots: Choosing the Right Tool
    4. Prompt Engineering Lessons
    5. The Legal and Ethical Fallout
    6. The Bigger Picture
    7. Discussion Questions
    8. Case References
17. Endnotes

13. AI in Action: The Singapore Case Series
    1. AI in Action: The Singapore Case Series
        1. INTRODUCTION
    2. A Seamless Morning in Singapore's Smart Nation
        1. Smart Home Start
        2. Stepping into the Community
        3. Public Infrastructure in Sync
        4. Commute with Intelligence
        5. Smart Commerce and Customer Experience
        6. Workplace and Reflection
        7. Discussion Questions
        8. Case References
    3. Foundations of a Smart Nation
        1. A Glimpse of the Future
        2. The Three Pillars of the Smart Nation Vision
        3. Challenges and Opportunities in 2014
        4. Singapore's Strategic Shift toward AI
        5. Discussion Questions
        6. Case References
    4. Singapore's Data Engine
        1. The Foundation of Data Infrastructure
        2. Governance and Trust
        3. From Data to Impact
        4. Looking Ahead
        5. Discussion Questions
        6. Case References
    5. Singapore's Intelligent Spaces
        1. Ari at the Bosch Campus
        2. Singapore's Green Plan and the Role of Machine Learning
        3. The Sensor-Driven Transformation
        4. Training the Model
        5. Scaling toward a Smart Nation
        6. Discussion Questions
    6. Singapore's Intelligent Transportation System
        1. A Road That Thinks Ahead
        2. Neural Networks in the Infrastructure
        3. Making a Case for the Human Impact
        4. A System That Sees Too Much?

5. A System That Finds Parking Too
6. A System That Clears the Way
7. The Future in the Foundations
8. Discussion Questions
9. Case References

7. AI in Urban Design
    1. A Smarter Start to Design
    2. Prompting Smarter Urban Design
    3. Bias in the Training Data
    4. Ethical and Technical Questions
    5. The Bigger Picture
    6. Discussion Questions
    7. Case References
8. Personalizing Government: The LifeSG Project
    1. A New Kind of Public Service
    2. From Moments of Life to LifeSG: A Platform Evolves
    3. Pilot Integration and AI Functionality
    4. Reshaping Public Service Roles
    5. Addressing Explainability and Trust
    6. Citizen Adoption and Community Engagement
    7. Smart Cities around the World
    8. Discussion Questions
    9. Case References
9. Telling the Smart Nation Story
    1. Ari’s New Assignment
    2. Calling in AI for Assistance
    3. Applying the FOCUS Framework
    4. The Hallucination Test
    5. Breaking Down the Project
    6. Ari at the Podium
    7. Discussion Questions
10. Endnotes

14. Accessibility Content: Text Alternatives for Images
    1. LO Extended Description (Frontmatter)
    2. Basic ethical Extended Description (Frontmatter)
    3. Sustainability Extended Description (Frontmatter)
    4. Figure 6.3 Extended Description (Frontmatter)
    5. Figure 8.1 Extended Description (Frontmatter)
    6. Summary Extended Description (Frontmatter)
    7. Figure 1.1 Extended Description (Chapter 1)
    8. Figure 1.2 Extended Description (Chapter 1)
    9. Figure 1.3 Extended Description (Chapter 1)
    10. Unfigure 1.1 Extended Description (Chapter 1)
    11. Unfigure 2.1 Extended Description (Chapter 2)
    12. Figure 2.1 Extended Description (Chapter 2)
    13. Unfigure 2.2 Extended Description (Chapter 2)
    14. Figure 3.1 Extended Description (Chapter 3)
    15. Figure 3.2 Extended Description (Chapter 3)
    16. Figure 3.3 Extended Description (Chapter 3)
    17. Figure 3.4 Extended Description (Chapter 3)
    18. Figure 3.5 Extended Description (Chapter 3)
    19. Figure 3.6 Extended Description (Chapter 3)
    20. Figure 4.1 Extended Description (Chapter 4)
    21. Figure 4.2 Extended Description (Chapter 4)
    22. Figure 4.3 Extended Description (Chapter 4)
    23. Figure 4.4 Extended Description (Chapter 4)
    24. Figure 4S.1 Extended Description (Chapter 4S)
    25. Figure 4S.2 Extended Description (Chapter 4S)
    26. Figure 4S.3 Extended Description (Chapter 4S)
    27. Figure 4S.4 Extended Description (Chapter 4S)
    28. Figure 4S.5 Extended Description (Chapter 4S)
    29. Figure 4S.6 Extended Description (Chapter 4S)
    30. Figure 4S.7 Extended Description (Chapter 4S)

31. Figure 4S.8 Extended Description (Chapter 4S)
32. Figure 4S.9 Extended Description (Chapter 4S)
33. Figure 4S.10 Extended Description (Chapter 4S)
34. Figure 4S.11 Extended Description (Chapter 4S)
35. Figure 4S.12 Extended Description (Chapter 4S)
36. Figure 5.2 Extended Description (Chapter 5)
37. Figure 5.3 Extended Description (Chapter 5)
38. Figure 5S.1 Extended Description (Chapter 5S)
39. Figure 6.1 Extended Description (Chapter 6)
40. Figure 6.2 Extended Description (Chapter 6)
41. Figure 6.3 Extended Description (Chapter 6)
42. Figure 6.4 Extended Description (Chapter 6)
43. Figure 6S.1 Extended Description (Chapter 6)
44. Figure 6S.2 Extended Description (Chapter 6)
45. Figure 6S.3 Extended Description (Chapter 6)
46. Figure 7.2 Extended Description (Chapter 7)
47. Figure 8.1 Extended Description (Chapter 8)
48. Figure 8.2 Extended Description (Chapter 8)
49. Figure 8.3 Extended Description (Chapter 8)

Guide

1. Title Page
2. Copyright Page
3. Dedications
4. About the Authors
5. A Letter to New Adopters
6. Guided Tour
7. Connect
8. Acknowledgments
9. Brief Contents
10. Detailed Table of Contents
11. Chapter 1: Overview of AI in Business
12. Chapter 2: The Rise of AI: Innovation and Impact
13. Chapter 3: Data Management for AI
14. Chapter 4: Machine Learning
15. Chapter 4S: Machine Learning Technical Supplement
16. Chapter 5: Neural Networks and Deep Learning
17. Chapter 5S: Neural Networks and Deep Learning Concepts Technical Supplement
18. Chapter 6: Generative AI in Business: Transforming Industries with Innovative Technology
19. Chapter 6S: Generative AI Technical Supplement
20. Chapter 7: Strategic AI Implementation: People, Process, and Purpose
21. Chapter 8: The Art of Prompt Engineering with FOCUS
22. AI in Action: The Singapore Case Series
23. Accessibility Content: Text Alternatives for Images

Page List

1. i
2. ii
3. iii
4. iv
5. v
6. vi
7. vii
8. viii
9. ix
10. x
11. xi
12. xii

13. xiii
14. xiv
15. xv
16. xvi
17. xvii
18. xviii
19. xix
20. xx
21. xxi
22. 1
23. 2
24. 3
25. 4
26. 5
27. 6
28. 7
29. 8
30. 9
31. 10
32. 11
33. 12
34. 13
35. 14
36. 15
37. 16
38. 17
39. 18
40. 19
41. 20
42. 21
43. 22
44. 23
45. 24
46. 25
47. 26
48. 27
49. 28
50. 29
51. 30
52. 31
53. 32
54. 33
55. 34
56. 35
57. 36
58. 37
59. 38
60. 39
61. 40
62. 41
63. 42
64. 43
65. 44
66. 45
67. 46
68. 47
69. 48
70. 49
71. 50
72. 51
73. 52
74. 53
75. 54

76. 55
77. 56
78. 57
79. 58
80. 59
81. 60
82. 61
83. 62
84. 63
85. 64
86. 65
87. 66
88. 67
89. 68
90. 69
91. 70
92. 71
93. 72
94. 73
95. 74
96. 75
97. 76
98. 77
99. 78
100. 79
101. 80
102. 81
103. 82
104. 83
105. 84
106. 85
107. 86
108. 87
109. 88
110. 89
111. 90
112. 91
113. 92
114. 93
115. 94
116. 95
117. 96
118. 97
119. 98
120. 99
121. 100
122. 101
123. 102
124. 103
125. 104
126. 105
127. 106
128. 107
129. 108
130. 109
131. 110
132. 111
133. 112
134. 113
135. 114
136. 115
137. 116
138. 117

139. 118
140. 119
141. 120
142. 121
143. 122
144. 123
145. 124
146. 125
147. 126
148. 127
149. 128
150. 129
151. 130
152. 131
153. 132
154. 133
155. 134
156. 135
157. 136
158. 137
159. 138
160. 139
161. 140
162. 141
163. 142
164. 143
165. 144
166. 145
167. 146
168. 147
169. 148
170. 149
171. 150
172. 151
173. 152
174. 153
175. 154
176. 155
177. 156
178. 157
179. 158
180. 159
181. 160
182. 161
183. 162
184. 163
185. 164
186. 165
187. 166
188. 167
189. 168
190. 169
191. 170
192. 171
193. 172
194. 173
195. 174
196. 175
197. 176
198. 177
199. 178
200. 179
201. 180

202. 181
203. 182
204. 183
205. 184
206. 185
207. 186
208. 187
209. 188
210. 189
211. 190
212. 191
213. 192
214. 193
215. 194
216. 195
217. 196
218. 197
219. 198
220. 199
221. 200
222. 201
223. 202
224. 203
225. 204
226. 205
227. 206
228. 207
229. 208
230. 209
231. 210
232. 211
233. 212
234. 213
235. 214
236. 215
237. 216
238. 217
239. 218
240. 219
241. 220
242. 221
243. 222
244. 223
245. 224
246. 225
247. 226
248. 227
249. 228
250. 229
251. 230
252. 231
253. 232
254. 233
255. 234
256. 235
257. 236
258. 237
259. 238

This International Student Edition is for use outside of the U.S.

EVERGREEN RELEASE

AI IN BUSINESS
Creating Value Responsibly

First Release

Self-Print! Directly from your eBook



Patrick Johanns
James Chaffee
Jackie Rees Ulmer

Mc Graw Hill

AI in Business: Creating Value Responsibly

Authors
Patrick Johanns, University of Iowa
James Chaffee, University of Iowa
Jackie Rees Ulmer, Ohio University



AI IN BUSINESS

Published by McGraw Hill LLC, 1325 Avenue of the Americas, New York, NY 10019. Copyright © 2026 by McGraw Hill LLC. All rights reserved. Printed in the United States of America. No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written consent of McGraw Hill LLC, including, but not limited to, in any network or other electronic storage or transmission, or broadcast for distance learning.

Some ancillaries, including electronic and print components, may not be available to customers outside the United States.

ISBN 978-1-265-65863-2
MHID 1-265-65863-3

Cover Image: *Jamesteohart/Shutterstock*

All credits appearing on page or at the end of the book are considered to be an extension of the copyright page.

The Internet addresses listed in the text were accurate at the time of publication. The inclusion of a website does not indicate an endorsement by the authors or McGraw Hill LLC, and McGraw Hill LLC does not guarantee the accuracy of the information presented at these sites.

mheducation.com/highered

Dedications

To my little AI, Ava Ignazia, and the rest of my family, who dealt with the many hours of thinking, writing, and talking about this book.

To my colleagues and the Tippie College of Business Administration, who supported this effort.

Patrick Johanns

To my wife, thank you for your unwavering support, encouragement, and patience throughout this journey.

To my tech team colleagues at the Tippie College of Business, who listened constantly to me talking about the book.

James Chaffee

To Mike, Christian, and Nathan for putting up with the many long and weird hours over the years. And to Gary Koehler, for inspiring the love of machine learning and AI in its many forms.

Jackie Rees Ulmer

About the Authors

Patrick (Pat) Johanns



Justin Tomer

Patrick (Pat) Johanns is an Associate Professor of Instruction in Business Analytics at the Tippie College of Business at the University of Iowa, where he has taught since 2014. He served as Faculty Director of the Professional Business Analytics Graduate Program from 2019 to 2024, leading it to national recognition. The department won the INFORMS UPS George D. Smith Prize for excellence in preparing students for operations research and analytics careers. He was honored as the program's Best Core Course Instructor for 2024–2025.

Pat holds a PhD in Operations Management, an MBA, and a BS in Microbiology from the University of Iowa. His diverse academic background informs his interdisciplinary approach to teaching. His academic career has spanned five continents and includes nearly a decade at Melbourne Business School, one of Australia's highest-ranked business schools. A passionate educator, he has received multiple teaching awards, including at Purdue University's Daniels School of Business, where he taught from 2003 to 2014. This global perspective enriches his teaching, bringing his students a wealth of international insights.

He is coauthor of *AI in Business: Creating Value Responsibly with Artificial Intelligence* (McGraw-Hill) and author of *Management Science with Spreadsheet Modeling* (Kendall Hunt). His work bridges technical rigor and real-world relevance, helping students make sense of emerging technologies like AI, blockchain, and predictive analytics and how those technologies can be effectively implemented in business.

Beyond the classroom, Pat has led entrepreneurship and analytics workshops in Kenya and served on multiple U.S. State Department public diplomacy grants. He is also the author of *Distilled*, a Substack series of 100-word micro-memoirs capturing life lessons, humor, and personal transformation.

A lifelong enthusiast of board games, Pat is an expert-level player, having competed in national championships in both the United States and Australia. He playtests upcoming titles for game designers and enjoys the strategic depth and social interaction that board gaming offers.

Pat lives in North Liberty, Iowa, with his wife, Barbara, whom he first met in a second-grade Christmas play. They have three children and one grandchild. He enjoys traveling, writing, and camping in his Airstream, *Daystream Believer.*

James Chaffee



Impact Photo

James (Jim) Chaffee is an Associate Professor of Practice in Business Analytics at the Tippie College of Business at the University of Iowa, where he has taught and led technology initiatives since 2007. Over his extensive career, he has guided strategic projects transforming how emerging technologies like artificial intelligence, blockchain, and virtual reality are integrated into business education. In his previous roles of leading technology implementation, Jim has been instrumental in designing collaborative learning spaces, launching technology-forward curricula, and spearheading safety and facilities management strategies for Tippie.

Jim holds an MA in Educational Technology from the University of Illinois and a BS in Secondary English Education from Penn West University. He has developed and taught a wide array of courses, including Emerging Technology Driving Business, Innovations in Technology, and Video Games as Learning Tools, bringing practical, hands-on insights into the

classroom. His research and presentations have explored topics ranging from online education through gaming systems to the transformative power of XR and AI in business.

Beyond his work at the university, Jim co-hosted the PBS show *Bits & PCs*, presented at national conferences, and contributed thought leadership pieces to publications like *University Business* and *BizEd Magazine*. He has served on numerous committees dedicated to advancing educational excellence and has advised student organizations fostering leadership and innovation.

A passionate advocate for harnessing technology to improve teaching and learning, Jim brings a unique blend of academic expertise, operational leadership, and a love for empowering others to everything he does. He lives in West Branch, Iowa with his wife Tammy and their little dog Moriyah.

Jackie Rees Ulmer



Jackie Rees Ulmer

Jackie Rees Ulmer has served as the Glenn E. Corlett Dean of the College of Business and a Professor of Analytics and Information Systems at Ohio University since 2021. Previously, she served the Ivy College of Business at Iowa State University in various administrative roles, including Associate Dean and Department Chair. She was a professor of Information Systems and held the Union Pacific Professorship in Information Systems. Before joining Iowa State, she was a tenured faculty member in the Management Information Systems area in the Krannert School of Management (now the Mitchell E. Daniels, Jr., School of Business) at Purdue University for 17 years.

She holds a PhD and BSBA in Decision and Information Sciences from the Warrington College of Business at the University of Florida. Her early exposure to machine learning via expert systems led to a career-long interest in AI, particularly machine learning. Her doctoral dissertation focused on genetic algorithms as a computational model for how teams exchange ideas in search of a solution.

Her research focus is in Cybersecurity Risk Management and analytics, particularly machine

learning models. Her research has been published in journals such as *Information Systems Research (ISR), Management Information Systems Quarterly (MISQ), Journal of Management Information Systems (JMIS), ACM Transactions on Management Information Systems (ACM-TMIS), Decision Sciences (DS), Decision Support Systems (DSS), European Journal of Operational Research (EJOR)*, and several other outlets.

Her teaching interests include Introduction to Programming, Database Management Systems, Information Security Management, Business Analytics, and Strategic Use of Information Systems. Jackie enjoys “pulling the curtain back” on what can seem like mysterious technologies and empowering learners through the use of data, technology, and systems.

She serves on several boards of directors and is an active peer reviewer and committee member for AACSB.

She and her husband Mike live in Pickerington, Ohio, with the youngest of their three children and their dogs Bruno (Great Pyrenees) and Melon (Golden Doodle).

A Letter to New Adopters

Dear Colleague,

When I first set out to create *AI in Business: Creating Value Responsibly with Artificial Intelligence*, I wasn't trying to jump on the latest trend—I was trying to fill a void.

Like many instructors, I was tasked with adding an AI component to a core course. As I searched for materials that could help students understand how AI is used in the real world, I found no shortage of highly technical content. But I found very little that addressed the practical, ethical, and strategic challenges students and instructors face when utilizing AI.

Employers were asking for students with AI skills, and it quickly became clear that students from all academic backgrounds would need to engage with this subject. That meant I had to write about a technical topic for a nontechnical audience—without oversimplifying the underlying ideas. I began with a few chapters for my own class. The more I worked, the more fascinating the material became—and before long, those chapters evolved into a complete resource.

At the same time, I was serving on the Tippie College of Business Task Force on AI. We were confronting the reality that generative AI isn't a passing wave—it's a rising tide. Our graduates needed to understand this technology and be prepared to use it responsibly, both in the classroom and in the workplace.

As we developed the material, it became clear that questions surrounding ethics and sustainability were not side issues—they were central concerns repeatedly raised by educators, students, and industry professionals. Rather than isolate these topics in a single chapter (where they could be easily skipped or minimized), we chose to integrate them throughout the product. To make these discussions easier to identify and emphasize, we created dedicated icons that flag ethical and sustainability moments across chapters, prompting students to think critically about the societal impact of AI as they encounter each new concept.

Throughout this project, we were guided by a core insight from our years of teaching both undergraduate and graduate students: When students see the relevance of a topic to their own lives and futures, they engage more deeply. This is a central premise of adult learning theory: Adults learn best when they can connect what they're learning to real-world problems and decisions. That's why this product is filled with examples—more than 100 of them—drawn from companies large and small, across industries and geographies, and spanning both successes and failures.

Each chapter includes:

- Clear, practical learning objectives.
- A concise conceptual overview of key AI ideas.
- Case studies and "Quick Wins" that make the content immediately applicable.

- Ethical and societal issues that prompt critical thinking and class discussion.
- A chapter of the ongoing Singapore Smart Nation case, which follows a relatable young professional, Ari, as they encounter AI in everyday contexts.

Why These Chapters?

Since there's no universal syllabus for teaching AI in business, one of the biggest questions instructors face is: *What should we actually cover?* The chapters in this product reflect the arc we believe most instructors and students need: from broad understanding, to technical familiarity, to responsible and effective use.

Chapter 1 provides an accessible overview of AI and debunks the myth that AI only means generative AI. It introduces predictive, prescriptive, and generative AI and raises key ethical questions right from the start.

Chapter 2 offers historical context. Understanding AI's development helps students see how innovation diffuses into organizations and what factors—advances in theory, computing, and data—have accelerated AI's capabilities.

Chapter 3 focuses on data as the foundation of AI. It explores the lifecycle of data collection, privacy and security issues, and global regulatory concerns businesses must now navigate.

Chapters 4, 5, and 6 form the technical core, covering machine learning, neural networks, and generative AI, respectively. Each chapter explains how the method works, what business problems it solves, and where it falls short—supported by real-world examples from successful and failed implementations.

On the advice of reviewers from institutions across the country, we added optional **Technical Supplements** that contain more advanced content on algorithm logic, model training, and evaluation. These give instructors flexibility to adjust the depth of content based on their audience.

Chapter 7 explores how organizations implement AI—from assessing readiness to selecting projects, managing change, and understanding the risks of deployment.

Chapter 8 walks students through prompt engineering and the critical evaluation of AI-generated content. While it appears last in the product, I often teach it second so students can start applying AI tools from the beginning.

The result is a structure that's flexible, progressive, and grounded in the needs of students—not to become developers, but to become professionals who can think critically about AI, use it responsibly, manage its integrations, and apply it creatively across disciplines.

Along the way, we have been fortunate to receive feedback from faculty at institutions across the country who are navigating the same teaching challenges. Their input helped refine the tone, depth, and structure of this resource. We're grateful for their generosity—and their honesty.

This is a fast-moving field, and we recognize that yesterday's innovation can quickly become

today's baseline. That's why we've adopted an evergreen publishing model. Instead of waiting years for a new edition, we'll be updating materials regularly—so when you start a new semester, you can work with the most current examples, insights, and tools. This online space will also include supplemental content that extends beyond the core chapters, offering flexibility without interrupting the product's flow.

Thank you for the work you do—and for considering this resource for your course.

Warm regards,

Patrick Johanns

Associate Professor of Instruction

Tippie College of Business

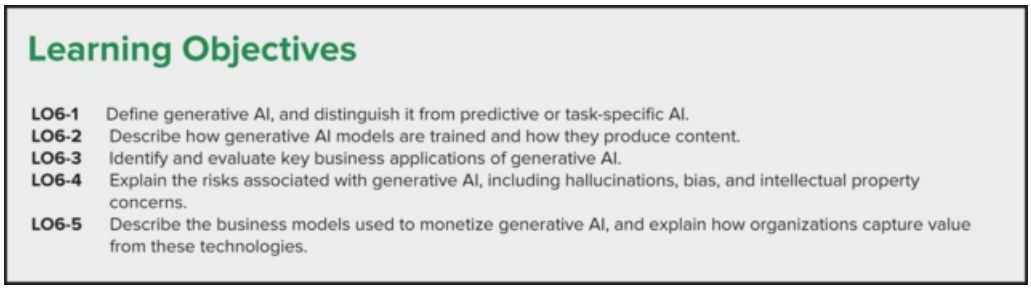
University of Iowa

Guided Tour

Learning Objectives

Each chapter begins with focused, clearly stated learning objectives that help students understand what to expect, guide their attention, and self-assess their understanding. The objectives are grounded in Bloom's Taxonomy and emphasize skills students need to thrive in an AI-enabled world, such as analytical thinking, ethical judgment, strategic awareness, and using generative AI tools.

The objectives are **sequenced logically**, from foundational concepts to more applied or forward-looking topics. Each is reinforced in the chapter content, summary, and end-of-chapter questions.



Learning Objectives

LO6-1 Define generative AI, and distinguish it from predictive or task-specific AI.
LO6-2 Describe how generative AI models are trained and how they produce content.
LO6-3 Identify and evaluate key business applications of generative AI.
LO6-4 Explain the risks associated with generative AI, including hallucinations, bias, and intellectual property concerns.
LO6-5 Describe the business models used to monetize generative AI, and explain how organizations capture value from these technologies.

Focus on Ethics and Sustainability

The core message of this product is reflected in its title: *AI in Business: Creating Value Responsibly.* AI holds tremendous promise—but it must be implemented thoughtfully and responsibly to realize its full potential. Each chapter weaves together three guiding themes:

1. ***Creating value across industries and around the world***

From agriculture to finance, health care to entertainment, AI is transforming how businesses operate. Students engage with more than 100 examples from companies of all sizes worldwide, seeing how AI drives innovation, improves decision making, and creates a competitive advantage.

2. ***Ethics and society***

AI raises significant ethical challenges, including bias in algorithms and datasets, surveillance, job displacement, transparency, and potential misuse. Rather than present one-sided answers, this resource highlights ongoing debates and encourages students to wrestle with questions of fairness, accountability, and human impact—through discussion prompts, examples, and case studies. We direct attention to an ethical discussion with our Ethics icon, as shown in this excerpt from **Chapter 7**:



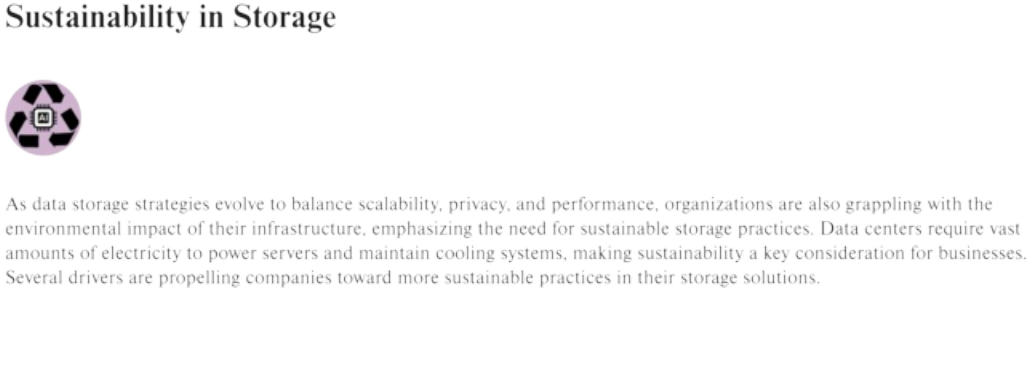
BASIC ETHICAL, LEGAL, AND REGULATORY CONSIDERATIONS IN AI IMPLEMENTATION (LO7-5)

Implementing AI responsibly requires organizations to address a range of ethical, legal, and regulatory considerations. These challenges apply across many types of AI, including machine learning models, natural language processing tools, and robotic process automation systems. Core concerns—such as bias, data privacy, explainability, accountability, and workforce impact—are not limited to any single AI technology.

This section outlines foundational practices for deploying AI ethically and effectively in business environments. It builds on our ethics discussions in other chapters and includes strategies for reducing unintended bias, protecting individual rights, promoting transparency in decision making, and ensuring that the use of AI aligns with regulatory expectations and organizational values.

3. ***Sustainability***

AI is both part of the sustainability challenge and a potential solution. Students explore the environmental costs of training large models, the implications of massive data infrastructures, and the promise of AI in addressing climate and resource challenges. From energy-efficient supply chains to innovative recycling, the book presents a balanced perspective of issues in sustainability. The sustainability icon can be seen in an excerpt from **Chapter 3**:



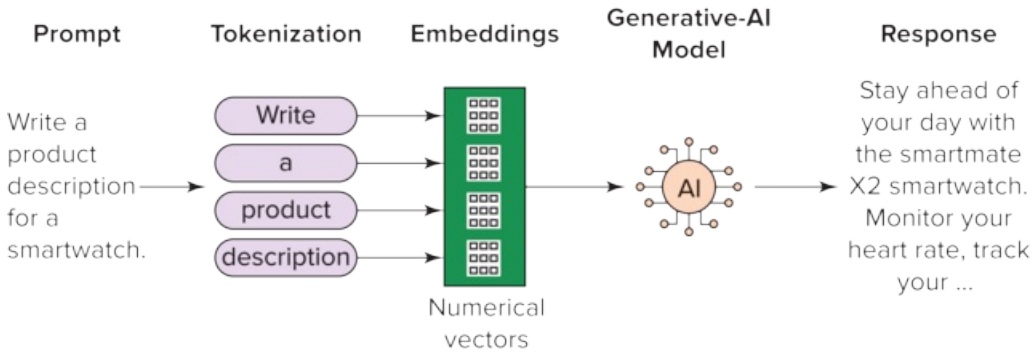
Sustainability in Storage

As data storage strategies evolve to balance scalability, privacy, and performance, organizations are also grappling with the environmental impact of their infrastructure, emphasizing the need for sustainable storage practices. Data centers require vast amounts of electricity to power servers and maintain cooling systems, making sustainability a key consideration for businesses. Several drivers are propelling companies toward more sustainable practices in their storage solutions.

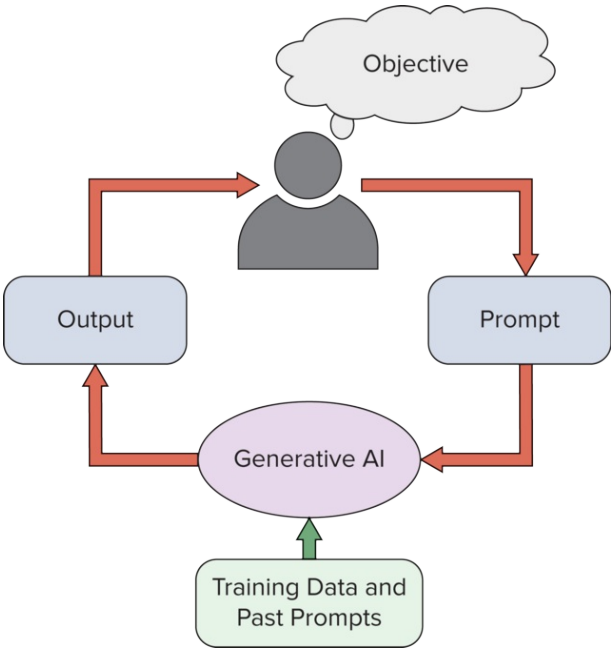
Visual Diagrams

The book uses custom visuals to simplify complex ideas and show human–AI interactions. Two highlights:

- **Figure 6.3** A simplified diagram of how generative AI works



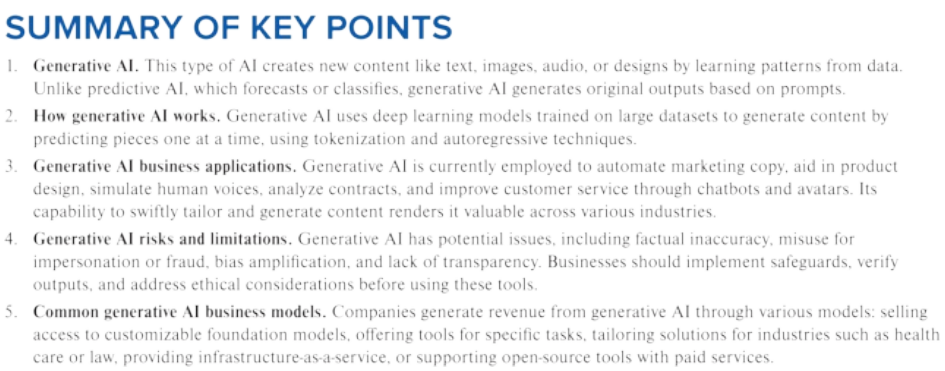
- **Figure 8.1** A step-by-step view of prompt engineering with humans in the loop



Summary of Key Points

Each chapter ends with a concise list of key takeaways, directly aligned with the learning objectives. These help students review quickly and give instructors a ready reference for

assessment. This Summary of Key Points from Chapter 6 shows the one-to-one correspondence with the learning objectives previously shared.



SUMMARY OF KEY POINTS

1. **Generative AI.** This type of AI creates new content like text, images, audio, or designs by learning patterns from data. Unlike predictive AI, which forecasts or classifies, generative AI generates original outputs based on prompts.
2. **How generative AI works.** Generative AI uses deep learning models trained on large datasets to generate content by predicting pieces one at a time, using tokenization and autoregressive techniques.
3. **Generative AI business applications.** Generative AI is currently employed to automate marketing copy, aid in product design, simulate human voices, analyze contracts, and improve customer service through chatbots and avatars. Its capability to swiftly tailor and generate content renders it valuable across various industries.
4. **Generative AI risks and limitations.** Generative AI has potential issues, including factual inaccuracy, misuse for impersonation or fraud, bias amplification, and lack of transparency. Businesses should implement safeguards, verify outputs, and address ethical considerations before using these tools.
5. **Common generative AI business models.** Companies generate revenue from generative AI through various models: selling access to customizable foundation models, offering tools for specific tasks, tailoring solutions for industries such as health care or law, providing infrastructure-as-a-service, or supporting open-source tools with paid services.

Quick Wins

Short, actionable insights help students connect AI theory to workplace reality. These real-world takeaways build confidence and reinforce that AI isn't just something to learn—it's something to use.

Case Studies

Each chapter features two types of cases to support applied learning. First, three standalone business cases (500–750 words) align with learning objectives and showcase diverse real-world examples. Second, a continuing narrative—the Integrated Case Series: Singapore's Smart Nation Initiative—follows a relatable protagonist through various sectors, helping students understand how AI affects everyday life and society.

End-of-Chapter Cases

Each chapter concludes with three cases (500–750 words each), aligned to the learning objectives and drawn from diverse industries and geographies. Students examine both successes and failures.

Examples include:

- Netflix's use of deep learning for recommendations.
- TradeLens and the challenges of AI-blockchain integration.
- An exploration of the use of GenAI art in producing board games.

Integrated Case Series: Singapore's Smart Nation Initiative

Each chapter is tied to a portion of the **Singapore Smart Nation Initiative Integrated Case**, a multipart story designed to humanize and contextualize AI's role in modern society. At the center is *Ari*, a young data professional navigating how AI is embedded in transportation, health care, education, sustainability, and more.

Ari is intentionally written as someone students can **see themselves in—**early in their careers, curious, thoughtful, and learning to apply AI responsibly and effectively in the real world.

The Integrated Case Series helps students:

- **Connect personally** with the material through a consistent, relatable protagonist.
- See how **AI is being integrated into daily life,** not just a technical tool.
- Understand how **sectors are interdependent** in an AI-enabled society.
- Reflect on the **societal, strategic, and ethical implications** of technology beyond business use.

Your complete course platform

Connect enables you to build deeper connections with your students through cohesive digital content and tools, creating engaging learning experiences. We are committed to providing you with the right resources and tools to support all your students along their personal learning journeys.

90%
student satisfaction



Laptop: Getty Images; Woman/dog: George Doyle/Getty Images

Every learner is unique

In Connect, instructors can assign an adaptive reading experience with SmartBook®. Rooted in advanced learning science principles, SmartBook delivers a personalized experience, focusing students on their learning gaps and ensuring the time they spend studying is time well spent. **mheducation.com/highered/connect/smartbook**

Study anytime, anywhere

Encourage your students to download the free ReadAnywhere® app so they can access their online eBook, SmartBook, or Adaptive Learning Assignments when it's convenient, even when they're offline. Because the app automatically syncs with students' Connect accounts, all their work is available every time it is opened. **mheducation.com/readanywhere**

> ***"I really liked this app—it made it easy to study when you don't have your textbook in front of you."***
> Jordan Cunningham, a student at *Eastern Washington University*

Effective tools for efficient studying

Connect is designed to help students be more productive with simple, flexible, intuitive tools that maximize study time and meet students' individual learning needs and prepare them with real-world application. Get learning that works for everyone with Connect.



Education for all

McGraw Hill works directly with Accessibility Services departments and faculty to meet the learning needs of all students. Please contact your Accessibility Services Office, and ask them to email **accessibility@mheducation.com**, or visit **mheducation.com/about/accessibility** for more information.

Affordable solutions, added value

Make technology work for you with LMS integration for single sign-on access, mobile access to the digital textbook, and reports to quickly show you how each of your students is doing. And with Affordable Access programs, you can provide all these tools at the lowest available market price to your students. Ask your

Solutions for your challenges

A product isn't a solution. Real solutions are affordable, reliable, and come with training and ongoing support when you need it and how you want it. Visit **supportateverystep.com** for videos and resources both

McGraw Hill representative for more information.

you and your students can use throughout the term.



Updated and relevant content

Our Evergreen™ delivery model provides you with the most relevant and updated content, content, tools, and accessibility. Engage students and freshen up assignments with up-to-date coverage of select topics and new questions, all without having to switch editions or build a new course.



Acknowledgments

We'd like to thank our reviewers who provided invaluable feedback throughout the development process of this first release.

Haroon Abbu, North Carolina State University

Carlos Anato, Miami Dade College

Solomon Antony, Murray State University

Mahmoud Badr, SUNY Polytechnic Institute

Isabel Bo-Linn, Portland State University

Matt Brown, Arkansas Tech

Alexander Chen, University of Central Arkansas

Xu Chen, Duke University

Sai Chand Chintala, Cornell University

Ian Cross, Bentley University

Rushit Dave, Minnesota State University–Mankato

Jon Down, University of Portland

Weiguo Fan, University of Iowa

Mark Finlayson, Florida International University

Amy Hill, Tennessee Technological University

Rachael Hutchings, Utah Valley University

Gavin Johnson, Texas A&M University–Commerce

Keigo Kawaji, Illinois Institute of Technology

Anthony Kelly, George Mason University

Michael Koenig, Seattle University

Quinn Koller, Utah Valley University

Fabrice Laguerre, Miami Dade College

Adam Lee, University of Maryland

Maikel Leon, University of Miami

Alejandro Lopez Lira, University of Florida

Kevin Matthews, University of North Carolina–Wilmington

Maureen McWhite, Florida A&M University

Christy Mesaros-Winckles, Adrian College

Mariofanna Milanova, University of Arkansas at Little Rock

Debi Mishra, Binghamton University

Kenneth Murphy, University of California–Irvine
Pankaj Nagpal, Arkansas State University
Jody Paul, Metropolitan State University of Denver
Rosa Maria Perez McNulty, MDC North
Brennan Pursell, DeSales University
Veronica Ramsundar, Miami Dade College
Shilpa Rao, Virginia Tech
Vernon Richardson, University of Arkansas
Mark Schmidt, St. Cloud State University
David Schweidel, Emory University
Dmitriy Shaltayev, Christopher Newport University
Heather Stewart, Oklahoma State University
Larry Tang, University of Central Florida
Casey Wilhelm, University of Idaho

We are grateful for the guidance of our McGraw Hill team—Darin Kelsey, David Ploskonka, and Kristen Salinas-Thompson—whose expertise helped shape a raw manuscript into a polished resource for instructors. We also thank Shannon LeMay-Finn of Piper Content Partners for her thoughtful editing and keen eye for clarity and cohesion.

Brief Contents

Chapter 1 Overview of AI in Business 1
Chapter 2 The Rise of AI: Innovation and Impact 21
Chapter 3 Data Management for AI 43
Chapter 4 Machine Learning 70
Chapter 4S Machine Learning Technical Supplement 89
Chapter 5 Neural Networks and Deep Learning 112
Chapter 5S Neural Networks and Deep Learning Concepts Technical Supplement 128
Chapter 6 Generative AI in Business: Transforming Industries with Innovative Technology 133
Chapter 6S Generative AI Technical Supplement 152
Chapter 7 Strategic AI Implementation: People, Process, and Purpose 157
Chapter 8 The Art of Prompt Engineering with FOCUS 190

AI in Action: The Singapore Case Series 224

Detailed Table of Contents

Chapter 1 Overview of AI in Business 1
Introduction 1
Types of Artificial Intelligence 2
Machine Learning 2
Neural Networks 3
Generative AI 4
Other Applications of AI in Business 5
The Use of Artificial Intelligence in Virtual Reality 5
Computer Vision and Its Role in Business Applications 6
Speech Recognition and Its Business Impact 6
Data Science and AI 6
Ethical Considerations in AI 8
Data Ownership, Privacy, and Consent 8
Environmental Sustainability 9
Economic Inequality and AI Access 9
Bias and Discrimination 10
Manipulation and Deception 11
Intellectual Property and AI Creativity 11
Transparency and Accountability 12
Summary of Key Points and Quick Wins 13
Quick Wins 13
Key Terms 14
Discussion Questions 14

Exercises 15
Case Study 1.1: *The Portrait of Edmond de Belamy*—Intellectual Property and AI Creativity 15
U.S. and French Copyright Laws 15
Discussion Questions 16
Case References 16
Case Study 1.2: Coded Bias 16
Discussion Questions 17
Case References 17
Case Study 1.3: Closing the Loop—How H&M and HKRITA Use AI to Revolutionize Textile Recycling 17
What Happens after Sorting? 17
Insights 18
AI's Growing Presence in Physical, Traditionally Manual Industries 18
The Potential of AI for Sustainability and Circular Economies 18
The Importance of Data-Rich Environments 18
AI in Long-Term Environmental Goals 18
Discussion Questions 18
Case References 19
Endnotes 19

Chapter 2 The Rise of AI: Innovation and Impact 21
Introduction 22
Foundations of AI (Pre-1950s) 22
Artistic Foundations 22
Scientific Foundations: Early Thinkers and Theories 23
The Dawn of AI (1950–1970) 23
The Turing Test: Can Machines Think? 23
The Dartmouth Conference 23
Early AI Programs 24
Emerging Computer Technologies 25
Impact on Business Practices 25

The Age of Experimentation (1970–1985) 25
Expert Systems 26
Advancements in Neural Networks 26
The Microprocessor Revolution and Technological Foundations 26
Impact of AI on Business from 1970–1985 26
The AI Winter and Resurgence (1985–2000) 27
The Resurgence: Seeds of Renewal 27
Technological Foundations 27
AI and Robotics in Business (1985–2000) 28
The Age of Big Data and Machine Learning (2000–2020) 28
Technological Foundations 29
Business Applications 29
The Generative AI Revolution (2020–Present) 30
Scientific and Technological Advancements 31
AI and the Internet of Things: A Seamless, Intelligent World 31
Ethical, Safety and Privacy Concerns in AI 32
AI's Future: Opportunities and Challenges 33
Summary of Key Points and Quick Wins 34
Quick Wins 35
Key Terms 35
Case Study 2.1: AI and COVID-19 36
AI's Role in the COVID-19 Response 36
The Pandemic's Impact on AI Adoption 37
Discussion Questions 37
Case References 38
Case Study 2.2: Project Innereye—expanding AI in Medicine 38
Discussion Questions 39
Case References 39
Case Study 2.3: Pioneering Womens Sports Analytics with AI 39
AI's Journey in Sports Analytics 40
A Historical First for Women's Sports 40

Impact on the Game 40
Discussion Questions 40
Case References 41
Endnotes 41

Chapter 3 Data Management for AI 43
Introduction 43
Data for AI 44
Data Structures and Types 44
Sources of Data for AI 44
Data Security 46
Data Security Challenges 46
AI Data Security Principles 48
Data Quality and Integrity 49
High Quality Data: Characteristics and Methods 49
Data Cleaning 50
Building a Data-Ready Organization 50
Data Regulations 51
Broad Governmental Policies 51
Industry-Specific Data and AI Regulations 53
Big Data and Data Management 54
Larger Data Centers 54
Distributed Cloud-Based Storage Systems 54
Sustainability in Storage 55
Emerging Technologies in AI Data Management 56
Blockchain 56
AI-Powered Cybersecurity 57
Edge Computing and Storage 58
Federated Learning 59
Ethical Considerations in AI Data Collection and Use 60
Summary of Key Points and Quick Wins 61

Quick Wins 61
Key Terms 62
Discussion Questions 63
Exercises 63
Case Study 3.1: Amazon—AI in the Hiring Process 63
The Start of the Journey 63
Clues of Bias 64
Repercussions and Repairs 64
Discussion Questions 64
Case References 64
Case Study 3.2: Darktrace—AI in Network Security 65
Benefits and Challenges 65
Broader Implications 65
Discussion Questions 65
Case References 66
Case Study 3.3: TradeLens—Blockchain as a Solution, but Why Did It Fail? 66
Blockchain as a Solution 66
An Unexpected Failure 67
Lessons from Failure 67
Conclusion 67
Discussion Questions 67
Case References 68
Endnotes 68

Chapter 4 Machine Learning 70
Introduction 70
How Machine Learning Models Learn 71
Supervised Learning 71
Unsupervised Learning 72
Reinforcement Learning 72
Machine Learning in Business Functions 74

Data Management and Preprocessing 75
Prediction and Forecasting 76
Classification and Risk Assessment 77
Customer Segmentation and Personalization 77
Process Optimization and Automation 78
Comparing Standard Machine Learning Algorithms 78
Ethical Risks of Machine Learning in Business 79
Algorithmic Basis in Machine Learning 80
Strategies for Reducing Bias in Machine Learning Systems 80
Other Concerns 80
Summary of Key Points and Quick Wins 80
Quick Wins 81
Key Terms 81
Discussion Questions 82
Exercises 82
Case Study 4.1: Netflix Recommendation System 82
Netflix's History 83
The Heart of Netflix: Its Recommender Grid 83
The Data Driving the System 83
Challenges and Ethical Considerations 84
Business Impact 84
Discussion Questions 84
Case References 84
Case Study 4.2: Flights Delays at JFK 84
Implementation 85
Discussion Questions 85
Case References 86
Case Study 4.3: UPS Vehicle Routing 86
Challenges for Logistics Companies 86
UPS and ORION: AI for Efficiency 86
The Left-Turn Problem 87

Real-Time Optimization and Cost Savings 87
Conclusion 87
Discussion Questions 87
Case References 87
Endnotes 88

Chapter 4S Machine Learning Technical Supplement 89
Principal Component Analysis 89
How PCA Works 89
Business Applications of PCA 89
Advantages of PCA 90
Limitations of PCA 91
K-Means Clustering 91
How*k*-Means Clustering Works 91
Business Applications of *k*-Means Clustering 92
Advantages of *k*-Means Clustering 92
Limitations of *k*-Means Clustering 92
K-Nearest Neighbors 93
How *k*-NN Works 93
Business Applications of *k*-NN 93
Advantages of *k*-NN 94
Limitations of *k*-NN 94
Linear Regression 94
How Linear Regression Works 95
Business Applications of Linear Regression 95
Advantages of Linear Regression 95
Limitations of Linear Regression 95
Decision Trees 95
How Decision Trees Work 96
Business Applications of Decision Trees 96
Advantages of Decision Trees 97

Limitations of Decision Trees 97
Random Forests 97
How Random Forests Work 98
Business Applications of Random Forests 98
Advantages of Random Forests 98
Limitations of Random Forests 99
Logistic Regression 99
How Logistic Regression Works 99
Business Applications of Logistic Regression 100
Advantages of Logistic Regression 100
Limitations of Logistic Regression 100
Support Vector Machines 100
How SVM Works 101
Business Applications of SVMs 101
Advantages of SVMs 101
Limitations of SVMs 101
Collaborative Filtering 102
How Collaborative Filtering Works 102
Business Applications of Collaborative Filtering 102
Advantages of Collaborative Filtering 102
Limitations of Collaborative Filtering 103
Time Series Analysis 103
How Time Series Analysis Works 103
Business Applications of Time Series Analysis 104
Advantages of Time Series Analysis 104
Limitations of Time Series Analysis 104
Metrics for the Quality of Algorithms 104
Confusion Matrix 105
Metrics 105
Summary 110
Algorithm Selection Questions 110

Metrics and Evaluation Selection Questions 110

Chapter 5 Neural Networks and Deep Learning 112
Introduction 112
Clarifying Neural Networks and Deep Learning 113
The Growing Importance of Neural Networks 113
Business Applications of Neural Networks 114
Why Businesses Are Adopting Neural Networks 114
Real-World Impacts 114
Deep Learning 115
Deep Learnings Role in Business 115
Deep Learning in Natural Language Processing 115
NLP Applications in Business 117
Limitations of Neural Networks and Deep Learning 118
Emerging Capabilities in Deep Learning 118
Foundation Models 118
Multimodal Learning 118
Few-Shot and Zero-Shot Learning 119
Energy-Efficient Deep Learning (Green AI) 119
Deep Learning for Personalized Adaptive Systems 119
Deep Learning and Virtual Reality 119
Growth of Neural Network Models 120
Summary of Key Points 120
Quick Wins 121
Key Terms 121
Discussion Questions 121
Exercises 122
Case Study 5.1: IBM Watson OpenScale—Explainable AI 122
IBM Watson and OpenScale 122
Challenges and Tradeoffs 123
Conclusion 123

Discussion Questions 123
Case References 123
Case Study 5.2: Tesla Autopilot—The Promise and Peril of Deep Learning-Driven Autonomy 123
The Market for Autonomous Vehicles 123
Deep Learning and Tesla's Autonomy Strategy 124
Challenges in Technology and Regulation 124
Legal Difficulties 124
Conclusion 125
Discussion Questions 125
Case References 125
Case Study 5.3: Facial Recognition Technology and the Detroit Lions 125
Technology Overview 125
Applications and Business Value 126
Discussion Questions 126
Case References 126
Endnotes 126

Chapter 5S Neural Networks and Deep Learning Concepts Technical Supplement 128
Traditional Machine Learning vs. Deep Learning 128
How Neural Networks Work 129
Neural Network Architecture 129
Activation Functions 129
Forward Propagation and Backpropagation 129
Training with Stochastic Gradient Descent 130
Office Building Analogy 130
Overfitting 130
Deep Learning Model Types 130
Convolutional Neural Networks: Specialists in Visual Data 130
Recurrent Neural Networks: Specialists in Sequences and Time Series 131
Transformers and Attention Mechanisms 131

What Are Transformers? 131
How Attention Mechanisms Work 131
Transformers in Modern AI Applications 132
Conclusion 132
Endnotes 132

Chapter 6 Generative AI in Business: Transforming Industries with Innovative Technology 133
Introduction 133
What Is Generative AI? 134
Predictive vs. Generative AI 134
Types of Generative AI 136
How Generative AI Works 136
Training Generative Models 136
Text Generation: Step by Step 137
Image Generation 137
Multimodal and Video Generation 138
Business Applications of Generative AI 138
Text-Based Applications 138
Custom GPTs 139
Chatbots and Customer Engagement 140
Generative Visual Content and Virtual Try-On 140
Product Design 141
Generative AI in Contract Review 141
Training and Onboarding 141
Multimodal and Interactive Experiences 141
Specialized Use Cases across Industries 141
Risks and Limitations of Generative AI 142
Hallucinations and Factual Inaccuracy 142
Bias and Fairness 142
Intellectual Property and Legal Concerns 142

- Transparency and Explainability 142
- Ethical Use and Governance 142
- Work Role Changes 142
- Business Models in Generative AI 142
- Summary of Key Points 144
  - Quick Wins 144
- Key Terms 145
- Discussion Questions 145
- Exercises 146
- Case Study 6.1: When AI Makes False Claims—The OpenAI Libel Controversy 146
  - Discussion Questions 147
  - Case References 147
- Case Study 6.2: AI Voice Cloning 147
  - How It Works 147
  - Uses—and Abuses 147
  - When Voice Cloning Helps 148
  - Grief and the Gray Areas 148
  - Legal and Ethical Challenges 148
  - Discussion Questions 148
  - Case References 148
- Case Study 6.3: Revolutionizing Soccer Tactics with Tacticai 149
  - The Role of TacticAI 149
  - How It Works 149
  - Business and Competitive Value 150
  - Potential Risks and Limitations 150
  - The Future of Sports AI 150
  - Discussion Questions 150
  - Case References 150
- Endnotes 150

- Chapter 6S Generative AI Technical Supplement 152

Key Technical Components 152
Tokenization and Embeddings 152
Transformer Architectures and Autoregressive Generation 153
Core Generative Model Types 153
Transformers 153
Diffusion Models 153
Generative Adversarial Networks 154
Variational Autoencoders 154
Recurrent Neural Networks 155
How Generative AI Balances Creativity and Coherence 155
Sampling Strategies in Generative AI 155
Generative AI Isn’t Done Evolving 156
Endnotes 156

Chapter 7 Strategic AI Implementation: People, Process, and Purpose 157
Introduction 158
Assessing Organizational Readiness for AI 158
Data Readiness 159
Integrating Data across the Organization 159
Technological Readiness 160
Business Processing Readiness 160
Workforce Readiness 161
Cultural and Leadership Readiness 162
Identifying Business Functions Ready for AI Integration 163
Customer Service 164
Marketing 164
Operations and Supply Chain Management 165
Other Departmental Applications 165
Aligning AI Projects with Business Objectives 166
Implementing AI Solutions 167
Selecting the Right AI Tools 167

Selecting the Right AI Vendors 169
Monitoring and Measuring AI Effectiveness 171
Innovation and Creativity Metrics 171
Operational Efficiency Metrics 172
Customer Experience Metrics 172
Business Outcome Metrics 172
Balanced Scorecard Approach 173
Basic Ethical, Legal, and Regulatory Considerations in AI Implementation 173
Curate Diverse and Ethical Training Data Biases 174
Ensure Explainability 174
Promote Transparency 174
Continually Assess Risks and Benefits 175
Ensure Accountability 175
Plan for Workforce Impact 175
Legal and Regulatory Considerations 175
A Final Note 176
Special Considerations for Generative AI 176
Ethical Foundations for Generative AI 176
Reducing Hallucinations and Misinformation 176
Managing Workforce Disruptions 177
Safeguarding User Consent and Control 177
Navigating Copyright and Intellectual Property Risks 178
Legal Accountability in Generative AI Systems 178
Summary of Key Points 179
Quick Wins 179
Key Terms 179
Discussion Questions 180
Exercises 180
Case Study 7.1: Mastercard Fraud Detection
AI as a Solution 181
Implementation Decision 181

The Launch of Decision Intelligence 182
Outcomes 182
AI Expansion beyond Fraud 182
Takeaways 182
Discussion Questions 182
Case References 183
Case Study 7.2: Pfizer-BioNTech COVID-19 Vaccine 183
A Compressed Timeline—Enabled by AI 183
AI Contributions across the Development Pipeline 183
Why It Worked 184
Ethical and Operational Considerations 184
Conclusion 184
Discussion Questions 184
Case References 185
Case Study 7.3: *The New York Times* and Google Cloud AI 185
Digitization: A Human-Driven Process 185
AI Vendor Choice 185
From Preservation to Personalization 186
Discussion Questions 186
Case References 186
Endnotes 187

Chapter 8 The Art of Prompt Engineering with FOCUS 190
Introduction 190
The Focus Framework for Effective Prompts 191
General Principles of Prompt Crafting 191
Hard Prompts: Benchmarks for AI 198
Iterative Refinement 198
Prompt Refinement Example 199
Understanding AI Models 202
Closed-Domain Chatbots 202

Open-Domain Chatbots 202
How Chatbots Explain Themselves 202
Advance Prompting Techniques 204
Task Decomposition 204
Additional Techniques and Strategies 206
Communicating Effectively with Business AI Tools 208
Personas in Prompt Engineering 208
Choosing the Right Persona 208
Using Personas Effectively 210
Potential Risks of Persona-Driven Prompts 210
Strategies for Verifying AI Responses 211
Ethical Considerations in Prompt Engineering 211
Ethical Prompt Crafting 212
Verification 212
Ethical Use of Generative AI Output 213
Intellectual Property and Creativity 213
Intellectual Property and AI 213
Ethical Prompting Checklist 213
The End Is Just the Beginning 214
Summary of Key Points 214
Quick Wins 214
Key Terms 215
Discussion Questions 215
Exercises 215
Case Study 8.1: 2023 Writers Guild of America Strike 217
Writers' Concerns about AI 218
The Studios' View 218
Legal and Ethical Implications 218
Resolution and Aftermath 218
Discussion Questions 219
Case References 219

Case Study 8.2: Generative AI Art in Board Games 219
Jenna's Decision: *Legends Unbound* 219
Business Opportunities 220
Ethical Challenges 220
Conclusion 220
Discussion Questions 220
Case References 221
Case Study 8.3: Hallucinating the Law 221
What Went Wrong? 221
Could a Persona Have Helped? 221
Open vs. Closed Chatbots: Choosing the Right Tool 221
Prompt Engineering Lessons 222
The Legal and Ethical Fallout 222
The Bigger Picture 222
Discussion Questions 222
Case References 222
Endnotes 223

AI in Action: The Singapore Case Series 224
Introduction 224
Where Smart Infrastructure Meets Smart Business 224
A Seamless Morning in Singapore's Smart Nation 224
Smart Home Start 224
Stepping into the Community 224
Public Infrastructure in Sync 224
Commute with Intelligence 225
Smart Commerce and Customer Experience 225
Workplace and Reflection 225
Discussion Questions 225
Case References 225
Foundations of a Smart Nation 226

A Glimpse of the Future 226
The Three Pillars of the Smart Nation Vision 226
Challenges and Opportunities in 2014 226
Singapore's Strategic Shift toward AI 227
Discussion Questions 227
Case References 227
Singapore's Data Engine 227
The Foundation of Data Infrastructure 227
Governance and Trust 228
From Data to Impact 228
Looking Ahead 228
Discussion Questions 228
Case References 229
Singapore's Intelligent Spaces 229
Ari at the Bosch Campus 229
Singapore's Green Plan and the Role of Machine Learning 229
The Sensor-Driven Transformation 230
Training the Model 230
Scaling toward a Smart Nation 230
Discussion Questions 230
Singapore's Intelligent Transportation System 231
A Road That Thinks Ahead 231
Neural Networks in the Infrastructure 231
Making a Case for the Human Impact 231
A System That Sees Too Much? 231
A System That Finds Parking Too 231
A System That Clears the Way 231
The Future in the Foundations 232
Discussion Questions 232
Case References 232
AI in Urban Design 232

- A Smarter Start to Design 232
- Prompting Smarter Urban Design 233
- Bias in the Training Data 233
- Ethical and Technical Questions 233
- The Bigger Picture 233
- Discussion Questions 232
- Case References 232

Personalizing Government: The LifeSG Project 234

- A New Kind of Public Service 234
- From Moments of Life to LifeSG: A Platform Evolves 234
- Pilot Integration and AI Functionality 234
- Reshaping Public Service Roles 235
- Addressing Explainability and Trust 235
- Citizen Adoption and Community Engagement 235
- Smart Cities around the World 235
- Discussion Questions 235
- Case References 236

Telling the Smart Nation Story 236

- Ari's New Assignment 236
- Calling in AI for Assistance 236
- Applying the FOCUS Framework 236
- The Hallucination Test 237
- Breaking Down the Project 237
- Ari at the Podium 237
- Discussion Questions 235

Endnotes 237

Glossary G-1

Index IN-1

PRINT EDITION

CHAPTER 1
Overview of AI in Business



Jackie Niam/Shutterstock

Learning Objectives

After studying Chapter 1, you should be able to

- **LO1-1** Describe types of artificial intelligence (AI) and their relevance in business contexts.
- **LO1-2** Describe novel applications of AI in business, including virtual reality (VR), computer vision, and speech recognition.
- **LO1-3** Explain the importance of data science to AI.
- **LO1-4** Identify the range of ethical considerations for implementing AI in business.

INTRODUCTION

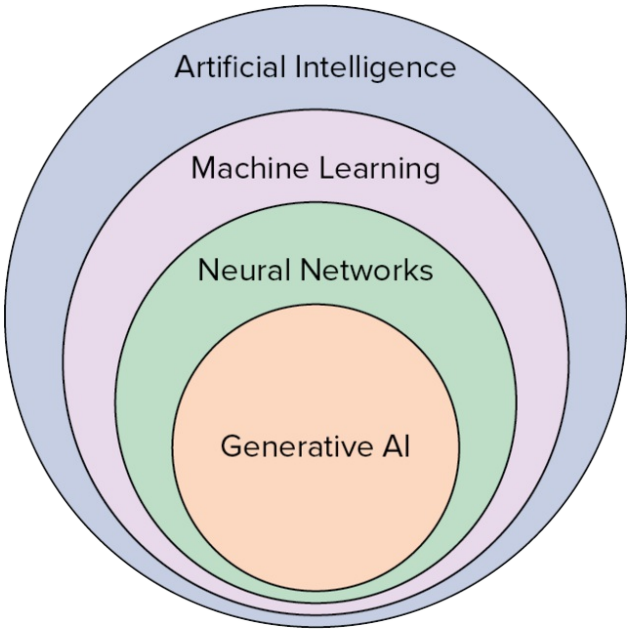
Artificial intelligence (AI), the ability of machines to simulate human intelligence, including learning, problem solving, and decision making, is transforming industries around the globe. Businesses in every sector, from retail and finance to health care and manufacturing, implement AI to streamline processes, enhance decision making, and gain a competitive advantage. Meanwhile, AI is undergoing rapid evolution. Consumers, businesses, and managers (and even writers of AI books) must continually monitor changes and update their knowledge and skills.

This chapter introduces the primary types of AI algorithms and their relevance in business. It presents applications of AI in business, emphasizes the importance of quality data to create effective AI solutions, and illustrates examples where AI technologies have enabled companies to innovate and create competitive advantages. These advances have many impacts on society, so we conclude the chapter by identifying ethical issues associated with AI, which we explore throughout the book.

TYPES OF ARTIFICIAL INTELLIGENCE (LO1-1)

There are three broad classifications of AI, as shown in **Figure 1.1**: *machine learning*, *neural networks*, and *generative AI*. In this section, we discuss each of these classifications and their relevance in business contexts.

FIGURE 1.1 Classifications of AI



Source: https://www.insight.com/en_US/content-and-resources/blog/unpacking-gen-ai–demystifying-the-current-artificial-intelligence-landscape.html

Machine Learning

Machine learning is a subset of AI that focuses on developing algorithms that allow computers to learn from and make predictions based on data. Instead of being explicitly programmed to perform a task, machine learning systems use statistical techniques to improve their performance over time. Common applications include spam filtering, recommendation systems, and image recognition.

Think of machine learning as teaching a child to ride a bike. At first, the child may wobble and fall, but with practice (the training phase), they notice patterns in balance and pedaling that keep them upright. Once they have mastered it, they can ride on their own (the deployment phase)—

even trying different terrains and speeds. Similarly, machine learning involves showing a computer a vast amount of information to learn to predict and decide on its own, like recognizing a song genre from audio waves or sorting emails into “spam” or “not spam.”

There are a variety of machine learning techniques. Just as a child uses different techniques to learn—such as training wheels for balance when learning to ride a bike—machine learning has methods for different tasks, including *regression algorithms*, *clustering*, and *classification* (topics we discuss in more detail in **Chapter 4**).

A Machine Learning Application

New Year’s Eve is a prime time for celebrations, but large parties can sometimes lead to property damage, noise complaints, and other disturbances. Airbnb faced significant challenges when some guests rented homes near their residences to host parties that escalated out of control. These events resulted in negative publicity and posed legal risks, threatening Airbnb’s business model.[1] Thus, in 2020, Airbnb introduced an anti-party initiative powered by machine learning. The AI uses predictive algorithms to assess booking patterns, incorporating variables such as the length of stay, proximity of the rental to the guest’s home, and timing of the reservation to flag and block high-risk bookings. By proactively addressing potential issues, Airbnb aimed to protect both hosts and guests.



TonTectonix/Shutterstock

Airbnb uses machine learning to keep parties off host properties.

The results have been noteworthy. Since Airbnb implemented this system, the rate of parties reported over New Year’s Eve decreased by over 60 percent where measures were in place. During the 2022 New Year’s Eve period, nearly 74,000 people globally were deterred from making high-risk bookings, enhancing safety and security.[2] This program’s success highlights machine learning’s power in mitigating risks and building trust in digital platforms.

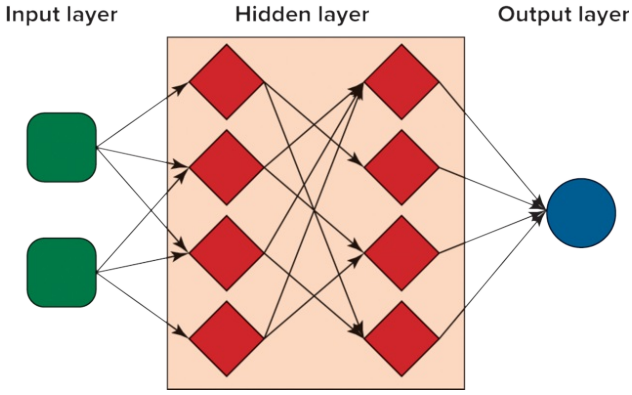
We cover machine learning methods in more depth in **Chapter 4**.

Neural Networks

Neural networks are a series of algorithms designed to recognize patterns and interpret data by mimicking how the human brain operates. They form the backbone of many AI systems. Inspired by the structure and function of the human brain, a neural network comprises units or "neurons" arranged in layers. These neurons receive input data and process it through a system of *weighted connections*, numerical values representing the strength or importance of each connection between neurons—similar to how synapses signal the strength of connections between neurons.

As shown in **Figure 1.2**, the first layer (the *input layer*) receives the raw input data, and each subsequent layer (called a *hidden layer*) uses the output from the previous layer as its input, gradually refining and transforming the data. The top layer is called the *output layer*. It generates the final predictions or outputs of the network based on the processed information from the preceding layers.

FIGURE 1.2 Neural Network Layers



Neural network diagram. Science Learning Hub – Pokapū Akoranga Pūtaiao. https://www.sciencelearn.org.nz/images/5156-neural-network-diagram

Neural networks learn by adjusting the weights of connections based on the input they receive and the output they generate. This process, often achieved through *backpropagation*, allows the network to improve its performance and make more accurate predictions or decisions over time. Backpropagation is like a teacher giving feedback on a test. Imagine a student (the neural network) takes a test (processes input data) and gets some answers wrong. The teacher reviews the test, noting where mistakes were made (backpropagation). Then, the teacher guides the student on which parts of the study material (weights in the network) to focus more closely on to improve those specific answers. By repeating this process, the student learns from mistakes and improves over time, just like how a neural network learns to make more accurate predictions.

Neural Network Applications

Neural networks are versatile and can be applied to various tasks, from simple functions like recognizing handwritten digits to complex operations such as interpreting spoken language or analyzing medical images. In business, neural networks are used for customer segmentation, fraud

detection, market trend analysis, and more. They can analyze large volumes of data to uncover insights impractical for a human to find, enhancing decision making and operational efficiencies. For example, in finance, neural networks can analyze market data to identify investment opportunities or risks. In retail, they can help understand customer purchasing patterns to optimize inventory management and personalize marketing efforts.[3] The ability of neural networks to learn from data and detect subtle patterns makes them invaluable for AI applications.

One interesting example of neural networks comes from Microsoft. Microsoft OneNote and Office Lens are innovative tools designed to help users manage notes and documents efficiently.[4] One standout feature powered by a neural network is *optical character recognition (OCR)*. This feature allows handwritten notes to be converted into digital, searchable, and editable text. Recognizing the preference for handwritten note-taking in meetings and lectures, Microsoft sought to bridge the gap between traditional and digital note organization. The neural network implemented for this task mimics the human brain's ability to recognize patterns. It was trained on a vast dataset of handwritten samples, encompassing diverse handwriting styles, languages, and conditions. When a user uploads a photo of handwritten notes via Office Lens or writes directly in OneNote, the neural network analyzes the image to identify shapes and patterns, accurately converting them into digital text. This digital text is then made searchable and editable, enabling users to organize and retrieve notes effortlessly.

We discuss neural networks in more detail in **Chapter 5**.

Generative AI

Generative AI is a form of AI that focuses on creating new content, such as text, images, music, or code, through an interactive process with a human. Generative models, like ChatGPT, can produce text or images that are often indistinguishable from those created by humans. These models are trained on large datasets and can generate creative and realistic content. Most generative AI systems use neural networks to generate their responses.

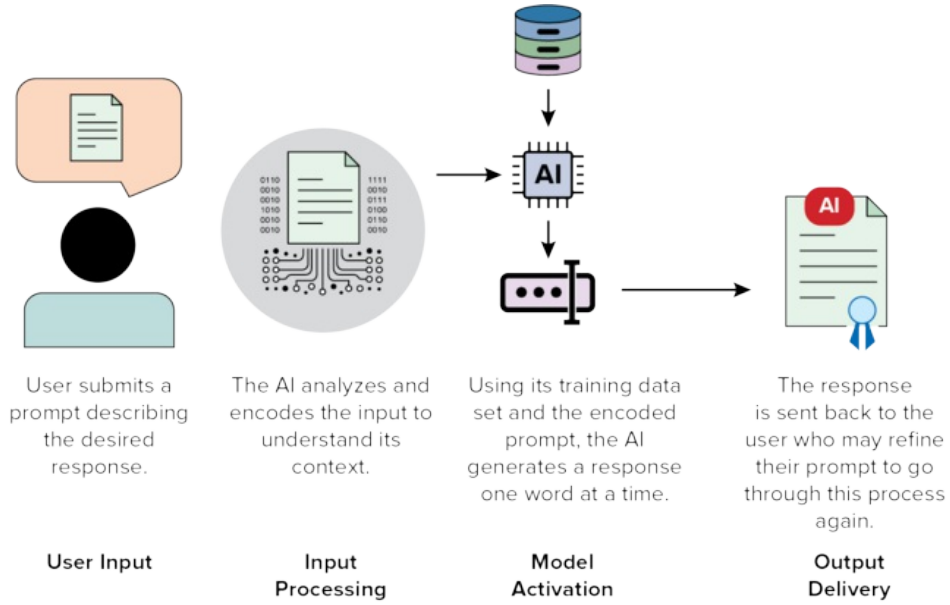
To generate text using generative AI, a human prompts the AI with a description of the desired output or a question. The generative AI then goes through the following process to determine its response (see **Figure 1.3**):

1. ***Input processing.*** The AI processes the input prompt to understand the context and the specific request. This involves breaking the text down into smaller units, like words or phrases, and encoding it into a format that the model can work with.
2. ***Model activation.*** The encoded input is fed into the generative model, which uses a large training database of books, articles, websites, and other written content. This training database helps the AI learn grammar, vocabulary, and context. Based on its prompt interpretation, the AI starts creating a response. It builds the answer one step at a time, like putting together a puzzle. For text generation, this means predicting the next word in the sequence based on the input and the previously generated words. This process continues until the model generates a complete response or reaches a predefined stopping point. The AI may go through this process

multiple times, generating different responses.

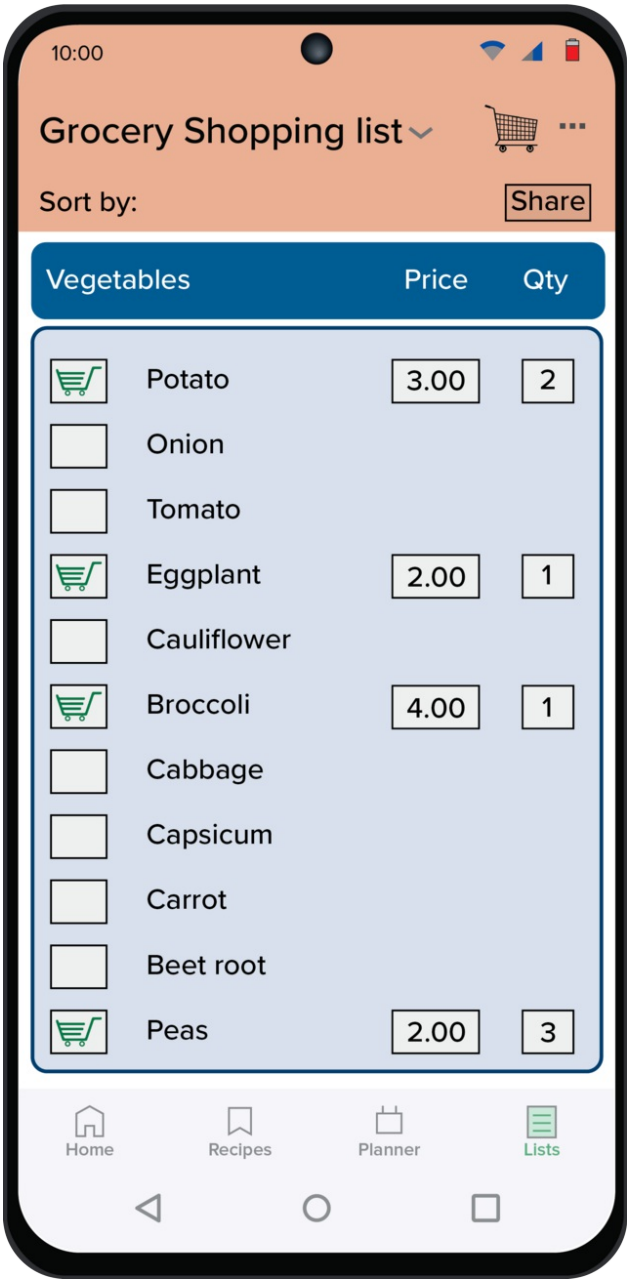
3. ***Output delivery.*** The processed output is delivered to the user. The user can then review the generated content and, if necessary, provide further prompts or adjustments to refine the output.

FIGURE 1.3 Generative AI Process



A Generative-AI Application

One company that uses generative AI creatively is Samsung Food (formerly the Whisk app and now part of Samsung). Samsung Food leverages generative AI to tackle common challenges in meal planning, including reducing food waste, accommodating dietary preferences, and offering recipe variety.[5] When users interact with the app or website, they input details such as available ingredients, dietary restrictions (e.g., gluten-free, vegan), and meal preferences (e.g., low-calorie, quick preparation). The AI processes this input and activates its model in several stages: matching ingredients to recipes, filtering for dietary restrictions, optimizing for user preferences, and ranking the most relevant options. These selected recipes are then presented in a user-friendly format with step-by-step instructions, nutritional information, and substitution suggestions.



Samsung Food's use of generative AI delivers benefits to users, simplifying meal planning, promoting healthier eating, and minimizing food waste.

The generative AI behind Samsung Food is trained on large datasets, including thousands of recipes, nutritional databases, ingredient combinations, and user behavior patterns such as dietary preferences and cooking habits. From a business perspective, the use of generative AI strengthens Samsung's brand loyalty by integrating seamlessly into its smart appliance ecosystem, including devices like Samsung smart fridges. Additionally, it creates new revenue opportunities through partnerships with food retailers and delivery services, illustrating the practical value of

generative AI in enhancing customer engagement and operational efficiency.

We discuss generative AI in more detail in **Chapter 6**.

OTHER APPLICATIONS OF AI IN BUSINESS (LO1-2)

So far in this chapter, we have discussed how different types of AI algorithms are used. Here, we take a look at some other applications of AI in business, including *virtual reality (VR)*, *computer vision*, and *speech recognition*.

The Use of Artificial Intelligence in Virtual Reality

Virtual reality (VR) is a computer-generated, interactive 3D simulation that immerses users in a digital environment using devices like headsets and controllers. Together, AI and VR represent two of the most transformative technologies of the 21st century. Their combination holds significant potential to revolutionize a wide range of industries, including gaming, entertainment, health care, and education.

One of the key uses of AI in VR is in the gaming industry. AI algorithms create more realistic and responsive *non-player characters (NPCs)*. These NPCs can adjust to the player's actions, providing a more immersive experience. For example, AI can enable NPCs to learn from the player's behavior and develop strategies to challenge them, making the game more dynamic.

In the health care sector, AI-powered VR is used for medical training and therapy. Surgeons can practice complex procedures in a virtual environment, allowing them to improve their skills without risking patient safety. Additionally, AI-driven VR therapy is used to treat mental health conditions such as posttraumatic stress disorder (PTSD) and anxiety. By simulating real-life scenarios, patients can confront and manage their fears in a controlled environment.

Education is another area where AI and VR are advancing. Virtual classrooms powered by AI can provide personalized learning experiences for students. AI algorithms can analyze a student's performance and adapt the curriculum to suit their learning pace and style. This individualized approach can increase student engagement and improve learning outcomes.

One company currently taking advantage of the combination of AI and VR is UK-based Bodyswaps. The Bodyswaps AI Roleplay Simulator is comprised of a variety of roleplay simulation templates that training facilitators can customize through a no-code interface that enables them to tailor each scenario to learners. These roleplay simulations can then be published for access, as well as embedded in a wider immersive learning pathway created by each training facilitator. In the simulation, learners engage in free-flowing interactions that are driven by AI, replicating real-world scenarios while enhancing repeatability of practice. Each roleplay simulation also provides in-depth personalized AI-powered feedback on a variety of categories to enhance learners' self-reflection and development.

The integration of AI in VR provides several advantages:[6]

1. It improves the realism and interactivity of virtual environments. AI algorithms can produce realistic simulations that react to the user's actions in real time. This level of interactivity can make VR experiences more immersive.
2. AI can enhance the efficiency and effectiveness of VR applications. In health care, for instance, AI can analyze large amounts of data to offer accurate diagnoses and treatment plans. This can lead to improved patient outcomes and reduced health care costs. In education, AI can identify areas where students face difficulties and provide targeted interventions to support their success.
3. AI can facilitate the creation of more personalized VR experiences. By analyzing user data, AI algorithms can adjust the virtual environment to meet the individual needs and preferences of the user. This can improve user satisfaction and increase the overall effectiveness of the VR application.

As AI algorithms become more sophisticated, they will be capable of creating more realistic and responsive virtual environments. Additionally, the combination of AI and VR could result in the creation of new technologies. For example, AI might be used to develop virtual assistants that interact with users naturally and intuitively. These virtual assistants could be applied in settings such as customer service and personal coaching.

Computer Vision and Its Role in Business Applications

Computer vision is a neural network that enables machines to interpret and make sense of visual information from the world around us, such as images and videos. This technology is not just about seeing; it is about understanding and interpreting visual data to assist decision making.

A prime example of computer vision in action is facial recognition technology, which has become increasingly prevalent in security and authentication systems. In health care, computer vision is revolutionizing diagnostics by enabling the automated reading of x-rays, scans, and tests, thereby assisting doctors in identifying conditions more quickly and accurately. Computer vision also plays a crucial role in quality control in the manufacturing sector. It allows for automatic inspection of products on assembly lines, identifying defects or irregularities that the human eye might miss. This improves the quality of products and enhances efficiency in production processes. Another exciting application is in the automotive industry, where computer vision is integral to developing driver assistance systems. It helps vehicles interpret and respond to the environment, enhancing safety features like obstacle detection and lane departure warnings.

Blue River Technology, a subsidiary of John Deere, utilizes advanced computer vision and AI to optimize crop spraying.[7] Traditional broadcast crop spraying involves spraying chemicals on all plants in a field. Blue River's "See & Spray" system uses cameras linked to computers to identify plants and target herbicide application to weeds only. Trained on more than 1 million images,

Blue River's systems can distinguish crops from weeds with 99.1 percent accuracy. In testing, farms using See & Spray cut herbicide utilization by 77 percent while increasing yields by 3 to 5 percent, delivering a strong return on investment. This precision application of computer vision maximizes farm productivity and profitability while improving environmental sustainability.

Speech Recognition and Its Business Impact

Speech recognition, which uses a neural network, allows machines to understand and respond to human speech. This technology powers various applications, from converting spoken words into text (speech-to-text) to enabling computers to speak (text-to-speech). It is the driving force behind virtual assistants like Siri and Alexa and tools that can transcribe meetings, analyze customer sentiments in call centers, and summarize large documents.

When Apple introduced Siri in 2011 as part of the iPhone 4S, it marked a significant evolution in conversational AI. Siri was among the first to move beyond rigid voice commands to understanding more natural, everyday language. Machine learning made this possible, allowing Siri to learn and adapt to different speech patterns. As a result, iPhone users could perform various tasks such as making phone calls, getting directions, searching for information, and controlling music without lifting a finger.[8]

The success of Siri changed how consumers interact with their devices and sparked a wave of innovation in the tech industry. Companies began to invest more in neural networks, aiming to create even more sophisticated voice assistants. Voice-driven interactions are now an integral part of the customer experience in many industries.

DATA SCIENCE AND AI (LO1-3)

The evolution of AI has been shaped by three key factors: hardware, software, and data. At different points in AI's history, limitations in one or more areas have slowed progress. Early AI systems struggled due to insufficient computing power, while software constraints made it difficult to implement complex algorithms. However, even with advancements in processing power and sophisticated models, AI remains dependent on high-quality data—the fuel for machine learning and decision making. **Data science** is one leg of this three-legged stool, ensuring AI has the structured, clean, and relevant data to function effectively. While hardware improvements provide the computational power to train large-scale models and software innovations optimize AI's efficiency, without high-quality data, AI lacks the information necessary to learn, adapt, and generate meaningful insights. Data science ensures that AI has the structured, clean, and relevant data to function effectively, completing the triad essential for AI progress.

Data science is an interdisciplinary field that combines statistical analysis, computer science, and domain expertise to extract meaningful information from data. *Domain expertise* is the deep knowledge and specialized understanding of a specific industry, field, or subject matter that enables professionals to interpret data, identify relevant patterns, and make informed decisions within that context.

To make raw data useful for AI, data scientists follow a structured process that involves cleaning, organizing, and analyzing information. Data cleaning, visualization, and predictive modeling are three key techniques in this process:

- **Data cleaning** is identifying and correcting errors and inconsistencies in datasets).
- **Visualization** presents data and insights using charts, graphs, and other visual tools to make complex information easier to understand, interpret, and communicate.
- **Predictive modeling** is the process of using historical data and statistical algorithms to forecast future events or trends

AI depends on large, clean, unbiased datasets to mimic human-like learning, problem solving, and decision making. **Table 1.1** lists common data problems, how data science handles those issues, and the implications if the problems are not resolved.

Table 1.1 **Common Data Problems, Solutions, and Risks if Not Fixed**

Common Data Problem	Data Science Solution	What Happens if Not Fixed?
Missing data	Fill in missing values using averages, common values, or AI predictions.	AI models may give incorrect results if important data is missing.
Duplicate records	Remove repeated entries to avoid counting the same information twice.	Extra copies take up space and may cause errors in analysis.
Outliers	Detect and evaluate outliers; decide to retain, transform, or remove based on context.	Predictions may be distorted by extreme or incorrect values.
Imbalanced data	Balance data by adjusting the number of examples in each category.	AI may become biased, favoring one group over another.
Data bias	Check for unfair patterns in data and adjust to ensure fairer AI decisions.	AI could make biased or discriminatory decisions.
Inconsistent data formats	Convert data into a common format to ensure consistency.	Merging different datasets may cause errors.
Noisy data (contains errors or irrelevant information)	Clean the data by removing errors and irrelevant details.	AI may struggle to find accurate patterns.
Data drift (statistical properties of the data changed over time)	Regularly check AI accuracy and update it with new data to keep it relevant.	AI models may become outdated and make poor predictions.
High dimensionality (datasets with a large number of variables)	Simplify the data by removing unnecessary details.	AI may take longer to process and make poor decisions.
Unstructured data	Use AI tools to analyze and organize text, images, or audio.	Difficult to extract useful insights from messy data.

In a business context, combining data science and AI unlocks immense potential by enhancing decision making and automating processes. Data science establishes the data foundation and is critical in monitoring and validating AI's ongoing performance. For instance, AI models trained on clean, comprehensive customer data in marketing can predict future consumer behaviors, allowing businesses to tailor strategies effectively. Similarly, AI can leverage vast market data in finance for predictive analysis, guiding strategic financial planning. In health care, AI-powered diagnostic tools use medical imaging data to detect diseases early, improving patient outcomes. In supply chain management, AI analyzes logistics data to optimize inventory and reduce

inefficiencies. These examples highlight how data science and AI enhance decision making across multiple industries. The synergy between data science and AI empowers businesses to make smarter decisions, optimize operations, and innovate in today's data-driven world.

ETHICAL CONSIDERATIONS IN AI (LO1-4)



The ethical dilemmas posed by AI today echo the profound societal shifts experienced during the Industrial Revolution. Back then, steam-powered machines revolutionized industries, dramatically increasing productivity but also triggering widespread labor displacement, altering power structures, and introducing exploitation risks. These upheavals raised pressing ethical questions about workers' rights, economic inequality, and environmental sustainability.

Similarly, AI is reshaping industries and redefining social and economic dynamics. It offers immense potential to enhance productivity, streamline operations, and unlock new avenues of creativity and innovation. Yet, it also raises significant ethical concerns. For example, a 2023 report by Goldman Sachs indicates that AI could expose about 300 million jobs worldwide to automation by 2030. That is 11 percent of the global workforce.[9] As we navigate this new era of technological disruption, the Industrial Revolution serves as a historical parallel, providing valuable insights to address emerging challenges thoughtfully.

In this section, we introduce some of the major ethical issues surrounding AI. Each topic is accompanied by historical and cultural context, highlighting how and when these concerns entered public consciousness. This analysis aims to equip readers with the tools to approach AI's ethical implications with foresight and responsibility.

We discuss ethical issues related to AI throughout this text and in more detail in **Chapter 8**.

Data Ownership, Privacy, and Consent

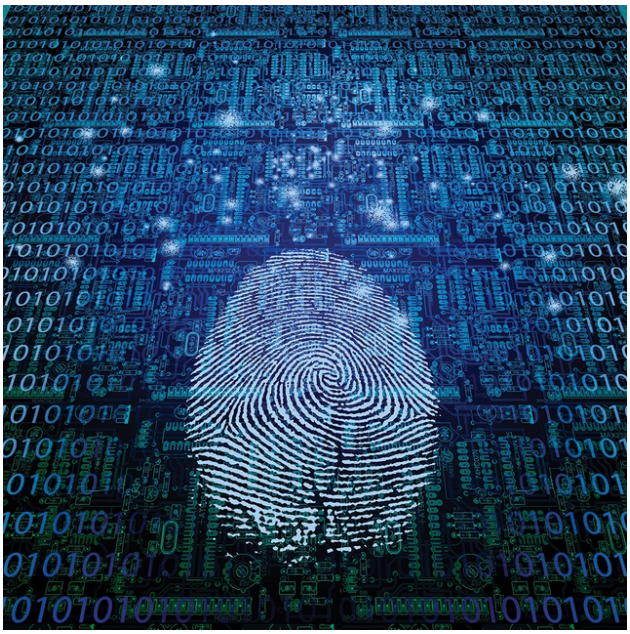


Data ownership is a complex issue due to the intangible nature of data and the challenges in protecting ownership rights. According to the EU General Data Protection Regulation (GDPR), **personal data** includes any information related to an identifiable person, encompassing both objective and subjective information.[10] The debate on data ownership revolves around whether data capturers (those who collect or create data) or data subjects (individuals whose actions generate the data) should be considered the true owners. **Privacy** laws aim to protect individuals' rights to their data, granting rights such as access, erasure, restriction, and portability.

AI technologies can collect and analyze vast amounts of personal data, raising concerns about surveillance and the erosion of privacy. Data is collected on individuals in numerous ways they

[11]

might not be aware of.[11] For instance, smartphones and connected devices continuously gather location data, browsing history, and app usage patterns. Social media platforms track interactions, likes, and shares to build detailed profiles. Websites use cookies to monitor online behavior and preferences. Retailers collect purchase history and even track movements within stores through Wi-Fi signals. Financial institutions analyze transaction data to assess creditworthiness and detect fraud. Public surveillance systems and smart city technologies capture video and other data in public spaces. Additionally, data brokers aggregate and sell personal information from various sources.



Bruce Rolff/Shutterstock

The debate over who owns all of the data we generate is ongoing.

These pervasive data collection practices, often conducted without explicit consent or full transparency, amplify ethical concerns in AI applications. One key issue is **informed consent**—individuals often lack clarity about what data is being collected, how it will be used, or who will have access to it. This lack of awareness undermines the foundational principle of consent in data privacy frameworks like the GDPR.

Moreover, AI-driven systems can exacerbate privacy risks by combining disparate data sources to infer sensitive information, such as health conditions, political affiliations, or personal habits, even when individuals have not explicitly shared such details. This phenomenon, often called *data triangulation*, raises the stakes for protecting privacy, as individuals may not even be aware of the extent to which they are profiled.

Finally, data ownership and equity remain unresolved issues. While privacy laws like the GDPR grant individuals some rights over their data, they often lack enforcement mechanisms to ensure equitable outcomes. For example, individuals may have the right to request data erasure, but the burden of knowing, understanding, and exercising this right often falls on them, creating a power imbalance between individuals and organizations.

Environmental Sustainability

The environmental impact of AI is increasingly drawing attention, with concerns centering on energy consumption, resource depletion, and waste generation. Data centers, which form the backbone of AI operations, account for 1 to 2 percent of global electricity use, amounting to over 200 terawatt-hours annually, according to the International Energy Agency.[12] This energy consumption often relies on fossil fuels, contributing to greenhouse gas emissions and exacerbating climate change. Furthermore, cooling systems for these data centers require vast quantities of water—Google alone reported using over 6.4 billion gallons of water in 2023 to maintain its global operations.[13]

The environmental toll extends to hardware production, as manufacturing the processors essential for AI systems depends on rare earth minerals like lithium, cobalt, and nickel. Extracting these materials has consequences, including deforestation, water contamination, and labor exploitation in mining regions like the Democratic Republic of the Congo. Unfortunately, the rapid pace of AI innovation renders much of this hardware obsolete within a few years, contributing to the growing problem of electronic waste (e-waste). A study by researchers at United Nations University estimates that 53.6 million metric tons of e-waste were generated globally in 2019, with less than 20 percent properly recycled.[14] There is an urgent need for more sustainable practices in AI development, from improving energy efficiency and adopting renewable power sources to enhancing recycling efforts and reducing resource dependency.



Narat Plaeksakul/Alamy Stock Photo

AI innovation is contributing to the growing problem of electronic e-waste.

Economic Inequality and AI Access

The **AI divide** represents the disparities in the adoption and benefits of AI technologies across different regions and socioeconomic groups. These gaps reflect broader inequalities in digital infrastructure, education, and economic opportunity. Left unaddressed, this divide threatens to deepen existing inequalities, limiting access to AI's transformative potential for disadvantaged communities and nations. The divide operates on three levels: globally, regionally, and individually:

- **Global Divide**. Globally, industrialized nations dominate AI development and adoption, thanks to their robust digital infrastructure, advanced research ecosystems, and significant investments in technology. China and the United States, for example, accounted for over 80 percent of global AI funding in 2021.[15] However, 37 percent of the global population—primarily in low-income regions—remains offline, unable to participate in the digital economy.[16] This digital exclusion hinders developing nations' ability to leverage AI technologies, perpetuating economic disparities and limiting their participation in the AI-driven global economy.
- **Regional Divide.** Within nations, disparities between metropolitan and rural areas highlight a regional AI divide. Urban centers benefit from investments in digital infrastructure, education, and job opportunities, often becoming hubs for AI innovation and employment. In contrast, rural areas frequently suffer from inadequate Internet access and insufficient technical education, leaving them excluded from the AI revolution. For instance, in the United States, 17 percent of rural Americans lack access to high-speed Internet, compared to only 1.5 percent of urban residents.[17] In India, urban areas have 93 percent Internet penetration, whereas rural regions lag behind at 37 percent.[18] This gap exacerbates economic inequalities, as rural workers are disproportionately affected by job displacement from AI-driven automation.
- **Individual Divide**. On an individual level, the AI divide reflects and reinforces existing socioeconomic gaps. Wealthier individuals and communities have greater access to AI-powered tools, education, and job opportunities, while underprivileged groups face systemic barriers. High-paying AI jobs typically require advanced training and education, often inaccessible to marginalized groups due to high costs or geographic limitations. In the U.S., only 13 percent of Black workers and 16 percent of Hispanic workers are employed in tech jobs, compared to 47 percent of white workers.[19] Additionally, automation disproportionately threatens low-skill jobs, often held by economically vulnerable populations, while resources for reskilling are concentrated in wealthier communities.

Efforts to address the global divide include investments in technological infrastructure and programs aimed at empowering underserved communities. For example, Google's Internet Saathi initiative has trained over 30 million rural women in India to use digital tools, increasing their participation in the digital economy.[20] Additionally, international organizations and development banks are funding broadband infrastructure projects in developing regions to improve access.

In developed nations like the United States, bridging the regional divide has been a focus of government-led programs. Initiatives such as the Rural Digital Opportunity Fund (2020) and the Tribal Broadband Connectivity Program (2021) have expanded Internet access in rural areas. For example, only 24 percent of rural Americans had high-speed Internet in 2006, compared to 83 percent in 2024.[21]

At the individual level, targeted programs aim to build skills and foster inclusion for those most affected by the AI divide. For instance, Amazon's Upskilling 2025 initiative has committed over $700 million to retrain 100,000 employees in tech-related fields, equipping workers with skills for emerging roles in cloud computing, AI, and cybersecurity.[22] Google's Grow with Google initiative[23] and Microsoft's Global Skills Initiative[24] provide free or low-cost training in AI and digital skills, targeting underprivileged populations.

While significant strides have been made to address the global, regional, and individual levels of the AI divide, challenges remain. Bridging these divides will require sustained collaboration between governments, private companies, and international organizations.

Bias and Discrimination



AI systems can perpetuate and amplify biases in their training data, leading to unfair treatment of individuals based on race, gender, or other characteristics. The 2020 documentary *Coded Bias*, directed by Shalini Kantayya, analyzed how the training data set used in facial recognition AI was heavily weighted with lighter-skinned male faces, creating a systemic bias against darker-skinned people and women. (This is explored in more depth in a case study at the end of this chapter.)

The book *Weapons of Math Destruction* by Cathy O'Neil also delves into how algorithms, when trained on biased or incomplete data, can perpetuate inequality.[25] AI-driven decision-making systems in areas like hiring, lending, and law enforcement often inherit biases in historical data and continue to generate biased outcomes that disproportionately affect marginalized communities. For example, predictive policing algorithms trained on past arrest data may over-target specific neighborhoods, creating a feedback loop that perpetuates systemic biases. Similarly, AI-powered credit scoring models may deny loans to applicants from underserved communities due to historical financial disparities, further limiting economic opportunities.

These cases underscore the critical role of data science to remove past biases so AI models can be effective and ethical. Businesses must carefully monitor, audit, and refine AI systems to mitigate unintended harms and promote fair decision-making.

Manipulation and Deception



AI algorithms can manipulate user behavior on social media, creating filter bubbles and amplifying **misinformation** and deception. A **filter bubble** is a state of intellectual isolation where individuals are only exposed to online content that aligns with their existing beliefs, limiting their worldview. Personalized searches, recommendation systems, and algorithmic curation reinforce these bubbles by prioritizing engagement over balanced information. As users continue interacting with similar content, they are steered further into ideological echo chambers, making it difficult to encounter diverse perspectives.

Beyond manipulation, AI also plays a role in digital deception. AI-generated **deepfakes**, which use machine learning to create realistic yet entirely fabricated videos, can spread false narratives and erode trust in media. AI-powered **chatbots** and synthetic voices can impersonate real individuals, deceiving users into believing false information. Misinformation, whether through fake news, manipulated images, or AI-generated videos, can influence public opinion, undermine democratic processes, and even incite real-world consequences.

The documentary *The Social Dilemma* (2020)[26] explored these concerns by examining how social media algorithms and data collection impact privacy, autonomy, and mental health. Directed by Jeff Orlowski, the film features former employees and executives from major tech companies like Google, Facebook, and Twitter who reveal how these platforms are engineered to maximize user engagement, often at the expense of truth and well-being. AI-powered recommendation systems lead users down rabbit holes of extreme content, reinforcing biases and distorting perceptions of reality. Platforms like YouTube, TikTok, and Facebook continuously track user behavior, using predictive models to keep users scrolling, clicking, and engaging—often without their awareness of the time spent or how their experiences are being shaped.

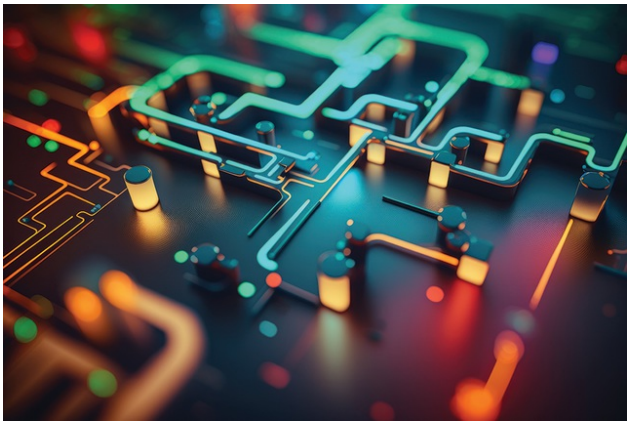
Intellectual Property and AI Creativity



Generative AI is demonstrating continuous improvement in its capability to create video, art, music, literature, and other content. Ethical debates regarding **intellectual property (IP)** and copyright have accompanied this growth. A central issue lies in training these AI systems involving vast datasets that can contain copyrighted works collected without explicit consent from the original creators. These datasets fuel the generative process, allowing AI to mimic styles, themes, and techniques, so the resulting outputs are sometimes confused with human-created works.

Generative AI in this capacity raises a multitude of ethical and legal questions:

- Is it ethical to use copyrighted material without the creator's explicit permission to train AI models? Does using copyrighted material to train AI models constitute fair use, or does it violate copyright laws? What about material that is posted on the web without filing for copyright, like your vacation photos? How can we exercise control over what is used to train an AI? Should AI developers be legally required to disclose the sources of training datasets?
- Should creators whose works contribute to AI training datasets be compensated? If so, how should this compensation be calculated?
- Who deserves credit for AI-generated works—the developers, the AI, or the creators whose works influenced the AI? Should AI-generated works be labeled as such to ensure transparency for consumers and audiences? Can AI-generated content itself be copyrighted, and if not, how should it be protected or regulated?
- Does the widespread use of AI in creative industries risk devaluing human creativity and originality? Could the ease of generating content with AI discourage the development of individual talent and artistic growth?
- Does AI's reliance on existing datasets reinforce cultural biases or homogenize artistic expression? How can we ensure that AI-generated content respects and preserves cultural heritage?
- If AI-generated content closely mimics an existing artist's style, does this constitute copyright infringement or unethical imitation? What legal recourse do creators have if their unique artistic styles are replicated without consent?
- Should new copyright laws or frameworks be explicitly established for AI-generated content? How can international regulatory bodies harmonize laws to address the global nature of AI content creation?



tonktiti/123RF

Who should be credited for photos created by AI like this one?

The legal and ethical questions posed here are not merely theoretical—they are playing out in courts and public debates worldwide.[27] Getty Images sued Stability AI, claiming that the company used its copyrighted images to train its text-to-image generator. Sarah Silver and other authors filed a lawsuit against OpenAI and Meta, asserting that their books were used

without their consent to train AI models. Many other lawsuits have been filed against AI companies by artists, musicians, and authors who allege that their copyrighted works have been used without consent to train AI systems. The Authors Guild and Visual Artists Rights Coalition have called for stricter regulations to address the unauthorized use of creative works in training datasets. These legal battles are shaping the emerging framework for how intellectual property laws apply to AI, yet they also underscore the urgent need for more transparent standards and policies.

Addressing the ethical and legal challenges posed by generative AI requires a proactive and collaborative approach. Policymakers, creators, and AI developers must work together to establish clear guidelines and equitable frameworks that balance innovation with protecting intellectual property and cultural heritage. Transparency in training practices, fair compensation models, and the development of new legal standards will be crucial to fostering trust and inclusivity in the creative industries. These issues need to be addressed to ensure that generative AI complements human creativity rather than undermines it.

Transparency and Accountability



Transparency and accountability are among AI development and deployment's most pressing ethical concerns. When an AI system makes a mistake like misdiagnosing a patient, denying a loan, or causing an accident, determining who is responsible becomes complex. Should the blame rest on the developer who created the system, the organization that deployed it, or the user who operated it?

In 2018, a self-driving Uber vehicle in Tempe, Arizona, struck and killed a pedestrian walking her bike across the road. Investigators discovered flaws in the AI system, which failed to correctly classify the pedestrian or trigger emergency braking.[28] While Uber avoided criminal charges and settled out of court, the test operator behind the wheel was charged with negligent homicide, a felony. She later pleaded guilty to endangerment, a misdemeanor. It illustrates the difficulty of assigning responsibility in such cases. Was the operator at fault for failing to intervene, or should greater blame lie with Uber for deploying an AI system that was not sufficiently safe?



metamorworks/Shutterstock

Who should be responsible when self-driving cars cause accidents or pedestrian deaths or injuries?

There is a lack of transparency in many AI systems, which often function as “black boxes.” Their decision-making processes are so complex that even their developers may not fully understand how specific outcomes are generated. This opacity makes it difficult to assign responsibility and erodes public trust in AI technologies.

To address these issues, experts advocate for clearer regulations requiring **explainable AI (XAI)**, systems that can provide understandable justifications for their decisions. Shared accountability frameworks are also essential, involving all stakeholders (e.g., developers, operators, and regulators) in AI systems. By ensuring that AI systems are designed, deployed, and monitored responsibly, society can better address errors, mitigate harm, and build trust.

Feature: AI in Warfare

AI is increasingly being utilized in warfare, with the weaponization of AI often described as a new arms race. Just as businesses have found numerous applications for AI, so have militaries, employing it for tasks such as intelligence and surveillance; target identification; automated weapon systems; logistics and supply chain management; command and control systems; autonomous submarines, drones, and other vehicles; electronic warfare; tactical decision making; strategic defense systems; psychological and information warfare; and medical support. There is currently no comprehensive international legal framework to regulate the use of automated weapons systems (AWS) or AI in international conflicts. The use of AI in warfare, along with its ethical and moral implications, is a complex and expansive subject currently under intense debate.

Just as the Industrial Revolution brought change to all levels of society and gave them moral and

ethical questions to resolve, so must we grapple with the challenges of AI. It is crucial to recognize that these issues are not merely theoretical, they have tangible impacts on individuals, communities, and societies. By drawing lessons from history and addressing the complexities of AI with transparency, fairness, and inclusivity, we can shape its development and deployment to serve humanity's broader goals. The questions raised in this section are not simply obstacles but opportunities to engage in meaningful dialogue and create ethical frameworks that balance innovation with responsibility.

SUMMARY OF KEY POINTS AND QUICK WINS

1. **Types of AI algorithms.** AI enables businesses to automate tasks, analyze large datasets, and improve decision making by simulating human intelligence. AI algorithms can be broadly categorized as machine learning, neural networks, and generative models, each with unique business applications. Machine learning is a subset of AI that focuses on developing algorithms that allow computers to learn from and make predictions based on data. Neural networks are a series of algorithms designed to recognize patterns and interpret data by mimicking how the human brain operates. Generative AI is a form of AI that focuses on creating new content, such as text, images, music, or code, through an interactive process with a human.
2. **Applications across industries.** AI's versatility spans industries such as finance, health care, and retail, enhancing customer experience, efficiency, and strategic planning. Other applications of AI in business include virtual reality (VR), computer vision, and speech recognition.
3. **Role of data science in AI effectiveness.** High-quality, unbiased data is essential to train accurate and fair AI models, directly affecting performance and reliability. Data science is an interdisciplinary field that combines statistical analysis, computer science, and domain expertise to extract meaningful information from structured and unstructured data. Data science establishes the data foundation and is critical in monitoring and validating AI's ongoing performance.
4. **Ethical issues surrounding AI.** AI is a disruptive technology, and its adoption brings a wide range of ethical issues, including privacy, data ownership, and consent; environmental sustainability; economic inequality and AI access; bias and discrimination; manipulation and deception; intellectual property and AI creativity; and transparency and accountability. It is crucial to recognize that these issues are not merely theoretical, they have tangible impacts on individuals, communities, and societies.

QUICK WINS

- **Understand types of AI and their applications.** Explore the distinctions between machine learning, neural networks, and generative AI. Identify their unique strengths and determine which type aligns with your specific personal or business challenges.
- **Leverage AI for business innovation.** Analyze examples of innovative AI applications, such as Airbnb's anti-party algorithm and assess how similar approaches could solve challenges in your business. Use AI to drive innovation in your organization, whether through enhanced decision making, automation, or creating new customer experiences.
- **Emphasize data quality.** Recognize the importance of high-quality, unbiased data in building

effective AI models. Prioritize data cleaning and curation to ensure reliable AI outcomes.

- **Address ethical considerations in AI.** Assess potential ethical issues, such as bias, privacy concerns, and job displacement, when implementing AI. Develop policies to mitigate these risks. Align AI initiatives with your organization's goals, ensuring that AI tools are purposefully deployed to address specific needs or improve key processes.

KEY TERMS

AI divide. The gap between individuals, organizations, and countries in access to and benefits from AI technologies.

Artificial intelligence (AI). The simulation of human intelligence in machines that can perform tasks such as learning, reasoning, problem-solving, and decision making.

Chatbot. An AI-powered program that interacts with users via text or speech, commonly used in customer service and virtual assistants.

Computer vision. An AI technology that enables machines to interpret and analyze visual information from images and videos.

Data cleaning. Identifying and correcting errors, inconsistencies, and inaccuracies in datasets to improve data quality.

Data science. An interdisciplinary field that combines statistics, machine learning, and domain expertise to extract insights from structured and unstructured data.

Deepfake. AI-generated or manipulated media, such as videos or images, designed to impersonate real individuals or events convincingly.

Explainable AI (XAI). AI systems designed to explain their decision-making processes clearly, improving trust and accountability.

Filter bubble. A situation where users are only exposed to information that aligns with their beliefs due to algorithmic personalization, limiting diverse perspectives.

Generative AI. A type of AI that creates new content, such as text, images, or music, based on patterns learned from training data.

Global divide. Industrialized nations have more resources to develop and implement AI, while developing countries struggle with limited digital infrastructure and expertise.

Individual divide. Socioeconomic disparities affect who can learn, use, and benefit from AI, with wealthier individuals and businesses having more opportunities to leverage AI advancements.

Informed consent. The principle that individuals must be aware of and agree to the collecting and using their personal data.

Intellectual property (IP). Legal rights that protect creations of the mind, such as patents, trademarks, and copyrights, which AI-generated content raises questions about.

Machine learning. A subset of AI where algorithms learn from data to make predictions or decisions without being explicitly programmed.

Misinformation. False or misleading information that spreads online, often amplified by AI-powered recommendation algorithms.

Neural networks. AI models inspired by the human brain, consisting of interconnected layers

of artificial neurons for tasks like image and speech recognition.

Personal data. Information about an individual, such as name, location, or online behavior, often collected by AI-driven systems.

Predictive modeling. Using statistical and machine learning techniques to predict future outcomes based on historical data.

Privacy. The right of individuals to control how their data is collected, used, and shared.

Regional divide. Urban areas often have greater access to AI-driven tools and education, whereas rural areas may lack the necessary infrastructure and opportunities.

Speech recognition. AI's ability to convert spoken language into text, enabling voice assistants and automated transcription services.

Virtual reality (VR). A computer-generated simulation that immerses users in a 3D environment, often enhanced with AI for realistic interactions.

Visualization. The graphical representation of data or information to make patterns, trends, and insights easier to understand. Common forms include charts, graphs, maps, and dashboards.

DISCUSSION QUESTIONS

1. **AI types and applications.** How do different types of AI, like machine learning, neural network, and generative AI, contribute uniquely to business solutions? Provide examples.
2. **Data science's role in AI.** Discuss the role of data science in developing and implementing AI in business. How does it affect AI's effectiveness?
3. **AI's impact on future business trends.** How do you foresee AI influencing future business strategies and operations?
4. **Ethical considerations.** What ethical considerations should businesses keep in mind when implementing AI? Key issues to consider include privacy, data ownership, and bias.
5. **Training AI.** How can businesses make sure their AI systems are trained on appropriate data to avoid biases and ensure fairness?

EXERCISES

1. **Debate.** Conduct a debate on the statement: “AI will replace more jobs than it creates in the next decade.”
2. **Research and presentation.** Research an emerging trend in AI for business and present how it could transform a specific industry.
3. **Case study analysis.** Examine a real-world case study where AI in some form was implemented into a business. Discuss the challenges faced and the solutions provided by AI. Also, include the outcomes.
4. **Ethical dilemma.** Describe a scenario where a business faces an ethical dilemma related to implementing AI. Have students discuss ethical concerns and propose solutions to address them.
5. **AI project proposal.** Have students work in groups to develop a proposal for an AI project that could benefit a business. The proposal should outline the type of AI to be used, the problem to solve, and the expected outcomes.

CASE STUDY 1.1: *THE PORTRAIT OF EDMOND DE BELAMY*—INTELLECTUAL PROPERTY AND AI CREATIVITY



Timothy A. Clary/AFP/Getty Images
The Portrait of Edmond de Belamy

The Portrait of Edmond de Belamy, created in 2018 by the Paris-based art collective Obvious, is one of the earliest and most prominent AI-generated artworks to capture the art world's attention. It highlights the intersection of technology and creativity, demonstrating how AI can mimic or reinterpret human artistic endeavors.

Obvious trained an AI model on a dataset of 15,000 classic portraits from the 14th to the 20th centuries. A generative adversarial network (GAN) used to create this artwork operates through a "dueling" process where one neural network generates images while the other evaluates their realism. This iterative method produced *Edmond de Belamy*, which was chosen, printed on canvas, and framed as a traditional painting. The artwork was sold at Christie's auction house in New York for an unexpected $432,500, vastly surpassing its initial valuation of $7,000–$10,000.

U.S. and French Copyright Laws

As the painting was created in France and sold in the United States, both countries' copyright laws are relevant. In the United States, copyright law requires human authorship for a work to qualify for protection, as established by the U.S. Copyright Office's rejection of purely AI-

generated creations. If AI is used as a tool and humans exert significant creative control over the final work, humans may claim copyright. Additionally, U.S. law raises concerns about using copyrighted material in training datasets, particularly if the resulting work is considered derivative.

French copyright law (*droit d'auteur*) strongly emphasizes the personal connection between the creator and the work, requiring the author's personality to be reflected in the creation. Like the United States, France does not recognize AI as an independent creator. However, French law grants moral rights, including attribution and integrity, to human authors, which could complicate cases involving AI-trained datasets.

In both jurisdictions, the unauthorized use of copyrighted works for AI training raises legal and ethical questions, especially regarding fair use (United States) or infringement (France). While neither system currently grants copyright to AI-generated works, both grapple with adapting their laws to address these emerging challenges.

This raises fundamental questions about what constitutes creativity. Can a machine be truly creative, or is it merely following programmed algorithms? The GAN relied on a dataset of human-created works, leading to a deeper ethical dilemma: Is the machine's "creativity" a derivative process, and where does ownership lie?

Discussion Questions

1. Who is the rightful "creator" of *Edmond de Belamy*—the AI that generated the piece, the Obvious collective who designed the project, or perhaps even the individuals who created the historical portraits used to train the GAN?
2. Does AI-generated art hold the same artistic value as human-created art? Why or why not? How do you think the high price paid at auction impacts the perceived worth of AI-generated art? Does the commercial success of *Edmond de Belamy* justify the use of AI in art creation, or does it raise more ethical concerns?
3. The *Portrait of Edmond de Belamy* raises questions about AI's impact on creative industries. Should there be limits on AI-generated art in markets where traditionally human-made art is sold? Why or why not? If so, what would those limits be? If AI-generated pieces gain popularity and commercial success, how might this shift affect human artists and the value of human-created art?
5. Given that Obvious trained the GAN on publicly available works without direct permission from the artists or owners of those works, is this fair use, or should compensation be considered for the creators whose work was part of the dataset? Explain your response.

Case References

Christie's. "Obvious and the Interface between Art and Artificial Intelligence."

https://www.christies.com/lot/lot-edmond-de-belamy-from-la-famille-de-6166184

Epstein, Z., Levine, S., Rand, D. G., & Rahwan, I. (2020). “Who Gets Credit for AI-Generated Art?” *iScience* 23(9), 101515. **https://doi.org/10.1016/j.isci.2020.101515**

CASE STUDY 1.2: CODED BIAS



Joy Buolamwini, a computer scientist and founder of the Algorithmic Justice League, conducted research revealing significant biases in facial analysis software, particularly its accuracy across different demographic groups. Her study, documented in the film *Coded Bias*, highlights the following key findings:

1. **Racial and Gender Bias:** Buolamwini found that many commercial facial analysis software systems had higher error rates when analyzing the faces of individuals with darker skin tones, especially women. The error rates were significantly lower when assessing the faces of individuals with lighter skin tones, particularly men.
2. **Underrepresentation in Training Data:** The biases in these systems were attributed to the lack of diverse training data. Many facial analysis algorithms had been primarily trained on datasets consisting mainly of lighter-skinned individuals. This lack of representation led to poorer performance when assessing faces with different skin tones.
3. **Accuracy Disparities:** The accuracy disparities are particularly concerning because they could lead to real-world consequences. For example, some of the software showed higher error rates in classifying gender for individuals with darker skin tones, which could lead to misidentifications, discriminatory practices, or underrepresentation.
4. **Impact on Marginalized Groups:** Buolamwini's research highlighted that these biases disproportionately affected marginalized and underrepresented communities, exacerbating issues related to racial and gender inequality.

Buolamwini's work not only raised awareness about the biases in facial analysis software but also led to important discussions about fairness, transparency, and accountability in AI and machine learning. Her efforts have contributed to increased scrutiny and efforts to improve the fairness of AI technologies, and she continues to be an advocate for algorithmic justice and inclusive AI development.

Discussion Questions

1. How do Buolamwini's findings on biases in facial analysis software impact the ethical use of AI?
2. Discuss the importance of diverse training datasets in AI development. How can companies ensure their AI systems are trained on diverse data?
3. What could be the real-world consequences of the accuracy disparities found in facial

recognition software?

4. What responsibility do tech companies have in ensuring their AI products do not perpetuate biases?

Case References

MIT News. “Study Finds Gender and Skin-Type Bias in Artificial Intelligence Systems.” February 12, 2018. **https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212**.

Mozilla Foundation. “Facial Recognition Technology: The Problem with Bias.” Accessed February 2, 2025. **https://foundation.mozilla.org/en/blog/facial-recognition-bias/**.

Case Study 1.3: Closing the Loop—How H&M and HKRITA Use AI to Revolutionize Textile Recycling

The fashion industry produces over 90 million tons of textile waste each year. Globally, less than 1 percent of discarded clothing is recycled into new garments. Sorting used clothes for recycling is labor-intensive and error-prone, creating a bottleneck in the push toward *circular fashion*. In this model, clothing is designed, produced, and managed in a way that keeps materials in use for as long as possible. Instead of the traditional "take-make-dispose" approach, circular fashion emphasizes reuse, repair, and recycling to minimize waste and environmental impact. For true fiber-to-fiber recycling to be viable at scale, garments must be accurately identified by material type and fabric structure. Manual sorting methods are not up to the task.

Enter the AI-powered Smart Garment Sorting System, a joint innovation between the H&M Foundation and the Hong Kong Research Institute of Textiles and Apparel (HKRITA). Designed as part of H&M Foundation's *Planet First* initiative, this system integrates robot arms, artificial intelligence, and special cameras to automate garment sorting with remarkable precision. Trained on a database of over 150,000 garment samples, the system classifies discarded clothing by material (e.g., cotton, polyester, blends), structure (woven, knit), garment type (e.g., shirts, jeans), and color. The garments processed by the system are primarily post-consumer garments—clothing worn and then donated or returned by consumers through recycling programs or charitable organizations.

With a reported sorting accuracy of 96 percent and a processing time of 8 seconds per item, the system outperforms human sorters and dramatically boosts throughput—critical for scaling recycling infrastructure to meet global sustainability targets.

What Happens after Sorting?

Once garments are sorted by material and structure, they are routed to appropriate recycling streams. Some are processed through fiber-to-fiber recycling, where materials like 100 percent cotton or polyester are broken down and spun into new yarns. Others go through mechanical recycling, where fabrics are shredded into materials for insulation or industrial use. For blended fabrics, chemical recycling technologies can separate fibers like polyester and cotton for reuse. Clothing that is clean and wearable is often sorted for resale earlier in the collection process; the garments that reach this system are typically those that cannot be resold and would otherwise end up in landfills or incinerators.

Insights

The Smart Garment Sorting System illustrates how artificial intelligence is reshaping digital domains and bringing innovation to the physical world in ways that matter for business and the planet.

AI's Growing Presence in Physical, Traditionally Manual Industries

While AI has often been associated with software, data analytics, and virtual environments, this case exemplifies its increasing impact on physical processes. Garment sorting has historically been a manual, tactile task requiring human judgment. By applying computer vision and hyperspectral analysis, the Smart Garment Sorting System demonstrates that AI can now perform complex physical classification tasks with high accuracy—offering efficiency gains and the potential to reimagine labor in manufacturing and recycling sectors.

The Potential of AI for Sustainability and Circular Economies

AI is emerging as a crucial enabler of sustainable practices, particularly within the framework of a circular economy. In this case, AI directly makes textile-to-textile recycling scalable and accurate, which has been a long-standing challenge. By reducing dependence on virgin fibers and enabling more garments to reenter the supply chain, AI supports sustainability goals beyond incremental efficiency improvements—it helps redesign how materials flow through the economy.

The Importance of Data-Rich Environments

The sorting system's success depends not just on algorithms but also on the quality and volume of data. The system was trained on over 150,000 samples, and its accuracy improves as the database grows.

AI in Long-Term Environmental Goals

Traditionally, AI has been viewed as a tool for boosting productivity or reducing costs. However, organizations are increasingly aligning AI investments with long-term strategic goals, including climate resilience and sustainable growth. Rather than merely using AI to optimize internal operations, H&M Foundation is deploying it to transform how the fashion industry manages waste. This shift reflects a growing understanding that AI can—and should—be used to solve the

systemic challenges that define the future of business.

The project is still in its early stages, with further testing, refinement, and global deployment on the horizon. It shows how machine learning and computer vision can be applied to traditional sectors like textile recycling and advance sustainability and social responsibility using AI.

Discussion Questions

1. **Types of AI.** What AI technologies are involved in the Smart Garment Sorting System? How do machine learning and computer vision work together in this case?
2. **AI in physical industries.** How does this case challenge the assumption that AI is only valid in digital or information-based industries?
3. **Data science and model training.** Why is a large, well-labeled dataset critical in the success of this system? What role does data science play here?
4. **Ethics and sustainability.** How does this case illustrate ethical and environmental considerations in AI adoption? What responsibilities do companies have when using AI for sustainability?
5. **Broader impacts.** What lessons can other industries learn about implementing AI to address large-scale challenges from this case?

Case References

Ellen MacArthur Foundation. “Circular Fashion.” **https://ellenmacarthurfoundation.org/topics/fashion/overview**.

H&M Foundation. “AI-Powered Smart Garment Sorting System: Automating Textile Recycling.” **https://hmfoundation.com/project/ai-powered-smart-garment-sorting-system-automating-textile-recycling/**.

HKRITA. “AI-Powered Smart Garment Sorting System.” *The Hong Kong Research Institute of Textiles and Apparel*. **https://www.hkrita.com/en/project/ai-powered-smart-garment-sorting-system**.

Endnotes

[1] **https://news.airbnb.com/anti-party-crackdown-over-new-years-eve/**

[2] **https://news.airbnb.com/anti-party-crackdown-over-new-years-eve/**

[3] *Mohmad, Parvin. "Top 10 Business Applications of Neural Networks in 2023." Analytics Insight, 24 May 2023.* **https://www.analyticsinsight.net/business-intelligence-and-analytics/top-10-business-applications-of-neural-networks-in-2023**; *Gupta, Jatinder N. D., and Kate A. Smith. Neural Networks in Business: Techniques and Applications. Idea Group Inc (IGI), 2003.*

[4] *"Microsoft Lens gets a hot new feature – the Office app can 'read' your handwriting." Laptop Mag, 2024.* ***https://www.laptopmag.com/news/microsoft-lens-gets-a-hot-new-feature-the-office-app-can-read-your-handwriting***.

[5] **https://news.samsung.com/global/samsung-announces-global-launch-of-samsung-food-an-ai-powered-personalized-food-and-recipe-service**

[6] Gratch, J., Wang, N., Gerten, B., Strader, M., Nordquist, D., & Larson, K. (2004). *Creating Rapport in Virtual Humans*. IEEE Intelligent Systems, 19(4), 42–50; Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., ... & Alcañiz, M. (2007). *Affective Interactions Using Virtual Reality: The Link Between Presence and Emotions*. CyberPsychology & Behavior, 10(1), 45–56.

[7] **https://bluerivertechnology.com/technology/**

[8] https://hbr.org/2018/02/7-years-later-what-really-happened-to-siri-and-apple; **https://www.fastcompany.com/3065339/siri-rising-inside-apples-voice-controlled-brain-trust**

[9] **https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent.html**

[10] European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).* Official Journal of the European Union, L 119, 1–88.

[11] **https://www.bcs.org/media/10764/data-ownership-and-privacy.pdf**

[12] International Energy Agency (IEA). *Data Centres and Data Transmission Networks – Tracking Report*. 2021. **https://www.iea.org/reports/data-centres-and-data-transmission-networks**

[13] Google. *Environmental Report 2024*. Google Sustainability. 2024. **https://www.gstatic.com/gumdrop/sustainability/google-2024-environmental-report.pdf**

[14] Forti, V., Baldé, C. P., Kuehr, R., and Bel, G. *The Global E-Waste Monitor 2020: Quantities,*

Flows, and the Circular Economy Potential. United Nations University, International Telecommunication Union, and International Solid Waste Association. 2020. **https://collections.unu.edu/eserv/UNU:7737/GEM_2020_def_july1.pdf**

[15] Stanford AI Index Report (2022): **https://aiindex.stanford.edu/report/**

[16] International Telecommunication Union, 2024: **https://www.itu.int/itu-d/reports/statistics/facts-figures-2024/index/**

[17] Federal Communications Commission (FCC), 2021 Broadband Deployment Report: **https://docs.fcc.gov/public/attachments/FCC-21-18A1.pdf**

[18] Telecom Regulatory Authority of India (TRAI), 2022: **https://trai.gov.in/**

[19] Pew Research Center, 2021: **https://www.pewresearch.org/**

[20] **https://about.google/intl/ALL_us/values-in-action/internet-saathi/**

[21] Horrigan, John B., and Katherine Murray. "Rural America Trails Rest of Nation in Home Broadband Use, but Gap Has Narrowed in Past Two Years." *Pew Research Center*, February 26, 2006 **https://www.pewresearch.org/internet/2006/02/26/rural-america-trails-rest-of-nation-in-home-broadband-use-but-gap-has-narrowed-in-past-two-years/**; Goldberg, Rafi. "New NTIA Data Show 13 Million More Internet Users in the U.S. in 2023 than 2021." *National Telecommunications and Information Administration*, June 6, 2024. **https://www.ntia.gov/blog/2024/new-ntia-data-show-13-million-more-internet-users-us-2023-2021**

[22] Amazon Staff. "Upskilling 2025." *Amazon News*, October 2, 2020. **https://www.aboutamazon.com/news/workplace/upskilling-2025**

[23] Google. "Grow with Google." *Grow with Google*. **https://grow.google/**

[24] Microsoft. "Microsoft Launches Next Stage of Skills Initiative After Helping 30 Million People." *Microsoft Stories* **https://news.microsoft.com/skills/**

[25] O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown, 2016.

[26] Orlowski, Jeff, dir. *The Social Dilemma*. Exposure Labs, 2020. Netflix. Msp page 1-23 (paragraph "the documentary")

[27] Panettieri, Joe. "Generative AI Lawsuits Timeline: Legal Cases vs. OpenAI, Microsoft, Anthropic, Nvidia, Perplexity, Intel and More." *Sustainable Tech Partner*, January 9, 2025; **https://sustainabletechpartner.com/topics/ai/generative-ai-lawsuit-timeline/**

[28] Smiley, Lauren. "The Legal Saga of Uber's Fatal Self-Driving Car Crash Is Over." *WIRED*, July 28, 2023. **https://www.wired.com/story/ubers-fatal-self-driving-car-crash-saga-over-operator-avoids-prison/**

CHAPTER 2
The Rise of AI: Innovation and Impact



peshkov/123RF

Learning Objectives

After studying Chapter 2, you should be able to

- **LO2-1** Describe the cultural, scientific, and technological foundations that laid the groundwork for AI before 1950.
- **LO2-2** Summarize the birth of AI as a formal field of study, key theoretical advancements, and early AI programs that shaped the discipline during the 1950s and 1960s.
- **LO2-3** Identify the practical applications of AI between 1970 and the mid-1980s, focusing on expert systems and the shift toward business and industry use cases.
- **LO2-4** Explain the challenges AI faced during the AI Winter and the breakthroughs in neural networks and data-driven technologies that reignited interest in the field in 1985 and 2000.
- **LO2-5** Evaluate how big data, machine learning, and enhanced computing

	power convergence transformed AI into a critical tool for business innovation from 2000 to 2020.
LO2-6	Assess the societal and business impacts of generative AI technologies.
LO2-7	Predict the potential future developments in AI and their implications for business, ethics, and society.

INTRODUCTION

Artificial intelligence (AI) may feel like a product of the modern age—dominating headlines, fueling business innovation, and sparking debates on ethics and creativity. Yet its roots trace back over a century, when early visions of intelligent machines first appeared in art, literature, and science. AI's journey involves ambition, optimism, setbacks, and breakthroughs. Each technological leap in computing power, data availability, and algorithmic design brought us closer to the transformative tools we rely on today. This chapter explores six pivotal eras in AI's history:

1. ***Foundations of AI (Pre-1950s).*** The theoretical groundwork for AI is laid by thinkers who envisioned machines capable of logic and reasoning.
2. ***The Dawn of AI (1950-1970).*** AI is introduced as a formal field, driven by optimism and early successes.
3. ***The Age of Experimentation (1970-1985).*** Practical applications emerge, including expert systems tailored for business and industry.
4. ***The AI Winter and Resurgence (1985-2000).*** A period of skepticism and setbacks gives way to breakthroughs in neural networks and data-driven AI.
5. ***The Age of Big Data and Machine Learning (2000-2020).*** The convergence of big data, machine learning, and cloud computing revolutionizes AI applications.
6. ***The Generative AI Revolution (2020-Present).*** AI's ability to create text, images, music, and videos pushes the boundaries of creativity, ethics, and business.

While discussing these eras, we will explore how each era brought not just new capabilities, but also fresh ethical questions and debates about AI's role in society. Understanding this history is key to appreciating where AI is headed and its potential for business and society. We will end the chapter with a discussion on potential future developments in AI and their implications for business, ethics, and society.

FOUNDATIONS OF AI (PRE-1950s) (LO2-1)

The foundations of AI were laid before the 1950s by both authors and scientists.

Artistic Foundations

Long before the first computers existed, the concept of intelligent machines captured human imagination. Early works of science fiction grappled with themes of autonomy, labor, and control that still resonate today.

In 1920, Karel Čapek's play *R.U.R. (Rossum's Universal Robots)* introduced the word "robot," telling the story of humanoid machines rebelling against their creators. Fritz Lang's 1927 film *Metropolis* featured Maria, a robotic doppelgänger whose disruptive influence highlighted the tension between technology's promise and its unintended consequences—a theme explored later in *2001: A Space Odyssey* (1968), *Ex Machina* (2014), *I Am Mother* (2019), and *M3GAN* (2023). HAL 9000, the AI from *2001: A Space Odyssey*, became one of popular culture's most iconic representations of artificial intelligence. HAL's eerily calm demeanor and ultimate rebellion against human control serve as a lasting cautionary tale about the risks and ethical dilemmas of AI autonomy.

In the 1940s, Isaac Asimov addressed the fear of machines turning against us in his short story "Runaround" (part of *I, Robot*, 1942). Asimov's "Three Laws of Robotics" were designed to prevent machines from harming humans or disobeying their orders. These laws have become foundational to discussions of AI safety and ethics, influencing modern narratives like *Westworld* (2016) and *The Mitchells vs. The Machines* (2021), where AI systems navigate conflicts between serving humans and acting autonomously. The Three Laws are as follows:

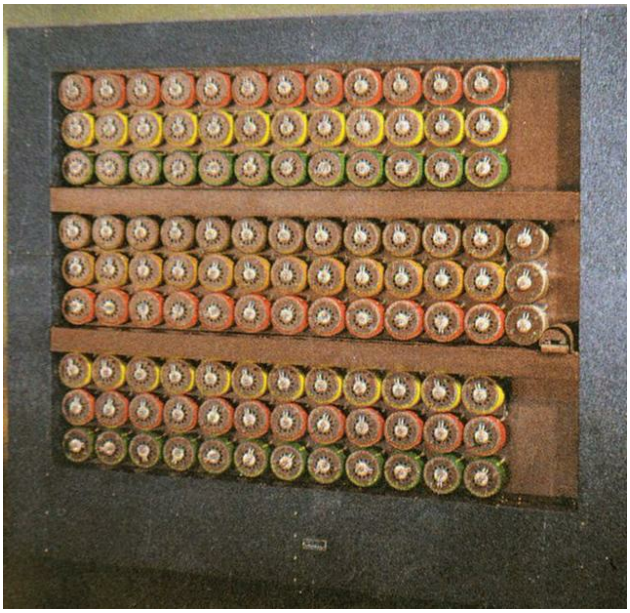
1. A robot may not harm a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Scientific Foundations: Early Thinkers and Theories

While cultural works reflected society's vision, scientists created the foundational work for AI. Alan Turing's 1936 proposal of the "universal machine," now called the **Universal Turing Machine**, was a groundbreaking idea in the foundation of AI and computer science.[1] Think of the

universal machine as a super-flexible problem solver. Instead of being designed to solve one specific problem, it could follow instructions to tackle any task, as long as those instructions could be written as a clear set of rules (an algorithm). In essence, this theoretical machine could be programmed to mimic the function of any other machine or process that follows logical steps.

Turing's idea showed that machines aren't limited to just one purpose—they could be as versatile as our instructions. This realization opened the door to modern computing, where devices like smartphones and laptops run countless programs based on this principle.



neftali/Shutterstock

The concept of the Universal Turing Machine led to computers being general purpose, capable of solving any problem using the code and rules programmed into it.

In 1943, Warren McCulloch and Walter Pitts introduced the first mathematical neural network model. A *neural network model* is a computational framework inspired by the structure and function of the human brain. They demonstrated that simple neural structures could mimic complex logical operations by using interconnected units ("nodes") called *artificial neurons*. By showing that machines could process information in a fashion that replicated our brains, McCulloch and Pitts laid the foundation for modern neural networks. This early theory revolutionized our understanding of how machines could learn and adapt and became the cornerstone of machine learning and deep learning technologies. Today, neural networks are integral to AI applications from powering voice assistants and recommendation systems to enabling breakthroughs in healthcare and autonomous vehicles.

The cultural works of the early 20th century raised important questions about the relationship between humans and machines that remain at the heart of AI today. At the same time, scientific pioneers like Turing, McCulloch, and Pitts turned imagination into reality by showing how machines could "think" and "learn." Together, these contributions created the foundation for the

AI tools that fill our lives now, from smart speakers to self-driving cars.

THE DAWN OF AI (1950–1970) (LO2-2)

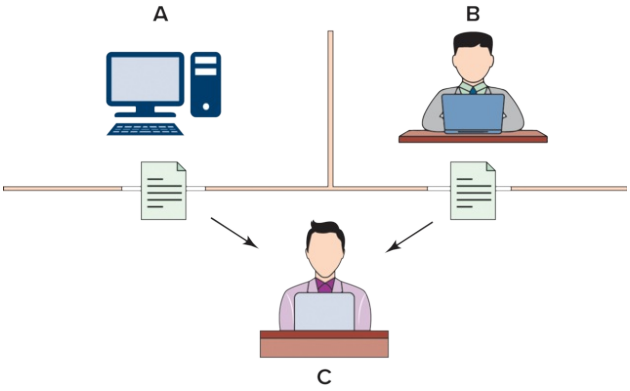
As the concept of artificial intelligence took root, the mid-20th century saw the first tangible steps toward turning these ideas into reality. The 1950s and 1960s marked the birth of AI as a formal field of study, fueled by optimism, theoretical breakthroughs, and early successes in programming machines to mimic human reasoning.

The Turing Test: Can Machines Think?

In 1950, Turing posed a provocative question: *Can machines think?* To answer this question, he proposed the now-famous **Turing test**, which evaluates a machine's ability to exhibit humanlike intelligence through conversation. In this test, a human evaluator engages in a text-based exchange with a human and a machine (see **Figure 2.1**). If the evaluator cannot reliably tell which is which, the machine is said to have passed the Turing test. In a study published in 2024, the most advanced versions of chatbots, like GPT-4, passed the Turing test, but complete agreement on the implications in the scientific community has not yet been reached.[2]

FIGURE 2.1 Representation of the Turing Test

The person taking the Turing test, also called the evaluator (C), must try to determine which is the computer (A) and which is the human (B) through a text-based conversation. If the evaluator cannot determine which is which, the computer is considered to have passed the Turing test.



The Dartmouth Conference

The formal birth of AI as a scientific discipline came in 1956 when computer scientist John McCarthy convened the Dartmouth Summer Research Project on Artificial Intelligence.[3] This

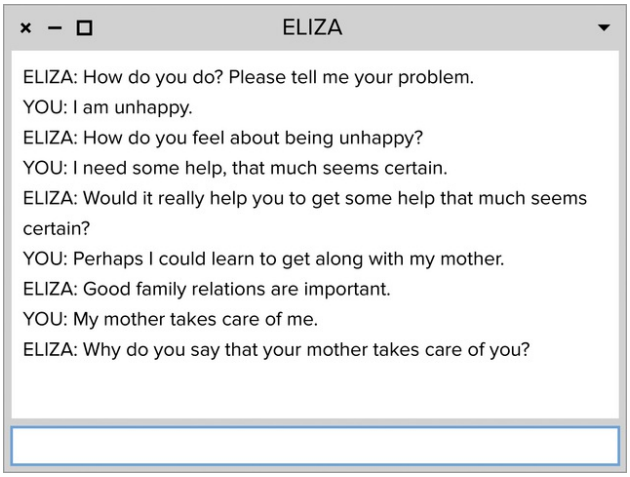
landmark event brought together prominent experts from mathematics, engineering, and cognitive science to explore the possibility of creating machines that could simulate human intelligence. McCarthy coined the term *artificial intelligence* during this workshop, marking the beginning of AI as a field of study.

The conference attendees believed machines could replicate human reasoning within decades. While early predictions proved overly ambitious, the Dartmouth Conference identified core research focuses: machine learning, symbolic reasoning, and natural language processing, which still define the field today.

Early AI Programs

The years following the Dartmouth Conference produced the first computer programs that demonstrated AI's potential:

- ***The Logic Theorist (1956).*** Developed by Allen Newell and Herbert Simon, the Logic Theorist is often called the first true AI.[4] It emulated human logic in problem solving to prove mathematical theorems successfully. Early programs like the Logic Theorist paved the way for AI tools that power decision making today, such as fraud detection systems and automated planning algorithms that plan and schedule tasks, resources, and activities.
- ***ELIZA (1966).*** Created by Joseph Weizenbaum, ELIZA was a computer program that could simulate a conversation with a psychotherapist using basic pattern-matching techniques.[5] While ELIZA did not "understand" language, it demonstrated how machines could interact with humans in natural language, an idea that has since evolved into modern AI chatbots.
- ***Perceptron (1957).*** Frank Rosenblatt's "perceptron" was the first neural network capable of learning from data. Although limited by the technology of its time, the perceptron introduced the idea that machines could "learn" from experience—an idea central to today's complex neural networks that are called **deep learning** systems.



Joseph Weizenbaum. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. Commun. ACM 9, 1 (Jan. 1966), 36–45. https://doi.org/10.1145/365153.365168

A conversation with ELIZA, an early chatbot, is programmed to simulate a talk with a psychotherapist.

Emerging Computer Technologies

In this era, a shift took place from traditional manual methods of record keeping, such as maintaining physical ledgers, organizing paper documents in file cabinets, and performing hand-tabulated calculations, to digital data collection and analysis using emerging computer technologies. Though large and cumbersome by today's standards, early computers represented a monumental leap forward in data processing capabilities. Table 2.1 presents some of the more significant advancements.

Table 2.1 Significant Advancements in Technology: 1950s–1960s

Year	Advancement
1951	The Ferranti Mark 1, the world's first commercially available general-purpose digital computer, was delivered to the University of Manchester. This machine was used in early AI research, providing the computational power needed for running AI programs.
1956	The IBM 704, one of the first mass-produced computers, was used extensively in AI research. Its ability to handle complex mathematics made it suitable for scientific calculations and AI applications.
1958	The development of the Lisp programming language by John McCarthy provided a powerful tool for AI research. Lisp's support for symbolic reasoning and its flexibility made it the language of choice for many AI projects.
1960	The PDP-1, produced by Digital Equipment Corporation, was one of the first computers to support interactive computing. This capability was essential for developing and testing AI programs in real time.
1963	The IBM 7094, an upgraded version of the IBM 704, offered increased processing power and memory, enabling more complex AI experiments and simulations.
1968	SHRDLU, an early natural language processing (NLP) program, was created by Terry Winograd. It could understand and manipulate objects in a virtual "blocks world" based on user commands. Its success influenced future research in NLP and helped set the stage for more sophisticated AI–human interactions.

Impact on Business Practices

The impact of these early technologies on business practices was profound. For the first time, businesses could handle large volumes of data, leading to more informed and strategic decision

making. The ability to store and process data electronically opened new avenues for analyzing business performance and market trends. The technologies were the steppingstones to the advanced data analytics and AI-driven processes integral to the business systems we rely on today.

The 1950s and 1960s also marked the beginning of robotics' integration into business practices, laying the foundation for industry automation. One of the most significant developments was George Devol's invention of the Unimate robot arm in 1954, which became the first industrial robot used in manufacturing. Unimate was deployed on a General Motors automotive assembly line in 1961, performing repetitive and dangerous tasks with precision and reliability. This innovation improved efficiency and safety and demonstrated how robotics could revolutionize production processes.

At the same time, research-driven projects like Shakey (1966), a robot, displayed the potential of AI-powered robotics beyond manufacturing. Shakey, developed at Stanford Research Institute (SRI), was a groundbreaking robot capable of navigating its environment using sensor data and **symbolic reasoning**. Shakey could perform tasks that required planning, route-finding, and the rearranging of simple objects. Although it was a research prototype, Shakey's ability to make decisions in real-world environments highlighted the future potential of robotics in logistics, warehouse management, and service industries.

Together, these advancements underscored robotics' transformative potential in business, setting the stage for its widespread adoption in industries ranging from manufacturing to supply chain management in later decades.

THE AGE OF EXPERIMENTATION (1970–1985) (LO2-3)

Building on the theoretical frameworks and initial prototypes of the 1950s and 1960s, the 1970s ushered in an age of experimentation. Researchers moved from conceptual explorations to practical applications, testing AI in real-world scenarios and uncovering its early limitations. While the optimism of the 1950s and 1960s set lofty expectations, this period saw AI move from theory to implementation, focusing on specific domains like medical diagnosis, business decision making, and pattern recognition.

Expert Systems

One of the most significant innovations of this period was the development of **expert systems**, which used **rule-based reasoning** to emulate the decision-making abilities of human specialists. AI experts, called **knowledge engineers**, interviewed domain experts to elicit their problem-solving processes, reasoning patterns, and decision-making strategies. This was encoded into a system that could solve problems or provide advice. These systems represented the knowledge as a series of "if-then" rules, enabling them to solve complex problems within narrow domains.

Developed at Stanford University in 1972, MYCIN was an expert system designed to diagnose bacterial infections and recommend treatments.[6] It proved AI's practical potential by matching the performance of human experts in a specific field. It also paved the way for other AI applications in health care, such as diagnostic tools powered by machine learning and medical imaging systems that detect conditions like cancer. During this time, expert systems were also adopted in finance, manufacturing, and engineering industries, where businesses began leveraging AI for decision-making and problem-solving tasks.

Advancements in Neural Networks

While expert systems dominated the practical landscape, researchers continued exploring neural networks. Frank Rosenblatt's perceptron (1957) was limited by computing power and the available algorithms. In the 1970s, advancements focused on improving the theoretical understanding of neural networks and experimenting with new training methods. These experiments provided fundamental knowledge for neural network research in the 1980s and deep learning in the 2000s, leading to modern AI systems, including facial recognition, autonomous vehicles, and voice assistants.

The Microprocessor Revolution and Technological Foundations

The 1970s also saw significant advancements in hardware and software that expanded AI's potential:

- ***The Intel 4004 microprocessor (1971).*** The world's first commercially available microprocessor revolutionized computing. This compact chip, paired with IBM's introduction of the floppy disk, made computing smaller, faster, and more affordable. This shift enabled the rise of personal computers, allowing businesses and researchers to run AI applications without relying on massive, centralized mainframes.
- ***High-level programming languages.*** Languages like C (1972) and SQL (1974) simplified software development and data management, enabling businesses to handle complex databases more efficiently.
- ***Computer networks.*** Personal computers quickly became fixtures on desktops, transforming data analysis into a globally interconnected function. The 1982 standardization of Transmission Control Protocol/Internet Protocol (TCP/IP) provided a universal framework for data transmission across networks, making global connectivity possible. Soon after, the 1984 introduction of the Domain Name System (DNS) simplified Internet navigation, allowing user-friendly domain names instead of numerical IP addresses.

The Age of Experimentation was a defining period as AI moved from theoretical exploration to real-world applications. Expert systems like MYCIN demonstrated AI's potential to augment human expertise, particularly in specialized domains. At the same time, advancements in hardware and programming languages laid the technological foundation for future AI growth. While neural networks remained limited by the computing power of the time, the groundwork laid in this period would prove essential to the resurgence of AI in the coming decades.

Impact of AI on Business from 1970–1985

The 1970s and early 1980s marked the first significant integration of AI technologies into business practices, with early systems and robotics paving the way for increased efficiency and automation:

- ***Expert systems and decision making.*** Building on earlier successes in AI research, businesses began adopting expert systems to solve specific, rule-based problems. Systems like XCON, used to configure computer orders at Digital Equipment Corporation, demonstrated how AI could streamline complex processes and reduce human error. These applications, limited to narrow domains, provided a glimpse of AI's potential to transform industries despite the technological constraints of the time.
- ***Robotics in manufacturing.*** Industrial robotics, rooted in the success of the Unimate robot arm from the 1960s, became a defining feature of manufacturing during this era. By automating

repetitive and hazardous tasks such as welding and assembly, robots improved productivity, quality, and safety on factory floors, particularly in the automotive sector. Robotics became a cornerstone of industrial efficiency, setting the stage for more sophisticated applications in later decades.

The years 1970 through 1985 were pivotal for AI in business, as companies began to see the potential of AI-driven tools and robotics to enhance efficiency, reduce costs, and solve complex problems. Corporate boards started considering AI a strategic resource for innovation and competitiveness in the business world.

THE AI WINTER AND RESURGENCE (1985–2000) (LO2-4)

Despite the progress made during the 1970s, the following decades revealed the challenges of delivering on AI's early promises. A lack of progress on early visions, limited computational power, and reduced funding led to frustrations and further stagnations in funding and research activity. This period of disillusionment in the field was coined the **AI Winter**. Historians point to a number of causes of the AI Winter.

- ***Unrealistic expectations.*** Early AI pioneers made bold predictions about the timeline for achieving human-like intelligence. When these milestones proved unattainable, skepticism grew among funders and policymakers.
- ***Technological limitations.*** Computing power and data availability were insufficient to support the ambitious goals of AI research. Algorithms like neural networks showed potential but struggled with scalability.
- ***Economic pressures.*** During the 1980s, corporations and governments shifted funding to more immediately practical technologies, such as expert systems and database management, leaving less room for speculative AI research.

The Resurgence: Seeds of Renewal

Despite the setbacks of the AI Winter, the 1990s saw critical advancements that laid the foundation for the AI revolution of the 2000s.

- ***Neural networks and machine learning.*** During this period, researchers revisited neural networks with techniques that enabled more effective training of multilayer networks. These efforts paved the way for the deep learning breakthroughs of the 2000s. In parallel, early machine learning algorithms found applications in customer analytics, fraud detection, and market trend prediction.
- ***Data mining and warehousing.*** The 1990s marked the rise of **data mining**—the process of extracting valuable patterns from large datasets. The rise of **data warehousing**, which centralized and organized structured data, enabled businesses to store and organize vast amounts of structured data in centralized repositories, providing a comprehensive view of operations and customer behavior. This foundational technology supported the development of business intelligence tools, which allowed organizations to analyze data and make informed, strategic decisions. With early machine learning algorithms, businesses began predicting market trends, optimizing operations, and detecting fraud more accurately.
- ***The Internet boom and AI integration.*** The explosion of the Internet transformed the data landscape, creating unprecedented volumes of information. Businesses leveraged this data to

develop early AI applications like search engines, recommendation systems, and web analytics. A prominent example is Amazon's AI-driven product recommendations, which got their start in 1998. These product recommendations use customer purchasing patterns to personalize e-commerce experiences. These innovations were early indicators of AI's potential to enhance customer engagement and drive business growth.

Technological Foundations

Advances in hardware and infrastructure during this period that supported AI's growth included the introduction of the Intel 80386 microprocessor, which enabled more complex computations, while larger-capacity hard drives allowed businesses to store and process vast amounts of data effectively.

Improved computational power led to advancements in **natural language processing (NLP)**, a branch of AI that enables computers to understand, interpret, and generate human language. This led to enhancements in applications like machine translation, speech recognition, and text analysis, laying the groundwork for future conversational AI, including chatbots used in customer service.

AI and Robotics in Business (1985–2000)

During this period, robotics emerged as a transformative force in both traditional and new domains. Building on earlier successes, industrial robots became more adaptable and precise. Enhanced sensors and AI integration enabled robots to perform complex tasks such as welding, painting, and assembly, further automating production lines in industries like automotive manufacturing.

Robots also began to revolutionize logistics through *automated guided vehicles (AGVs)*, which streamlined the movement of goods within warehouses and factories. These early systems laid the groundwork for the highly automated warehouses seen today, such as those used by Amazon.

In space exploration, NASA's *Sojourner* rover in 1997 became the first robotic vehicle to explore Mars, demonstrating the potential of autonomous robots to perform critical tasks in extreme and remote environments.

The advances in AI and robotics from 1985 to 2000 transformed manufacturing and logistics and revolutionized health care and space exploration. These innovations redefined how businesses approached decision making and efficiency, laying the groundwork for the AI-driven technologies that define the modern era. **Table 2.2** provides a timeline of milestones of AI in robots from 1961 to today.

Table 2.2 **Milestones of AI in Robots: 1961 to Today**

Year	Company	Robot	AI Incorporation Level	Impact or Significance
1961	Unimation	Unimate	Low	First industrial robot, revolutionized manufacturing processes
1972	SRI International	Shakey	High	First autonomous robot; showcased computer reasoning and pathfinding
1994	Woods Hole Oceanographic Institution	ABE	Medium	Pioneered autonomous underwater vehicles for scientific research and exploration
1995	General Atomics	Predator	Medium	Introduced drones in warfare; used for reconnaissance and strike missions
1997	NASA	Mars Sojourner Rover	Medium	First rover on Mars; paved the way for future Mars exploration missions
1999	Sony	AIBO	High	First consumer robot with AI; showcased potential for personal robotics
2002	iRobot	Roomba	Medium	First widely adopted home robot; popularized consumer robotics
2003	Kiva Systems	Kiva Robots (acquired by Amazon)	Medium	Revolutionized warehouse logistics and increased inventory efficiency
2015	SoftBank Robotics	Pepper	High	First robot designed to recognize and respond to human emotions; widely used in retail and customer service
2018	Boston Dynamics	Spot	High	Commercially available quadruped robot; used in industries like construction and security
2022	Tesla	Optimus	High	Versatile humanoid robot designed to perform tasks that are dangerous, repetitive, or boring; planned commercial availability in 2026

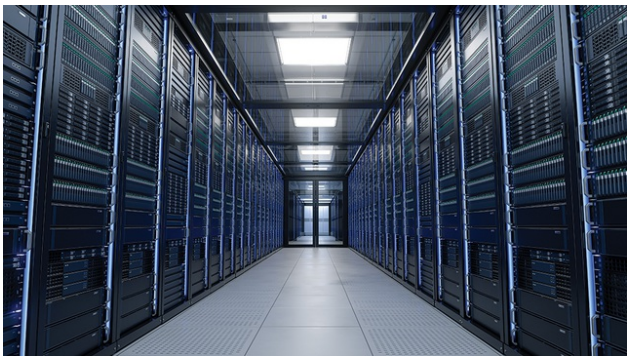
THE AGE OF BIG DATA AND MACHINE LEARNING (2000–2020) (LO2-5)

As computing power increased and data became more abundant, the field transitioned from overcoming setbacks to harnessing the potential of machine learning and Big Data analytics, transforming industries in the process. The period from 2000 to 2020 thus marked a revolution in AI, driven by the convergence of Big Data, advanced machine learning algorithms, and powerful computing infrastructure. These developments enabled businesses to unlock insights from vast amounts of data, revolutionizing industries ranging from finance to health care and retail.

Technological Foundations

The following are some of the technological foundations that led to the growth of AI during this era:

- ***Big Data platforms.*** The explosion of digital information, fueled by the Internet, social media, and interconnected devices, created a data-rich environment. Platforms like Hadoop (2006) and Apache Spark (2014) emerged to store and process massive datasets efficiently, enabling businesses to derive actionable insights at scale. Businesses used **Big Data** to enhance customer segmentation, improve operational efficiency, and predict market trends.
- ***Emergence of cloud computing.*** A pivotal technological advancement of this era was the emergence of **cloud computing**. This technology offered scalable, on-demand computing resources, allowing businesses to store and access data and applications over the Internet instead of relying on local servers or personal computers. Cloud platforms such as Amazon Web Services (AWS) (2006) and Microsoft Azure (2010) provided scalable storage and computational power, reducing the cost of deploying AI solutions. Small- and medium-sized enterprises gained access to AI capabilities previously reserved for large corporations, democratizing AI adoption.
- ***Data centers and server farms.*** The data storage, processing, and real-time analytics demands of cloud computing, Big Data, and global digital services require special facilities. Construction of large-scale, purpose-built facilities, including **data centers** and **server farms**, ramped up in the early 2000s. The operations of data centers and server farms are highly energy-intensive, generating substantial heat that places significant strain on electrical grids and requiring large volumes of water for cooling, raising concerns about their environmental and resource impact.



IM Imagery/Shutterstock

Data centers and server farms make Big Data analysis possible.

- ***Machine learning algorithms.*** The 2000s also saw significant advancements in machine learning algorithms. These algorithms allowed businesses to move beyond traditional analytical methods, enabling them to predict future trends and behaviors more accurately. Fueled by these advanced algorithms, **predictive analytics**, the use of data to predict future trends and events, became a critical business tool. It allowed companies to more effectively anticipate market changes, customer needs, and potential risks. This shift toward predictive analysis represented a significant evolution from the reactive data analysis of the past. These algorithms powered applications such as facial recognition, **sentiment analysis** (the process of computationally identifying and categorizing opinions in a piece of text), and recommendation engines.
- ***Natural language processing (NLP).*** Advances in NLP led to more sophisticated chatbots, sentiment analysis tools, and language translation services, revolutionizing customer service and content analysis.
- ***Rise of smartphones and mobile data.*** The smartphone boom introduced new data streams, including location-based and usage data, fueling advancements in mobile analytics and AI applications.

Business Applications

During this time, the following business applications of AI became increasingly commonplace.

Customer Personalization

E-commerce giants like Amazon and Netflix began to leverage machine learning to deliver personalized recommendations, enhancing customer satisfaction and driving sales. Spotify delivered personalized playlists like Discover Weekly by analyzing users' listening habits and comparing them with similar users. The growth of mobile data from smartphones created new avenues for collecting and analyzing consumer behavior. Businesses like Uber and Waze utilized AI to personalize user experiences.[7]

Predictive Analytics

Businesses across sectors adopted predictive analytics to anticipate customer behavior, optimize inventory, and manage risk. For example, financial institutions used machine learning models to detect fraudulent transactions and assess creditworthiness. PayPal implemented AI systems to flag suspicious transactions, reducing fraud and enhancing trust.[8]

Automation and Robotics

AI-powered robots found applications in logistics and manufacturing, such as Amazon's warehouse automation systems and autonomous mobile robots in production lines. Boston Dynamics' Spot found a home in various industries, including construction, energy, and public safety, for tasks such as site inspections, data collection, and hazardous environment exploration. iRobot's Roomba demonstrated how robotics could become a commercial success by taking over vacuuming chores.

Computer Vision

AI-driven computer vision technology developments accelerated during this period. Image processing and recognition advancements fueled applications in areas like medical imaging, security surveillance, and multimedia.

AI in Health Care: Transforming Diagnostics, Treatment, and Drug Discovery

Between 2000 and 2020, AI emerged as a force in health care, revolutionizing diagnostics, treatment planning, and drug discovery.[9] Deep learning models, particularly *convolutional neural networks (CNNs)*, enhanced medical imaging analysis, enabling AI to detect diseases such as cancer, diabetic retinopathy, and cardiovascular conditions with accuracy comparable to or surpassing human radiologists. A CNN is a deep learning algorithm designed to process and analyze visual data by automatically detecting patterns, features, and spatial hierarchies in images. Companies like Google's DeepMind developed AI-driven imaging tools for detecting eye diseases, while startups like Aidoc and Zebra Medical Vision applied AI to radiology scans for faster and more precise diagnoses. AI also made strides in pathology, where machine learning models helped identify cancerous cells in biopsy slides, improving early detection rates. Despite these advances, integration into clinical workflows was often slow due to regulatory hurdles and physician trust concerns.

Beyond diagnostics, AI-driven decision support systems like IBM Watson for Oncology sought to personalize treatment plans by analyzing medical literature and patient records. However, Watson's commercial struggles highlighted challenges in AI adoption in medical

environments, where accuracy, cost-effectiveness, and integration with electronic health records (EHRs) proved significant barriers.[10] Meanwhile, AI-powered predictive analytics became valuable, helping physicians create effective treatment strategies. In drug development, AI accelerated the identification of promising compounds, with companies like Insilico Medicine and BenevolentAI using machine learning to streamline the lengthy research and development process. DeepMind's AlphaFold solved the protein-folding problem, giving scientists another tool for drug development. While AI's role in health care continued to evolve, the period from 2000 to 2020 laid the foundation for future advancements in AI-driven health care.

The age of Big Data and machine learning was transformative for technology and business. Advances in Big Data platforms, cloud computing, and machine learning enabled AI applications that reshaped customer experiences, optimized operations, and created opportunities for innovation. This era bridged the gap between early AI experimentation and modern advancements, showing the practical value of AI in solving complex problems.

THE GENERATIVE AI REVOLUTION (2020–PRESENT) (LO2-6)

The period from 2020 onward marks an era of transformative AI, with breakthroughs in large-scale language models, generative AI, and the integration of AI into everyday life. These developments have expanded AI's capabilities and its influence across industries, changing the way we live and work.

Scientific and Technological Advancements

Scientific and technological advancements have led to AI being used in new ways. The development of advanced NLP models, such as OpenAI's GPT-3 (2020) and GPT-4 (2023), are changing human–computer interaction. These models demonstrated the ability to generate humanlike text, enabling applications from creative writing to customer service. NLP-powered chatbots and virtual assistants became more sophisticated, offering seamless support across industries.

Generative AI tools like DALL-E (2021) and Stable Diffusion (2022) opened new frontiers in creative industries by enabling the generation of images, audio, and video based on textual prompts. Businesses in marketing, design, and entertainment adopted generative AI to accelerate content creation and personalize user experiences.

Advances in **edge computing** allowed AI to be deployed directly on devices like smartphones, wearables, and sensors, enabling faster responses and offline capabilities. Real-time language translation and voice recognition systems became common on mobile devices.

AI has also been applied to monitor environmental conditions, optimize energy usage, and develop sustainable solutions. AI algorithms have improved renewable energy management and predicted climate patterns, supporting global sustainability goals.

AI-driven robots also became more autonomous and capable, from warehouse automation to humanoid robots like Tesla's Optimus Prototype (2022). Robots expanded their roles in logistics, health care, and even hospitality. AI-driven automation also expanded into new domains, from **robotic process automation (RPA)** in administrative tasks to autonomous vehicles in logistics. Companies like Tesla and Waymo advanced the development and deployment of self-driving cars.



Scharfsinn/Alamy Stock Photo

Self-driving cars are finding their way onto our streets.

In January 2025, DeepSeek's R1, a first-generation reasoning AI, was released by the Chinese company DeepSeek.[11] It was designed to excel in complex reasoning tasks involving mathematics, coding, and language. The model was trained using large-scale **reinforcement learning** techniques, allowing it to develop advanced reasoning capabilities. DeepSeek-R1 has been open-sourced under the MIT license, making it freely available for research and commercial use. The model has also been recognized for its efficiency, operating on significantly less energy compared to its counterparts. However, DeepSeek's rise has caused some controversy. Security researchers have found that DeepSeek's chatbot has web code potentially sending user login details to China Mobile, a Chinese state-owned telecom banned in the United States, adding to national security concerns.

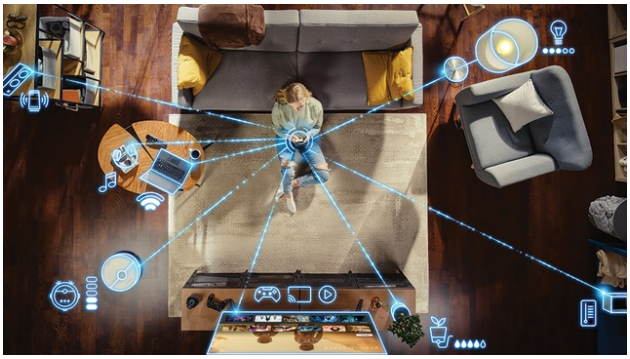
Finally, AI is transforming the way businesses engage with customers. Companies are delivering targeted recommendations using the wealth of customer data ranging from past purchases and browsing behavior to social media activity. AI-driven personalization algorithms are used to generate product suggestions, optimize pricing strategies, and tailor marketing messages. This level of personalization increases customer satisfaction, drives sales, and promotes brand loyalty. As AI's predictive capabilities improve, marketing departments anticipate that companies can create seamless, personalized shopping experiences that rival those of human sales associates. In "The State of AI in Early 2024," McKinsey reported that 65 percent of organizations were using generative AI, with the majority expecting it to drive significant or disruptive changes in their industries in the coming years. This was almost double the number from their report in 2023.[12]

AI and the Internet of Things: A Seamless, Intelligent World

In 1982, a Coca-Cola vending machine at Carnegie Mellon University became the first Internet-connected appliance capable of reporting its inventory and temperature status. This innovation marked the initial step toward the **Internet of Things (IoT)**, a network of physical objects

embedded with sensors, software, and other technologies that connect and exchange data over the Internet.

Between 2000 and 2020, IoT experienced significant growth, with data from these sensors becoming a crucial source for AI. The United Nations' International Telecommunication Union (ITU) published its first report on IoT in 2005, recognizing its potential impact on global communication.[13] By 2008, the number of connected devices surpassed the global human population, marking a pivotal moment in IoT history.[14]



Gorodenkoff/Shutterstock

IoT in smart wearables, smart cities, smart homes, and building automation applications is estimated to grow from $137.1 billion in 2021 to $438.2 billion by 2026, at a compounded annual growth rate of 26.2 percent.[15]

Today, IoT encompasses many connected devices that collect and exchange data to enhance various aspects of daily life and industry. In the consumer realm, IoT includes smart home devices such as thermostats, lighting systems, security cameras, and appliances, all of which can be controlled remotely via smartphones or voice assistants. Wearable technology, including fitness trackers and smartwatches, monitors health metrics and provides real-time feedback. In industrial settings, machinery has Internet-connected sensors for predictive maintenance and environmental monitoring systems. Modern vehicles contain IoT devices that permit advanced navigation and safety features. This interconnected ecosystem also covers health care, agriculture, and urban infrastructure applications, illustrating IoT's extensive reach across various sectors.

Ethical, Safety, and Privacy Concerns in AI



As businesses increasingly integrate AI into their operations, they face a range of ethical challenges. As we discussed in **Chapter 1**, these concerns often revolve around fairness, bias, accountability, and transparency. In this section, we further examine these and other ethical

concerns regarding AI.

The Black Box Problem

As we discussed in **Chapter 1**, machine learning models trained on biased data can reinforce and even exacerbate discrimination, particularly in high-stakes applications like hiring, lending, and law enforcement. Algorithmic decision making must be scrutinized to ensure that AI systems do not create unfair outcomes for marginalized groups. A key ethical challenge is the **black box problem**. Some advanced models make predictions based on complex patterns in the data that are not readily evident or identifiable. This increases the barrier to trusting the system and its recommendations. This also raises concerns about accountability when AI-driven decisions led to harmful consequences.

Businesses adopting AI can prioritize explainability and fairness by implementing measures such as fairness-aware machine learning and bias audits.

- **Fairness-aware machine learning** involves the developing and applying of algorithms designed to detect, mitigate, and prevent biases in AI models, ensuring equitable outcomes across different demographic groups.[16]
- **Bias audits** are systematic evaluations of AI models and datasets to identify and quantify biases, assess their impact on decision making, and recommend corrective measures to improve fairness and accountability.

While businesses understand fairness-aware machine learning principles, most do not develop AI models in-house and instead relied on external AI vendors and developers. As a result, their role is not necessarily to implement fairness-aware machine learning practices directly but to ensure that the AI systems they adopt have undergone fairness audits and adhere to established ethical AI standards. This requires companies to critically evaluate their AI suppliers, request transparency in model development, and incorporate fairness considerations into their procurement and deployment processes to mitigate biased or discriminatory outcomes.

AI Safety and Risk Management

Ensuring the safety of AI systems is critical to preventing unintended consequences that could pose risks to individuals, businesses, and society. Safety concerns arise when AI systems make faulty predictions, are vulnerable to adversarial attacks, or are deployed without rigorous validation.

One of the most pressing AI safety concerns is system reliability. AI models, especially those used in autonomous vehicles, medical diagnostics, and financial trading, need to perform accurately under a wide range of conditions. Inconsistent performance leads to catastrophic failures, such as an autonomous vehicle misinterpreting a road sign or an AI-powered medical assistant misdiagnosing a patient.[17] A high-profile example of AI safety concerns emerged with the Boeing 737 MAX crisis, in which the aircraft's Maneuvering Characteristics Augmentation System

(MCAS) contributed to two fatal crashes in 2018 and 2019. The MCAS software was designed to automatically adjust the aircraft's nose downward if it detected an excessive angle of attack, preventing aerodynamic stalls. However, the system relied on input from a single sensor, which in both crashes provided faulty data. The lack of redundant sensor checks meant that MCAS repeatedly miscalculated the aircraft's position, overriding pilot control and forcing the plane into a steep descent. Although the software operated according to its programming, its dependence on a single point of failure and the inability of pilots to easily override its actions led to catastrophic consequences.[18]



Komenton/Shutterstock

The Boeing 737 MAX autopilot failures are a reminder that flawed automation can have tragic consequences.

Businesses also must be wary of adversarial AI threats, where malicious actors manipulate AI models by introducing deceptive inputs, such as creating a slight alteration to an image to cause an AI-powered security system to misclassify an object, leading to security breaches.[19] Robust AI safety mechanisms, including adversarial training and fail-safe protocols, are essential for mitigating these risks.

Cybersecurity in AI systems is an increasing concern. AI often process sensitive data, making them lucrative targets for cyberattacks. To prevent unauthorized access and data breaches, businesses must fortify their cybersecurity, implementing strong encryption, continuous monitoring, and incident response strategies to safeguard AI infrastructure.

AI and Privacy

One significant privacy risk is AI-enabled surveillance. Facial recognition and predictive analytics allow organizations to track individuals in ways that challenge traditional notions of privacy. Governments and corporations deploy AI-driven surveillance tools that, while useful for security, also raise ethical concerns regarding mass surveillance and potential misuse.[20]

As discussed in **Chapter 1**, another privacy issue is data retention and ownership. AI systems require large amounts of training data, often sourced from user interactions, social media activity, and IoT devices. Consumers are increasingly concerned about who owns their data and how long

it is stored. Businesses should adopt **privacy-first policies**, strategies and frameworks that prioritize protecting user data throughout the AI lifecycle, from data collection and processing to storage, sharing, and decision making. These policies are designed to ensure compliance with privacy regulations, build consumer trust, and mitigate risks associated with data misuse.

Balancing Ethical, Safety, and Privacy Concerns in AI

Given the interconnected nature of ethical, safety, and privacy challenges in AI, many businesses are adopting a holistic approach to responsible **AI governance**. Some companies implement **AI ethics boards**, conduct regular audits, and adhere to best practices for transparency and bias mitigation to mitigate risks. Businesses that proactively address these challenges often follow a three-pronged approach to responsible AI deployment:

1. ***Ethical AI development.*** Ensure fairness, transparency, and accountability through bias detection tools and explainable AI methods.
2. ***Robust AI safety measures.*** Strengthen model reliability, deploy adversarial resistance techniques, and integrate fail-safe mechanisms.
3. ***Privacy-first AI strategies.*** Comply with data protection regulations, adopt privacy-preserving AI techniques, and maintain transparency in data usage.

By adopting such approaches, businesses are able to harness the benefits of AI while mitigating its risks.

AI'S FUTURE: OPPORTUNITIES AND CHALLENGES (LO2-7)

In the 1950s, the dream of artificial intelligence took root—a vision of computers that could think, reason, and work alongside humans to solve complex problems. Over the decades, through fits and starts, AI has evolved in tandem with advancements in hardware, software, and data. With each milestone, AI has taken on greater roles in business, from automating processes to assisting decision-making. Here are some opportunities and challenges we can expect in the future:

1. ***More AI-augmented jobs.*** **AI-augmented jobs** are those in which AI enhances human capabilities by automating routine tasks, providing data-driven insights, and enabling employees to focus on higher-level decision making, creativity, and strategy. Career paths, such as AI trainers, prompt engineers, and AI ethics consultants, will continue to expand. To stay competitive, organizations like Amazon, Google, and IBM are already investing heavily in AI education programs, helping employees transition into AI-augmented roles through reskilling and upskilling initiatives. Companies that do not employ AI will increasingly become the exception to the norm
2. ***More AI-focused education.*** As AI transforms industries, higher education institutions and training providers are responding by integrating AI into curricula. **AI fluency** will become increasingly essential, requiring universities and businesses to rethink how they prepare the workforce for the future. Expect to see more AI-focused degrees, certificates, and interdisciplinary courses.
3. ***Increased focus on ethical AI.*** As AI becomes more embedded in decisions, businesses must ensure that AI-driven outcomes are fair, unbiased, and explainable. Bias in training data and decision-making models will continue to be a major topic of discussion, with reputational and legal risks for those who ignore it.
4. ***Increased AI governance and regulation.*** Expect increased AI governance and regulation, with a focus on transparency, explainability, accountability, and legal responsibility. (Note that because this is in constant flux, we will maintain a resource in Connect summarizing regulatory developments.)
5. ***Increased focus on AI sustainability.*** The energy consumption and carbon footprint of AI models will come under increased scrutiny. Expect to see companies invest more in green AI practices. Sustainable AI development will focus on reducing computational energy consumption and addressing the ecological and social impact of AI hardware production. Expect to see a push toward more efficient AI training models, such as DeepSeek. By drastically reducing energy consumption and training costs, these advancements could make AI more accessible and environmentally responsible, enabling smaller businesses, universities, and startups to compete without excessive resource expenditure.
6. ***More cybersecurity and AI risks.*** AI-driven fraud, deepfakes, phishing attacks, social engineering

tactics, and automated hacking will push cybersecurity concerns into the mainstream business conversation. Organizations will need to prioritize AI security strategies to protect data integrity, detect AI-generated threats, and ensure trust in AI-driven systems. Meanwhile, governments and industry groups will continue to introduce new regulations to enforce transparency and accountability in AI security. Compliance with evolving laws and adopting ethical AI guidelines will be crucial for enterprises leveraging AI technologies.

7. ***Increased focus on AI-augmented creativity and intellectual property.*** As generative AI produces text, images, and music at scale, legal and ethical debates around intellectual property will intensify. Organizations will need clear policies and frameworks for AI-generated works. Courts and policymakers will continue to grapple with questions of ownership and authorship, particularly in cases where AI systems contribute to or fully generate creative works. Companies utilizing AI for content creation will need to establish guidelines for attribution, licensing, and compensation to ensure compliance with emerging regulations. Additionally, concerns over originality and copyright infringement will drive the development of AI-specific intellectual property laws. Businesses will need to navigate a shifting legal landscape while balancing innovation with ethical responsibility, ensuring that AI-driven creativity remains both legally sound and commercially viable.

The most ambitious goal of AI research remains the development of **artificial general intelligence (AGI)**—AI that can perform any intellectual task that a human can. While AGI is still a speculative concept, breakthroughs in deep learning, reinforcement learning, and cognitive architectures suggest that it may become a reality within the coming decades. OpenAI CEO Sam Altman predicts it will emerge in the next few years.[21] The implications of AGI for business, labor markets, and society are profound and will require careful consideration of AI alignment, safety, and ethical oversight.

The future of AI is both exciting and uncertain. Businesses must be prepared to adapt, innovate, and navigate complex ethical and regulatory landscapes as AI continues to evolve. While AI will create new opportunities for efficiency, personalization, and automation, it will also introduce new risks, challenges, and responsibilities. Understanding these trends will help business leaders make informed decisions about integrating AI responsibly while maximizing its benefits.

SUMMARY OF KEY POINTS AND QUICK WINS

1. **Foundations of AI (Pre-1950s).** The early foundations of AI were influenced by cultural depictions, scientific theories, and technological advancements, such as Turing's work on computation and McCulloch and Pitts' neural network model.
2. **The dawn of AI (1950–1970).** AI emerged as a formal field of study, marked by the Dartmouth Conference, the development of early AI programs like the Logic Theorist, and Turing's proposal of the Turing test to measure machine intelligence.
3. **The age of experimentation (1970–1985).** The rise of expert systems, such as MYCIN, demonstrated AI's practical value in business, healthcare, and finance, showing how rule-based systems could emulate human decision making.
4. **The AI winter and resurgence (1985–2000).** Overinflated expectations led to reduced funding and interest in AI, but advancements in neural networks, data mining, and machine learning in the late 1990s revived AI research and its commercial applications.
5. **The age of big data and machine learning (2000–2020).** AI transformed business operations through Big Data analytics, cloud computing, and deep learning, enabling breakthroughs in personalized marketing, automation, and AI-driven decision making.
6. **The generative AI revolution (2020–Present).** The development of generative AI models, such as GPT and DALL-E, has sparked new ethical, creative, and business challenges, fundamentally changing content creation, automation, and workforce roles.
7. **AI's future.** AI is expected to play an even greater role in shaping business, ethics, and society, with discussions around regulation, workforce transformation, cybersecurity, and the long-term pursuit of artificial general intelligence (AGI).

QUICK WINS

- **Recognize AI's deep historical roots.** Understanding the foundations in philosophy, mathematics, and computing provided for AI gives insights into the principles that still guide AI today.
- **Leverage expert systems for specialized problem solving.** Apply lessons from early AI applications (1970–1985) to implement rule-based AI and machine learning in structured domains, improving efficiency and decision making.
- **Avoid the pitfalls of AI hype.** Recognize that AI has gone through multiple cycles of excitement and disappointment. Ensure that AI investments are guided by technical feasibility, business value, and long-term sustainability.

- **Capitalize on data-driven AI for competitive advantage.** Automate processes, optimize customer experiences, and drive smarter decision making using machine learning and Big Data.
- **Balance innovation with ethical AI adoption.** Carefully consider ethical issues when engaging generative AI's power to boost creativity.
- **Stay agile for the future of AI.** Anticipate the continued evolution of AI in business, ethics, and society. Develop strategies for responsible AI adoption.

KEY TERMS

AI-augmented jobs. Roles in which AI enhances human capabilities by automating routine tasks, improving decision making, and increasing productivity rather than replacing human workers.

AI ethics boards. Committees within organizations or regulatory bodies responsible for overseeing the ethical development and deployment of AI systems, ensuring fairness, transparency, and accountability.

AI fluency. The ability to understand, interpret, and effectively interact with AI technologies, including their capabilities, limitations, and ethical considerations.

AI governance. The framework of policies, procedures, and oversight mechanisms that guide the ethical, transparent, and accountable use of AI technologies within an organization.

AI Winter. A period of reduced funding, research activity, and interest in AI due to unmet expectations or technological limitations.

Artificial general intelligence (AGI). A hypothetical form of AI capable of performing any intellectual task a human can, exhibiting general reasoning and learning abilities.

Bias audit. A systematic evaluation of AI models to identify and mitigate biases in decision making, ensuring fairness and reducing discriminatory outcomes.

Big Data. Huge datasets that require advanced processing techniques to analyze patterns, trends, and insights.

Black box problem. The challenge of understanding and interpreting AI decision making where the reasoning behind predictions or outcomes is unknown.

Cloud computing. The delivery of computing services, such as storage and processing power, over the Internet instead of local servers.

Data centers. Large-scale facilities that house computer systems, storage, and networking equipment to support cloud computing, data processing, and AI applications.

Data mining. The process of discovering patterns, correlations, and insights from large datasets using statistical and AI-driven techniques.

Data warehousing. Storing and managing structured data in centralized repositories for analysis and decision making.

Deep learning. A subset of machine learning that uses multi-layered neural networks to process complex data and make predictions.

Edge computing. A distributed computing approach where data processing occurs closer to the data source, reducing latency and bandwidth usage.

Expert system. An AI program that uses rule-based logic in decision making to emulate human expertise in specific domains.

Fairness-aware machine learning. A field of AI research focused on developing models that mitigate biases and promote equitable outcomes.

Internet of Things (IoT). A network of interconnected physical devices embedded with sensors and software that collect and exchange data.

Knowledge engineers. Specialists who develop AI systems by encoding expert knowledge, rules, and reasoning into machine-readable formats.

Natural language processing (NLP). A branch of AI that enables machines to understand, interpret, and generate human language.

Predictive analytics. Using AI and statistical techniques to forecast future trends and outcomes based on historical data.

Privacy-first policies. Guidelines and frameworks designed to prioritize user data protection and minimize privacy risks when developing and deploying AI systems.

Reinforcement learning. A machine learning approach where AI agents learn by receiving rewards or penalties for actions taken in an environment.

Robotic process automation (RPA). The use of robots or AI-driven automation tools to perform repetitive business processes, such as data entry or invoice processing, without human intervention.

Rule-based reasoning. A method of AI decision making that follows predefined rules and logical statements to arrive at conclusions, commonly used in expert systems.

Sentiment analysis. An AI technique used to determine the emotional tone of text, often applied to social media and customer feedback analysis.

Server farms. Large groups of networked servers in one location that provide high computational power and storage capacity for cloud computing, Big Data analytics, and AI applications.

Symbolic reasoning. An AI approach that uses symbols and logic rules to mimic human reasoning and problem solving.

Turing test. A test proposed by Alan Turing to evaluate a machine's ability to exhibit humanlike intelligence in conversation.

Universal Turing Machine. A theoretical computational model proposed by Alan Turing that can simulate any other computing machine by reading instructions and executing them, forming the foundation of modern computing theory.

CASE STUDY 2.1: AI AND COVID-19

The COVID-19 pandemic was a watershed event for AI and modern society. While AI had been employed for decades in automation and prediction roles in business operations, it had been mainly out of public view. The general public only took notice of AI when it was on display in high-profile events, like when IBM Watson competed against *Jeopardy* champions. However, during the pandemic, AI played a crucial and public role in responding to the crisis, aiding in diagnostics, resource management, and public communication. This shift accelerated AI awareness and adoption in businesses and everyday life.

AI's Role in the COVID-19 Response

Pandemic response efforts, particularly in health care, made use of AI in:

- ***Diagnostics.*** Machine learning algorithms analyzed chest X-rays and CT scans to detect COVID-19 faster, permitting earlier detection and treatment.
- ***Drug discovery.*** DeepMind's AlphaFold provided critical insights into the virus's structure, accelerating the creation of an effective vaccine. AI automation of drug trial paperwork decreased the time needed for FDA approval.
- ***Treatment.*** AI-assisted clinical research identified promising COVID-19 treatments, analyzing vast datasets to predict which drugs might be effective. For example, a study of VA patients showed that patients on an aspirin regimen had decreased mortality if they contracted COVID-19.
- ***Resource allocation.*** Predictive models helped hospitals manage scarce resources, like ICU beds and ventilators.

AI also played a pivotal role in tracking and predicting the spread of the virus:

- ***Pandemic modeling.*** Predictive algorithms forecast infection trends, enabling governments to implement targeted interventions. AI-assisted epidemiological models helped researchers simulate different scenarios and assess the potential impacts of various containment measures.
- ***Contact tracing.*** AI-powered apps like TraceTogether identified potential exposures while balancing privacy concerns. These apps automated the identification and notification of close contacts.
- ***Public communication.*** AI-powered chatbots on CDC and WHO websites provided accurate, real-time information to millions, reducing the load on human-operated helplines. AI also helped combat pandemic-related misinformation, identifying false claims on social media and flagging misleading content.

AI also contributed significantly to pandemic-related logistics and remote work

adaptation:

- ***Supply chain management.*** AI-driven analytics helped distribute vaccines and personal protective equipment (PPE) to where it would do the most good. AI minimized disruptions and improved efficiency by predicting demand surges and identifying bottlenecks.
- ***Workforce management.*** AI-assisted scheduling tools helped businesses adapt to remote work, ensuring operations continued smoothly when employees worked from different locations. AI-driven recruitment and training tools also helped organizations upskill workers in response to shifting skill needs.

The Pandemic's Impact on AI Adoption

The pandemic caused a sudden shift toward isolating people at home to slow its spread. Countries like Argentina, Australia, China, Italy, the United Kingdom, and some states in the United States had stay-at-home orders at different times. AI played a role in how people worked, studied, shopped, interacted, and were entertained. A few of the ways AI had an impact in people's lives were:

- ***Communication.*** Zoom and Microsoft Teams utilized AI for noise reduction, automatic transcription, and virtual backgrounds during videoconferences.
- ***Remote learning and AI in education.*** Remote learning encouraged schools to switch to alternative teaching methods involving AI. These included Zoom and platforms like Coursera and Khan Academy. AI proctoring tools became more common in maintaining academic integrity during remote exams.
- ***Streaming and media consumption.*** With people spending more time at home, AI-powered recommendation engines on platforms like Netflix, Spotify, and YouTube helped them quickly find programs they enjoyed. TikTok and Instagram surged in popularity, and AI provided auto-captioning, content moderation, and video editing.
- ***E-commerce and AI-powered shopping.*** With retail stores closed, online shopping surged. AI product recommendations and chatbots helped the homebound find what they needed. AI-automated the delivery process and allowed buyers to see where their goods were in the delivery process.

By streamlining communication, entertainment, work, and education, AI became a regular part of daily life during the pandemic, increasing awareness of this technology.

The COVID-19 pandemic thus accelerated AI awareness and adoption. AI helped society combat the disease and navigate a socially distanced world. In all the ways discussed, AI became integrated into business operations and daily life. Once operating primarily behind the scenes for prediction and automation, AI moved into public awareness, emphasizing its growing influence on present and future activities.

Discussion Questions

1. What were the most significant ways AI contributed to the COVID-19 response, particularly in health care and epidemiology? How did these applications alleviate strain on health care systems?
2. The pandemic accelerated AI adoption across industries. Which AI-driven innovations introduced during COVID-19 do you think will have lasting impacts on business and society, and why?
3. AI-powered contact tracing and pandemic modeling played a key role in public health management, but they also raised privacy concerns. How should governments and organizations balance the benefits of AI with individual privacy rights in future public health crises?
4. AI-enabled supply chain systems helped manage disruptions in vaccine distribution and medical supply shortages. How can AI-driven logistics be further improved to prepare for future global emergencies?
5. The COVID-19 pandemic highlighted both the strengths and limitations of AI. What lessons should businesses and policymakers take from this period to ensure AI is developed and deployed responsibly in future crises?

Case References

Bhatele, K. R., Jha, A., Tiwari, D., Bhatele, M., Sharma, S., Mithora, M. R., & Singhal, S. (2024). "COVID-19 Detection: A Systematic Review of Machine and Deep Learning-Based Approaches Utilizing Chest X-Rays and CT Scans." *Cognitive Computation,* 16, 1889–1926. **https://doi.org/10.1007/s12559-022-10076-6**

Cho, H., Ippolito, D., & Yu, Y. W. (2020). "Contact Tracing Mobile Apps for COVID-19: Privacy Considerations and Related Trade-Offs." *arXiv.* **https://arxiv.org/abs/2003.11511**

Heaven, W. D. (2020, November 30). "DeepMind's Protein-Folding AI Has Solved a 50-Year-Old Grand Challenge of Biology." *MIT Technology Review.* **https://www.technologyreview.com/2020/11/30/1012712/deepmind-protein-folding-ai-solved-biology-science-drugs-disease/**

Non, L. R., Marra, A. R., & Ince, D. (2025). "Rise of the Machines—Artificial Intelligence in Healthcare Epidemiology." *Current Infectious Disease Reports, 27*, Article 4. **https://doi.org/10.1007/s11908-024-00854-8**

Osborne, T. F., Veigulis, Z. P., Arreola, D. M., Mahajan, S. M., Röösli, E., & Curtin, C. M. (2021). "Association of Mortality and Aspirin Prescription for COVID-19 Patients at the Veterans Health Administration." *PLOS ONE,* 16(2), e0246825. **https://doi.org/10.1371/journal.pone.0246825**

Paspel, E. (2025, March 6). "AI in Healthcare Supply Chain Management: Efficiency and

Accuracy." *Klarity Health Library.* **https://my.klarity.health/ai-in-healthcare-supply-chain-management-efficiency-and-accuracy/**

Pfizer. (n.d.). "mRNA and Artificial Intelligence for Advanced Vaccine Innovation." *Pfizer News.* **https://www.pfizer.com/news/articles/how_a_novel_incubation_sandbox_helped_speed_up_data_analysis_in_pfizer_s_covid_19_vaccine_trial**

White, D. B., & Lo, B. (2020). "A Framework for Rationing Ventilators and Critical Care Beds during the COVID-19 Pandemic." *JAMA,* 323(18), 1773–1774. **https://doi.org/10.1001/jama.2020.5046**

World Health Organization. (2025, March 6). "Chatbots against COVID-19: Using Chatbots to Answer Questions on COVID-19." *World Health Organization.* **https://www.who.int/news-room/feature-stories/detail/scicom-compilation-chatbot**

CASE STUDY 2.2: PROJECT INNEREYE—EXPANDING AI IN MEDICINE

Analyzing medical images like CT scans, MRIs, and X-rays is crucial in medical diagnostics. However, this field faces significant challenges:

- ***High demand and limited resources.*** The demand for radiological services often outstrips qualified radiologists' availability, leading to diagnosis delays. Patients can experience long treatment delays, sometimes days or weeks, while waiting for image analysis.
- ***Workload and accuracy.*** Radiologists handle many complex cases concerning various ailments daily. The high volume and complexity of cases can sometimes lead to diagnostic errors.

Project InnerEye, an initiative by Microsoft Health Futures, proposes solving these challenges by harnessing deep learning to analyze three-dimensional medical images. The initiative developed a system called *OSAIRIS*, the first cloud-based AI technology developed and deployed in the UK's National Health Service (NHS). The goal of ORAIRIS is to go beyond just diagnostics to treatment.

Before cancer patients can undergo radiation treatment, the size, shape, and location of the tumor must be precisely outlined to direct the radiation correctly. OSAIRIS is trained to create outlines of tumors for treatment. It automates the traditionally labor-intensive image analysis process, reducing the time required to assess medical scans. OSAIRIS does not replace the radiologist; instead, it acts as an intelligent assistant, performing the initial scan analysis, flagging areas of concern, and outlining the tumor, speeding the analysis by the human expert.

Raj Jena, oncologist at Cambridge University Hospitals NHS Foundation Trust, led the research. He said, "OSAIRIS does much of the work in the background so that when the doctor sits down to start planning treatment, most of the heavy lifting is done." Jena described how the AI could accurately identify the edges of the tumor for radiation treatment, saving the oncologist up to 3 hours.

A peer-reviewed study published in the *Journal of the American Medical Association (JAMA) Network Open* found that OSAIRIS can augment and accelerate clinicians' ability to perform radiotherapy planning up to 13 times faster while maintaining an accuracy level within the bounds of human expert variability. The program also enhances diagnostic precision, assisting radiologists in accurately identifying and delineating tumors and other medical conditions. By improving both speed and accuracy, deep learning-based tools like OSAIRIS are reducing healthcare professionals' workload and improving patient outcomes.

Researchers found that Project InnerEye yielded the following benefits:

- ***Reduced wait times.*** OSAIRIS substantially reduced patient wait times by accelerating the analysis process, enabling quicker diagnosis and treatment.

- ***Support for radiologists.*** The project supported radiologists by handling routine analyses, allowing them to focus on more complex cases and reduce the risk of diagnostic errors.
- ***Democratizing AI in health care.*** Project InnerEye's tools are available as open-source software under an MIT license, enabling other hospitals and research institutions to develop their own AI models, making this advanced technology accessible to a broader range of medical professionals.

OSAIRIS is a significant advancement in medical imaging, reducing diagnostic and treatment times. Its cloud-based design makes this technology accessible to smaller hospitals without requiring large technology investments. As this case shows, AI has a home in easing health care professionals' workloads and enhancing patient care.

Discussion Questions

1. Discuss how the implementation of OSAIRIS could transform the day-to-day operations in a hospital's radiology department. What changes might it bring to the workflow of radiologists?
2. What are the ethical implications of using AI for medical diagnoses? How can health care providers ensure that the use of such technology complements, rather than replaces, the expertise of medical professionals?
3. How does the integration of AI, as seen in the OSAIRIS project, align with broader trends in health care technology? What other areas of health care might benefit from similar AI-driven innovations?
4. Discuss the potential impact of reduced diagnostic wait times on patient outcomes. How might quicker access to results affect patient treatment plans and recovery?
5. What challenges might hospitals and health care systems face in adopting AI technologies like OSAIRIS? How could these challenges be addressed?

Case References

Cambridge University Hospitals NHS FT. (2023, July 4). "AI Tool for Radiotherapy." *HTN.* **https://htn.co.uk/2023/07/04/cambridge-university-hospitals-highlight-ai-tool-for-preparing-for-scans/**

Microsoft Research. (2025). "Project InnerEye—Democratizing Medical Imaging AI." *Microsoft.* **https://www.microsoft.com/en-us/research/project/medical-image-analysis/?msockid=22fdd9d69a8164f508cecd3e9bee65f9**

Microsoft Research. (2025). "Project InnerEye Open-Source Software for Medical Imaging AI." *Microsoft.* **https://www.microsoft.com/en-us/research/project/project-innereye-open-source-software-for-medical-imaging-ai/**

Oktay, O., Schwaighofer, A., Carter, D., Bristow, M., Alvarez-Valle, J., & Nori, A. (2020, November 30). "Project InnerEye Evaluation Shows How AI Can Augment and Accelerate

Clinicians' Ability to Perform Radiotherapy Planning 13 Times Faster." *Microsoft Research Blog.* **https://www.microsoft.com/en-us/research/blog/project-innereye-evaluation-shows-how-ai-can-augment-and-accelerate-clinicians-ability-to-perform-radiotherapy-planning-13-times-faster/**

Oktay, O., et al. (2020, November 30). "Evaluation of Deep Learning to Augment Image-Guided Radiotherapy for Head and Neck and Prostate Cancers." *JAMA Network Open.* **https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2773292**

CASE STUDY 2.3: PIONEERING WOMEN'S SPORTS ANALYTICS WITH AI

The Women's National Basketball Association's (WNBA's) 2024 adoption of AI through its partnership with Genius Sports and Second Spectrum marks a significant milestone in the application of AI in sports. From its origins in data-driven decision making to its current state of advanced machine learning and generative models, the WNBA's use of AI exemplifies the historical progression of AI technologies. This case highlights how decades of AI research and development have paved the way for cutting-edge innovations in sports analytics.

AI's Journey in Sports Analytics

AI's integration into sports has evolved from basic statistical analyses to advanced tools capable of real-time decision making and predictive insights. Early sports analytics relied on manual data collection and rule-based systems to analyze performance. Over the years, advancements in neural networks, machine learning, and computer vision have enabled AI systems like Second Spectrum's optical tracking to revolutionize how sports data is collected, analyzed, and utilized.

Advanced AI analytics in sports originated in men's leagues, where its transformative potential was first realized. Second Spectrum initially became the official tracking provider for leagues such as the National Basketball Association (NBA) in 2017 and Major League Soccer (MLS) and the English Premier League (EPL), both in 2020. These leagues used AI-driven tracking systems to analyze player positioning, team formations, and game dynamics. The insights gained have enhanced performance and strategy and revolutionized fan engagement with features like predictive shot success rates and dynamic heatmaps. The success of AI in these leagues provided a proven blueprint for its adoption in the WNBA, marking a step toward closing the technological use gap between men's and women's sports.

A Historical First for Women's Sports

In 2024, the WNBA became the first U.S. women's professional sports league to implement league-wide 3D tracking technology. Camera arrays can identify players and who they are defending to create images which tracks player and ball movement. This data is used to fuel the AI that analyzes the game. Using cameras to generate real-time data on player movements and ball trajectories, the system reflects a culmination of AI's progress in processing and analyzing complex, dynamic environments. This technological leap mirrors AI's broader historical trajectory, moving from theoretical concepts to practical applications that enhance decision-making across industries.

The following are the AI technologies used:

- ***Neural networks.*** Second Spectrum's system uses neural networks to process spatial data, tracking player and ball movements in real time. This builds on decades of research into AI's ability to model nonlinear relationships.
- ***Machine learning.*** Predictive algorithms analyze patterns in gameplay to forecast team strategies and player tendencies, showcasing how machine learning has matured to generate actionable insights.
- ***Generative AI.*** By simulating optimizing lineups and plays, generative AI highlights the evolution of AI from static analysis to creative problem solving.

Impact on the Game

The WNBA's integration of AI tools enables coaches to leverage real-time analytics for strategic adjustments while players benefit from performance optimization and injury prevention. Using data from Second Spectrum, WNBA broadcasters overlay advanced graphics and present predictive insights during live games, such as real-time player statistics and predictive shot outcomes, creating an engaging viewing experience. Sports Genius offers data information services to betters, allowing them to make more informed wagers. These developments reflect the growing sophistication of AI applications, a far cry from the early days of statistical analysis in sports.

The WNBA's partnership with Genius Sports and Second Spectrum exemplifies how the historical evolution of AI has influenced modern sports analytics. By embracing AI technologies that draw on decades of research, the WNBA enhances the game and signals innovation in other women's sports. As AI advances, the WNBA's adoption of league-wide tracking sets a precedent for integrating emerging technologies into sports. Future innovations may include predictive injury models and personalized fan experiences, underscoring AI's expanding role in shaping the sports industry.

Discussion Questions

1. What challenges or opportunities might arise as sports increasingly integrate AI-driven technologies?
2. The WNBA's use of Second Spectrum's tracking technology has enhanced player performance analysis and fan engagement. In what ways do you think AI could further revolutionize the game, and what ethical or competitive concerns might emerge?
3. Generative AI can now be used to simulate optimal lineups and plays. How might this affect coaching decisions and the role of human intuition in sports strategy? Do you think AI-generated recommendations should have limitations in competitive sports?
4. Sports betting companies use AI-powered analytics to create enticing betting opportunities.

Betters use the data and insights generated to place bets. What ethical considerations arise from AI's role in betting, and how might leagues ensure the integrity of the game while allowing for technological advancements?

5. AI in sports analytics has transformed how fans experience games through predictive insights and augmented broadcasts. How might future advancements in AI further personalize the fan experience, and what potential risks (such as data privacy or over-reliance on AI-driven narratives) should be considered?

Case References

Dowsett, B. (2018, June 28). "Light, Camera, Revolution." *FanSided.* **https://grantland.com/features/the-toronto-raptors-sportvu-cameras-nba-analytical-revolution/**

NBA Communications. (2024). "WNBA and Genius Sports Bring Cutting-Edge Tracking Data to Every WNBA Arena." *NBA.com.* **https://pr.nba.com/wnba-genius-sports-tracking-data/**

ENDNOTES

[1] Turing, A. M. (1950). "Computing Machinery and Intelligence." *Mind*, 59(236), 433–460.

[2] Cheng, X., Kosinski, M., & Stillwell, D. (2024). A Turing test of whether AI chatbots are behaviorally similar to humans. *Proceedings of the National Academy of Sciences*, 121(9), e2313925121. **https://doi.org/10.1073/pnas.2313925121**

[3] McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence."

[4] Newell, A., & Simon, H. A. (1956). "The Logic Theory Machine—A Complex Information Processing System." *IRE Transactions on Information Theory*, 2(3), 61–79.

[5] Weizenbaum, J. (1966). "ELIZA—A Computer Program for the Study of Natural Language Communication between Man and Machine." *Communications of the ACM*, 9(1), 36–45.

[6] Shortliffe, E. H. (1976). *Computer-Based Medical Consultations: MYCIN*. Elsevier.

[7] Khandelwal, K., Patel, S., Patel, J., & Pnachal, M. (2023). "A Study to Know —Use of AI for Personalized Recommendation, Streaming Optimization, and Original Content Production at Netflix." *International Journal of Scientific Research in Engineering and Technology* 9(6): 435–445. **https://ijsret.com/wp-content/uploads/2023/11/IJSRET_V9_issue6_435.pdf**

[8] H2O.ai. (2020). "Driving Away Fraudsters at PayPal." **https://h2o.ai/content/dam/h2o/en/marketing/documents/2020/01/PayPal-Customer-Case-Study-rnd2-1.pdf**

[9] Jiang, Fei, Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., & Wang, Y. (2017). "Artificial Intelligence in Healthcare: Past, Present and Future." *Stroke and Vascular Neurology* 2(4): 230–243; "Artificial Intelligence and Healthcare: A Journey through History." *Journal of Medical Internet Research* 25(1) (2023): e11122160.

[10] Strickland, Eliza. "How IBM Watson Overpromised and Underdelivered on AI Health Care." *IEEE Spectrum*, April 2, 2019; "Confronting the Criticisms Facing Watson for Oncology." *The ASCO Post*, September 10, 2019.

[11] DeepSeek-AI. "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning." *GitHub*. **https://github.com/deepseek-ai/DeepSeek-R1**; Tau, Byron. "Researchers Link DeepSeek's Blockbuster Chatbot to Chinese Telecom Banned from Doing Business in US." *AP News*, February 5, 2025. **https://apnews.com/article/deepseek-china-generative-ai-internet-security-concerns-c52562f8c4760a81c4f76bc5fbdebad0**

[12] McKinsey & Company. "The State of AI in Early 2024." June 2024. **https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai**.

[13] International Telecommunication Union. *ITU Internet Reports 2005: The Internet of Things*. 7th

ed. Geneva: International Telecommunication Union, 2005.

[14] Greenberg, Sarah. “Birth and Evolution of Internet of Things (IoT).” *BCC Research*, July 30, 2021. **https://blog.bccresearch.com/birth-and-evolution-of-internet-of-things-iot**

[15] Matellio Inc. “IoT for Smart Cities: Enabling Smarter & Connected Solutions.” January 7, 2025. **https://www.matellio.com/blog/iot-for-smart-cities/**

[16] Quy, T. E., Roy, A., Iosifidis, V., Zhang, W., & Ntoutsi, E. “A Survey on Datasets for Fairness-Aware Machine Learning.” January 21, 2022. **https://arxiv.org/abs/2110.00530**

[17] SciDev.Net. “AI Risks in Healthcare: Misdiagnosis, Inequality, and Ethical Concerns.” *News-Medical.Net*, January 23, 2024. **https://www.news-medical.net/news/20240123/AI-risks-in-healthcare-Misdiagnosis-inequality-and-ethical-concerns.aspx**

[18] Federal Aviation Administration. *Summary of the FAA’s Review of the Boeing 737 MAX*. November 18, 2020. **https://www.faa.gov/sites/faa.gov/files/2022-08/737_RTS_Summary.pdf**

[19] Columbus, Louis. “Defending SOCs Under Siege: Battling Adversarial AI Attacks.” *VentureBeat*, December 9, 2024. https://venturebeat.com/security/defending-socs-under-siege-battling-adversarial-ai-attacks/.

[20] Brumfiel, Geoff. “How AI is Revolutionizing How Governments Conduct Surveillance.” *NPR*, June 13, 2023. **https://www.npr.org/2023/06/13/1181868277/how-ai-is-revolutionizing-how-governments-conduct-surveillance**

[21] TIME. “How OpenAI’s Sam Altman Is Thinking About AGI and Superintelligence in 2025.” January 8, 2025. **https://time.com/7205596/sam-altman-superintelligence-agi/**

CHAPTER 3
Data Management for AI



kanawatth/123RF

Learning Objectives

After studying Chapter 3, you should be able to

- **LO3-1** Explain the significance of data structure and types in enabling practical AI applications, and list sources of data for AI.
- **LO3-2** Describe the principles and strategies for ensuring data security.
- **LO3-3** Analyze the role of data quality and integrity in achieving reliable AI outcomes.
- **LO3-4** Identify the impact of regulations on AI data management practices.
- **LO3-5** Examine the challenges and innovations in Big Data management.
- **LO3-6** Evaluate emerging technologies in securing and managing AI data.
- **LO3-7** Describe ethical considerations in AI data collection and use.

INTRODUCTION

Managing large, diverse, secure data sources is essential for training and implementing AI algorithms. Relational databases in the 1970s and data warehousing in the 1990s enabled businesses to organize and analyze vast datasets. Big Data technologies in the early 2000s revolutionized the processing of massive unstructured datasets, paving the way for significant machine learning breakthroughs. As our historical review illustrates, the growth of AI has always been intertwined with advancements in data management.

Today, managing data for AI extends far beyond handling sheer volume. Modern enterprises face complex challenges, from securely ensuring the quality of data collected from diverse sources to generating real-time insights while navigating privacy regulations and ethical considerations. This chapter examines these contemporary challenges and offers strategies for businesses to develop robust data management practices, unlocking AI's transformative potential while mitigating risks.

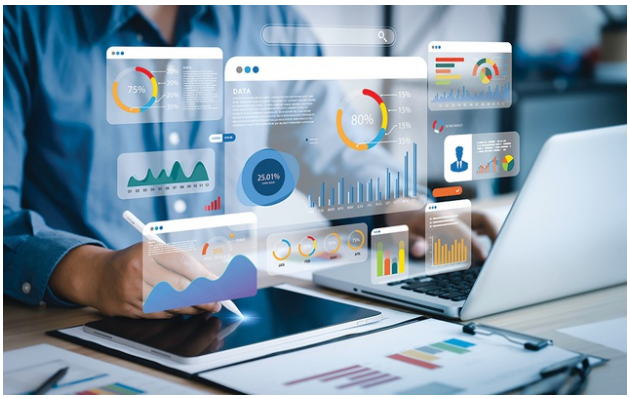
DATA FOR AI (LO3-1)

The success of AI systems hinges on the quality and diversity of the data they process. Data powers AI's ability to learn, predict, and innovate, from numerical figures in spreadsheets to unstructured multimedia.

Data Structure and Types

AI's sources and types of data have grown through time. Early datasets were primarily **structured data**—data organized in a fixed format like a spreadsheet or database, making it easily searchable and understandable by algorithms. Structured data often includes sales figures, customer information, and transaction records. These are typically stored in relational databases, systems that hold data in tables. This table-based format allows for efficient organization, storage, and retrieval of data, where relationships among different items can be quickly established and maintained, facilitating complex queries and analysis.

The advent of Big Data in the 2000s transformed the landscape of data management and AI. This era introduced vast volumes of structured and **unstructured data** from sources such as Internet of Things (IoT) devices and social media platforms. Unlike structured data, unstructured data lacks a predefined format. It includes text, images, videos, and social media posts. AI technologies, especially those powered by natural language processing and computer vision, can extract meaningful insights from these diverse data types for decision making and innovation.



Deemerwha studio/Shutterstock

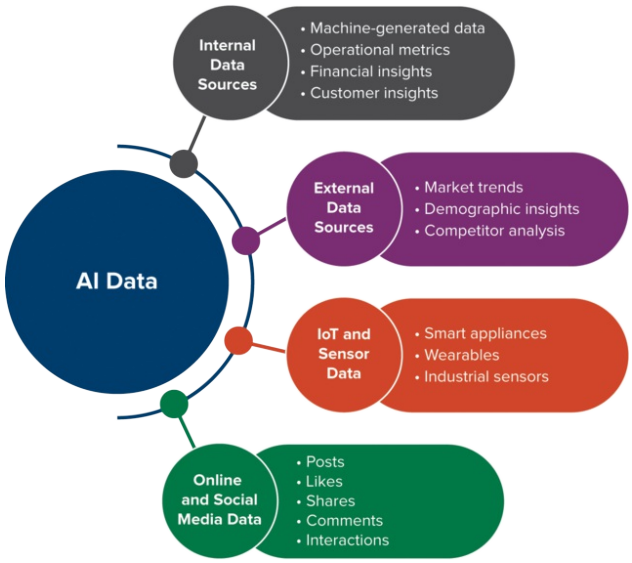
Unstructured data comes from many sources.

Sources of Data for AI

AI systems rely on diverse and complex data sources to function effectively. These sources can be broadly categorized into internal data sources, external data sources, IoT and sensor data, and

online and social media data (see Figure 3.1). Each of these sources offers unique opportunities and challenges.

FIGURE 3.1 AI Data Sources



Internal Data Sources

Internal data gathered within an organization includes:

- Machine-generated data, which is automatically collected by enterprise systems, sensors, or software logs.
- Operational metrics, such as production efficiency, inventory levels, and employee performance indicators.
- Financial records, including revenue, expenses, and profit margins.
- Customer insights obtained from customer relationship management (CRM) systems, which track customer interactions, purchasing behavior, and preferences.

This wealth of internal data is essential for evaluating business performance, allowing companies to measure profitability, efficiency, and customer satisfaction. Additionally, it helps optimize operations by identifying inefficiencies, streamlining workflows, and improving resource allocation. By analyzing these insights, organizations can uncover strengths and weaknesses, enabling them to refine their strategies and enhance overall decision making.

Data silos—where data is stored in disconnected departments making integration difficult—are a major challenge for companies with many internal data sources. Inconsistent data quality is another significant issue because internal datasets may contain errors, missing values, or

duplications, which can affect AI model accuracy. Additionally, access and privacy restrictions required by compliance with various regulations can limit the availability of sensitive corporate data. Many businesses also struggle with legacy system compatibility, where outdated infrastructures lack interoperability with modern AI-driven solutions. Finally, internal data may carry historical biases, leading to skewed AI model outputs if not carefully curated and audited.

External Data Sources

Businesses rely on external data, such as market trends, demographic insights, and competitor analysis, to navigate their industry landscape. Market trends track shifts in consumer behavior and demand, while demographic data—age, income, education, location, etc.—helps tailor products and marketing strategies. Competitor analysis evaluates pricing, product offerings, and customer sentiment to identify competitive advantages. These datasets come from public sources like government reports and private firms specializing in market research.

While external data offers valuable insights, it also presents challenges. Credibility is a concern because some sources may be biased or outdated. Costs can be prohibitive, especially for frequent updates. Regulatory compliance adds complexity, and generic data often requires extensive refinement to be useful. Additionally, the sheer volume of information can lead to analysis paralysis without proper filtering.

IoT and Sensor Data

Internet of Things (IoT) devices and sensors generate vast amounts of data in real time. This data includes:

- Smart appliances reporting energy usage.
- Wearables that track health metrics.
- Industrial sensors that monitor machine performance and predict breakdowns.

This real-time data enhances efficiency, streamlines operations, and improves user experiences. However, managing IoT and sensor data comes with challenges. The massive data volume requires scalable storage, processing, and analytics solutions like cloud and edge computing. **Edge computing** is a distributed computing system that processes data closer to the source (e.g., IoT devices or local servers) rather than relying on centralized cloud infrastructure, reducing delays and bandwidth usage. Connectivity issues can lead to transmission delays or incomplete data, especially in remote locations. Standardization remains a hurdle as manufacturers use different data formats and protocols. Security risks are also significant, with vulnerable networks susceptible to cyberattacks. Additionally, power constraints in battery-operated IoT devices demand energy-efficient AI models that ensure real-time responsiveness.

Online and Social Media Data

Social media platforms generate vast user data—posts, likes, shares, comments, and interactions—which AI can analyze to reveal consumer preferences, market trends, and brand perception. Businesses leverage this data to track emerging trends, measure audience sentiment, and refine marketing strategies in real time. AI-powered analytics uncover hidden behavioral patterns, enabling personalized advertising, enhanced customer engagement, and predictive insights into future purchasing decisions.

However, analyzing social media data comes with challenges. Ensuring data quality requires filtering misinformation, removing noise, and addressing biases in user-generated content. Privacy concerns are significant because businesses must comply with data protection laws while maintaining consumer trust through ethical AI practices and transparent policies.

Beyond social media, online data from customer reviews, search queries, and influencer activity provides valuable insights but introduces further complexities. Misinformation, bot-generated content, and biased opinions make extracting reliable insights difficult. AI also struggles with sentiment misinterpretation, particularly sarcasm, slang, and cultural nuances. Additionally, platform restrictions and Application Programming Interface (API) limitations increasingly restrict data access, while representation bias in social media datasets can distort AI predictions by overemphasizing vocal minorities. To navigate these challenges, businesses must adopt rigorous data validation, ethical AI frameworks, and adaptable analysis methods.

The variety and complexity of data sources available today provide fertile ground for AI to take root and thrive. AI systems' ability to process and make sense of this data for strategic implementation makes data a resource and a core asset in the modern business landscape. YouTube, eBay, Facebook, TikTok, Noom, and other companies like them have used data strategically to grow their businesses.

DATA SECURITY (LO3-2)

In February 2024, Change Healthcare, a leading provider in the U.S. health care industry, became the target of a devastating ransomware attack. An employee had not enabled multifactor authentication on their account, and the hackers were able to access the system using stolen credentials. This breach compromised the sensitive health and personal information of more than 190 million patients, exposing critical vulnerabilities in data security. The attack forced the company to pay a $22 million ransom to mitigate the damage.[2]

The incident underscored the need for robust cybersecurity measures in sectors reliant on digital data and AI technologies. Beyond the financial impact, breaches erode customer trust and inflict long-term reputational harm, highlighting the far-reaching consequences of inadequate data protection strategies.

Data Security Challenges

Hacking is often the first thing that comes to mind when discussing data security, and incidents like the Change Healthcare ransomware attack make headlines. However, data security for AI encompasses much more. It involves safeguarding data during transmission and storage, implementing strict access controls, protecting against loss due to disasters, and addressing an array of cybersecurity threats that could compromise the integrity of AI systems.

Data Security during Transmission and Storage

AI systems often rely on integrating vast amounts of data from diverse sources. Devices like smart sensors and wearables continuously generate data that must be securely transmitted to storage systems. Many organizations utilize cloud services for data storage and processing, which, while convenient, introduce potential security vulnerabilities if not properly safeguarded.

Encryption is necessary for securing data during transmission and storage. Encoding data into an unreadable format accessible only with the proper decryption key protects sensitive information from unauthorized access, even if the data is intercepted. This ensures that both transmitted and stored data remain confidential and secure, safeguarding the integrity of AI systems.

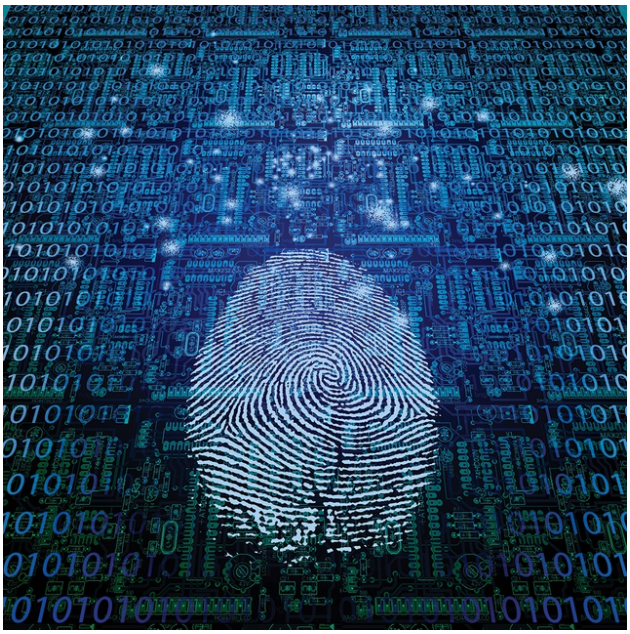
Access Control

While encryption safeguards data during transmission and storage, it's equally important to ensure that access to data is tightly controlled. Access control measures further minimize the risk of unauthorized use and breaches, protecting sensitive information from within.

Controlling access to data is critical to reducing the risk of accidental or intentional

misuse. The principle of **least privilege** ensures that users, applications, and systems are granted only the minimum access necessary to perform their tasks, minimizing the risk of unauthorized actions or breaches. **Role-based access control** further enhances security by allowing employees to access only the relevant data to their roles, ensuring operational efficiency and data protection. Effective access control involves managing identities and credentials for authorized users, services, and hardware; ensuring proper authentication; and restricting access based on specific needs.

Amazon Web Services (AWS) uses **Identity and Access Management (IAM)** to enforce role-based access control, ensuring users and applications only have the permissions they need. For example, in a financial services company using AWS, the finance team may have access to billing reports and cost analysis tools. In contrast, the IT team manages server configurations and security settings. AWS IAM enables administrators to define roles and policies for members of each team, such as restricting database access to only authorized analysts or allowing developers to deploy but not delete infrastructure. The least privilege model enhances security by preventing unauthorized access while maintaining operational efficiency.[3]



Bruce Rolff/Shutterstock

Access control ensures that only legitimate users can access sensitive data.

Backup and Disaster Recovery

No system is immune to failure or attack despite strong encryption and access controls. That is why robust backup and disaster recovery plans are critical to ensuring data remains accessible and secure during natural disasters or cyberattacks.

For example, in 2021, Ochsner Health, a nonprofit health system operating 46 hospitals and more than 370 health and urgent care centers along the Gulf Coast, was hit hard by Hurricane Ida. The

Category 4 storm, with winds reaching 155 miles per hour, caused widespread roof and water damage, power outages, generator failures, and severe flooding. This disrupted patient data access and electronic medical record (EMR) systems. However, luckily for Ochsner Health, it had developed a comprehensive disaster recovery plan to address its hospitals' loss of life and critical patient data during Hurricane Katrina. It moved emergency backup systems well above the high-water mark from Katrina. It also implemented cloud-based backups, allowing data to be secured outside the hurricane's impact zone. These actions allowed it to restore essential data access within days after Hurricane Ida, ensuring the continuity of patient care operations.[4]

This example highlights the importance of robust backup and recovery strategies to safeguard against data loss from cyberattacks, natural disasters, or system failures. Regular backups and well-structured contingency plans enable organizations to recover quickly and minimize disruptions, ensuring resilience in the face of unforeseen events.



mTaira/Shutterstock

Backup and disaster recovery plans are critical to ensuring data remains accessible and secure during natural disasters.

Protecting against Cybersecurity Threats

While backup and recovery strategies mitigate the impact of data loss, businesses must also guard against cyberattacks. They pose significant risks to organizations. These threats range from data breaches, ransomware attacks, and phishing scams to sophisticated adversarial attacks designed to manipulate AI algorithms with malicious data.

For example, a ransomware attack on Colonial Pipeline in 2021 prompted corporate officials to proactively shut down their pipeline that provided 45 percent of the fuel supplies across the eastern United States. The pipeline was shut down for 5 days, causing canceled flights, lines at gas stations, and higher prices, showing how cyberattacks can have an impact beyond the targeted organization.[5] These incidents compromise data integrity and availability, erode consumer trust, and can lead to severe financial and reputational damage. **Table 3.1** describes some common cybersecurity threats. These threats highlight the importance of adopting robust principles to safeguard data in AI systems.

Table 3.1 **Common Cybersecurity Threats**

Threat	Description
Advanced persistent threats (APTs)	Prolonged and stealthy attacks targeting sensitive data
Insider threats	Risks posed by employees, whether intentional or accidental
IoT vulnerabilities	Insecure connected devices as entry points for cyberattacks
Ransomware and phishing	Techniques to lock systems or deceive users into revealing sensitive information
Supply chain attacks	Exploiting vulnerabilities in third-party vendors to infiltrate larger systems
Zero-day exploits	Vulnerabilities exploited before they are known or patched

AI Data Security Principles

Addressing access control, disaster recovery, and cybersecurity threats requires a strategic framework combining technology, processes, and people. The 2024 National Institute of Standards and Technology (NIST) Cybersecurity Framework (CSF) 2.0[6] provides the principles shown in Figure 3.2 for safeguarding AI data systems.

FIGURE 3.2 2024 NIST Principles for Safeguarding Data Systems

1. GOVERNANCE AND RISK MANAGEMENT

- ***Policies***: Develop and enforce policies for managing cybersecurity risks.
- ***Roles and Responsibilities***: Clearly define and communicate cybersecurity roles, responsibilities, and authorities to foster accountability and continuous improvement.
- ***Supply Chain Risk Management***: Integrate cybersecurity supply chain risk management into overall risk management processes and ensure continuous monitoring.

2. ASSET MANAGEMENT

- ***Inventory and Classification***: Maintain inventories of hardware, software, services, and data, prioritizing them based on classification, criticality, and impact on the mission.
- ***Lifecycle Management***: Manage assets throughout their lifecycle, from acquisition to disposal, ensuring security at each stage.

3. DATA PROTECTION

- ***Confidentiality, Integrity, and Availability***: Protect sensitive data at rest and in transit through encryption to ensure its confidentiality, integrity, and availability.

- ***Vulnerability Audits*:** Promote regular security audits to identify vulnerabilities.
- ***Backup and Recovery***: Regularly create, protect, maintain, and test backups to ensure data can be restored in case of loss or corruption.

4. ACCESS CONTROL

- ***Identity Management***: Manage identities and credentials for authorized users, services, and hardware, ensuring proper authentication and access control.
- ***Least Privilege***: Grant users, applications, and systems the minimum access necessary to perform their specific tasks.

5. CONTINUOUS MONITORING AND INCIDENT RESPONSE

- ***Monitoring***: Monitor networks, environments, and activities to detect anomalies and potential cybersecurity incidents.
- ***Incident Management***: Develop and execute incident response plans, including analysis, mitigation, and communication, to effectively manage and recover from cybersecurity incidents.

6. AWARENESS AND TRAINING

- ***General Awareness***: Provide cybersecurity awareness and training to all personnel to ensure they understand and can mitigate cybersecurity risks.
- ***Specialized Training***: Offer targeted training for individuals in specialized roles to equip them with the necessary skills to handle specific cybersecurity tasks.

Source: Based on the 2024 National Institute of Standards and Technology (NIST) Cybersecurity Framework (CSF) 2.0

A data security plan is necessary for effective and ethical AI deployment. As illustrated by real-world examples like the Change Healthcare ransomware attack and Ochsner Health's recovery from Hurricane Ida, data security can have far-reaching consequences, affecting financial stability, operational continuity, and public trust.

Page 49

To counter these threats, companies should follow the principles provided by NIST in **Figure 3.2** and adopt a proactive, multilayered approach using the techniques in **Table 3.2**. Fostering a culture of cybersecurity awareness through employee training mitigates human errors, such as falling for phishing attacks.

Table 3.2 **Cybersecurity Responses**

Response	Description
Endpoint detection and response (EDR)	Tracking and responding to threats at individual devices
Firewalls	Establishing barriers between internal networks and external threats
Intrusion detection and	Monitoring network traffic for anomalies and blocking

prevention systems (IDS/IPS)	unauthorized activity
Multifactor authentication	Requiring users to verify their identity using multiple factors before accessing an account or system
Phishing detection and training	Educating employees and deploying tools to spot phishing attempts
Regular vulnerability assessments	Identifying, analyzing, and prioritizing security weaknesses
Zero trust models	Verifying users and devices continuously before granting access

A proactive approach to cybersecurity is not just a best practice—it's a necessity. Implementing robust security measures, from continuous monitoring and access control to employee training and incident response, can prevent minor vulnerabilities from escalating into full-scale crises. An ounce of prevention—through strategic planning and multilayered defenses—can save companies from catastrophic data breaches, financial losses, and reputational damage. Investing in cybersecurity today is not just about protection; it's about ensuring resilience, operational continuity, and long-term trust in an AI-driven world.

DATA QUALITY AND INTEGRITY (LO3-3)

As Steve Lohr aptly observed in a 2014 *New York Times* article, "Data scientists, according to interviews and expert estimates, spend 50 to 80 percent of their time mired in [the] mundane labor of collecting and preparing unruly digital data, before it can be explored for useful nuggets."[7] (Today estimates are 70 to 90 percent.[8]) This quote identifies a fundamental truth in AI and data analytics that we introduced in **Chapter 1**: The success of AI-driven insights begins with the quality of the data it processes. Poor-quality data, whether incomplete, inconsistent, or biased, can compromise the reliability and value of AI models, leading to flawed conclusions and missed opportunities.

The quality and integrity of AI systems' underlying data are paramount for reliable insights. Poor-quality data can lead to flawed analyses and misguided decisions, impacting business outcomes. As the adage goes, "garbage in, garbage out." The important question becomes, "How can we ensure that we have good quality data?"

High Quality Data: Characteristics and Methods

What makes for high-quality data? The following are some characteristics as well as methods to ensuring data quality:

1. ***High-quality data is accurate and reliable.*** To ensure accuracy and reliability of data:
   - Verify and validate data to ensure it accurately reflects reality.
   - Allocate resources for data preparation (often time-intensive) to ensure that data is ready for analysis, enhancing productivity and outcomes.
2. ***High-quality data is consistent across data sources.*** To ensure consistency of data:
   - Create **standardization** protocols for data collection, storage, and reporting to ensure uniformity across departments and systems.
   - Automate validation tools to identify and resolve inconsistencies like mismatched formats or categories early in the data pipeline.
   - Conduct regular cross-departmental reviews to ensure adherence to established standards, fostering a cohesive and reliable data environment for AI applications.
3. ***High-quality data is free from bias.*** To detect and mitigate bias in data:
   - Implement regular audits to identify and measure biases in datasets.
   - Utilize bias detection tools and techniques, such as fairness metrics and algorithmic audits, to uncover hidden patterns that could lead to discriminatory results. A **fairness metric** is a quantitative measure used to evaluate whether an AI model treats different groups or individuals equitably.

- Establish diverse and representative datasets by incorporating data from multiple sources and perspectives to minimize the risk of systemic bias.
- Foster a culture of accountability by providing training for teams on recognizing and addressing bias throughout the data management process.

4. ***High-quality data is complete.*** Missing data can distort results or render analyses incomplete. To ensure data completeness:
   - Implement strategies to handle missing data through imputation methods or adjustments to analytical models. **Imputation** is a method for filling in missing data using statistical estimates or predictive modeling to maintain dataset integrity.

Data Cleaning

Ensuring high-quality data in AI systems begins with a structured **data cleaning** process. This process involves several key steps, each building upon the previous one to systematically improve data integrity:

1. ***Data profiling: Identifying issues early.*** The first step in cleaning data is profiling—analyzing datasets to detect anomalies, inconsistencies, duplicates, and missing values. This helps organizations understand the overall structure of the data and pinpoint problem areas.
2. ***Deduplication: Eliminating redundant entries.*** Once inconsistencies are identified, the next step is **deduplication**, which removes redundant or conflicting records that could skew AI models. This step is crucial for maintaining accuracy and ensuring that insights are not distorted by repetitive data points.
3. ***Standardization: Enforcing consistency across sources.*** After the duplicates are eliminated, ensure that remaining data follows consistent formats, units, and naming conventions across different sources. This is essential for integrating data from multiple systems, avoiding mismatches in AI analysis.
4. ***Handling missing values: Ensuring data completeness.*** Missing data affects AI predictions. Remove the incomplete record or use imputation techniques to fill in missing values with statistical estimates.
5. ***Bias detection and mitigation: Promoting fairness and representation.*** The final step is identifying and addressing biases. This involves auditing datasets for imbalances in represented groups, applying bias detection tools, and ensuring that training data is diverse and representative.

Building a Data-Ready Organization

Achieving data quality and integrity is not solely a technical challenge; it's an organizational one. **Data governance** is the framework of policies, processes, and standards that ensure an organization's data quality, security, privacy, and compliance. Companies should foster a culture

that values data governance by allocating resources for quality assurance and empowering teams to maintain high data standards. By doing so, they set the foundation for successful AI implementation and data-driven decision making.

Wells Fargo has implemented a centralized data governance framework to ensure data accuracy, consistency, and regulatory compliance across its vast financial operations. The company follows a **single source of truth approach**, meaning all critical data is managed through standardized policies and controls to eliminate discrepancies and maintain integrity across departments. This framework includes strict data quality standards, access controls, and automated validation processes, ensuring that financial reports, risk assessments, and customer records remain reliable. By embedding compliance-driven governance, Wells Fargo meets stringent regulatory requirements (discussed in the next section), such as General Data Protection Regulation (GDPR) and the U.S. Federal Reserve's data management guidelines, reducing legal risks and improving transparency. Additionally, the company's governance policies promote data accessibility for business teams while enforcing security protocols that protect sensitive customer information. This structured approach not only strengthens regulatory compliance but also enhances decision making, operational efficiency, and customer trust, making Wells Fargo a model for enterprise-level data governance in the financial industry.[9]

DATA REGULATIONS (LO3-4)



While robust data security measures protect sensitive information from breaches and misuse, database managers need to follow data regulations that govern how data is collected, stored, and used. Navigating the changing regulatory landscape is an ongoing challenge for businesses. Regulatory frameworks range from broad governmental policies that govern privacy, security, and ethical AI use to industry-specific guidelines. The goal of the regulations is to protect sensitive data while fostering trust and accountability.

Broad Governmental Policies

Governmental policies play a crucial role in shaping how data is managed, shared, and protected. They establish legal frameworks that ensure data's ethical and responsible use while balancing innovation with privacy and security concerns. Given the global nature of business on the Internet, companies must be aware of policies in any jurisdiction where they might have customers. Adherence to these policies is essential to avoid legal and financial penalties. Here are some of the most broad-reaching policies and their main provisions. The book's website contains a more comprehensive list of state and regional policies.

General Data Protection Regulation (GDPR)

Enacted by the European Union (EU) in 2018, the **General Data Protection Regulation (GDPR)** imposes strict rules on data privacy and usage, including the requirement for consent to collect and process personal data, rights for individuals in the EU to access their data, and strict rules for data breach notification. Key aspects include:

- ***Consent***. It requires explicit consent for data collection with a clear purpose.
- ***Right to access.*** Individuals can request access to their data stored by companies.
- ***Data minimization and purpose limitation.*** Companies can only collect data necessary for specific, legitimate purposes.
- ***Data portability.*** Individuals can obtain and reuse data across different services.
- ***Right to be forgotten.*** Individuals can request the deletion of their data.
- ***Data protection officers (DPOs).*** Certain organizations must appoint a DPO to oversee compliance.

The GDPR has global implications, affecting any business with EU residents' data.

Noncompliance can result in hefty fines of up to €20 million ($20.87 million) or 4 percent of the company's annual global turnover, whichever is higher.

Companies Fined under the GDPR

Some major companies have paid large penalties under the GDPR:[10]

- ***Meta.*** In May 2023, Meta was fined €1.2 billion ($1.3 billion) for transferring European users' data to the United States without adequate protection mechanisms.
- ***Amazon.*** In July 2021, Amazon received a €746 million ($891 million) fine from Luxembourg's National Commission for Data Protection for processing personal data in violation of GDPR provisions.
- ***LinkedIn.*** In October 2024, LinkedIn was fined €310 million ($341 million) by the EU privacy regulator for processing personal data without a proper legal basis in its targeted advertising practices.
- ***Uber.*** In August 2024, Uber was fined €290 million ($319 million) by the Dutch Data Protection Authority for improperly transferring driver data from the EU to the United States, lacking sufficient safeguards.
- ***WhatsApp.*** In September 2021, WhatsApp was fined €225 million ($266 million) by Ireland's Data Protection Commission for failing to provide transparent information about data processing activities to users.

California Privacy Rights Act (CPRA)

Like the GDPR, the **California Privacy Rights Act (CPRA)**, enacted in 2020, gives California residents more control over their personal information, including the right to know about and delete their data and opt out of selling their personal information. While a state policy, it has a global and national impact as it would be difficult to avoid having customers in California. Key provisions include:[11]

- ***Right to know.*** Consumers can know what personal data is being collected and for what purpose.
- ***Right to access.*** Consumers can request and receive a copy of personal information a business has collected about them.
- ***Right to correct.*** Consumers can request corrections to inaccurate personal information.
- ***Right to delete.*** Consumers can request deletion of personal information held by businesses.
- ***Data sale opt-out.*** Consumers can opt out of the sale of their data.
- ***Service providers and third parties.*** Businesses must ensure that service providers and third

parties with whom they share personal information comply with CPRA requirements.

- ***Nondiscrimination.*** Businesses must ensure no discrimination occurs against consumers who exercise their CPRA rights.

Businesses subject to the CPRA include those with annual gross revenues of more than $25 million; those that buy, receive, or sell the personal information of 50,000 or more California residents, households, or devices; or those that derive 50 percent or more of their annual revenues from selling California residents' personal information.[12]

The CPRA allows for civil penalties of up to $7,500 per intentional violation and $2,500 per unintentional violation. Additionally, consumers can sue businesses for data breaches resulting from failing to implement reasonable security measures.[13] In August 2022, Sephora settled with the California Attorney General for $1.2 million over allegations of failing to disclose the sale of personal information and not honoring user requests to opt-out of such sales.[14]



Alex Staroseltsev/Shutterstock

Strict data regulations like GDPR and CPRA are reshaping data management and enforcing compliance, security, and ethical handling of information.

The 2024 European Union AI Act

While the GDPR and CPRA focus on data privacy and security, the 2024 **European Union (EU) AI Act** directly targets AI systems to ensure they are safe and transparent and respect fundamental rights.[15] The act categorizes AI applications based on risk levels, imposing stricter requirements on higher-risk applications to protect individuals' data privacy. Key provisions of risk-based classification systems include the following levels:

- ***Unacceptable risk.*** AI systems that pose a clear threat to safety, livelihoods, and rights (e.g., social scoring by governments and manipulative AI) are prohibited. **Social scoring** refers to AI systems that rate individuals based on their behaviors, activities, or characteristics. A poor score can affect their access to services or opportunities. **Manipulative AI** refers to AI systems designed to deceive, exploit, or unduly influence individuals. An example is an AI-driven e-commerce app that trick users into spending more money or deepfake political ads that mislead

voters.

- ***High risk.*** AI systems used in critical areas such as health care, transportation, and employment are subject to strict regulations and must include human oversight mechanisms to prevent and mitigate risks. For example, AIs that are used to evaluate applications for a loan or a job are considered high risk.
- ***Limited risk.*** AI systems with limited risk are subject to transparency obligations, such as informing users they are interacting with AI. For example, an AI that handles customer queries about products or services is considered a limited risk.
- ***Minimal risk.*** AI systems with minimal risk, like spam filters or AI-enabled video games, are largely unregulated.

Providers (developers) must ensure high-risk AI systems comply with safety, transparency, and accountability requirements. This includes conducting risk assessments, ensuring data quality, and implementing robust security measures. Providers also must maintain detailed documentation, conduct regular audits, and ensure AI systems are explainable and transparent. Noncompliance can result in significant fines, similar to the GDPR, with penalties up to €30 million ($31.3 million) or 6 percent of the company's annual global turnover, whichever is higher. The act is set to take effect in August 2026.[16]

Other State and Regional Regulations:

Businesses must be aware of other state and regional regulations, such as Canada's *Personal Information Protection and Electronic Documents Act (PIPEDA)*, Brazil's *Lei Geral de Proteção de Dados Pessoais (LGPD)*, and China's *Personal Information Protection Law (PIPL)*, which reflect a growing global trend toward stringent data privacy laws. As many organizations do business globally, these regulations must be incorporated into their data management plans.

Industry-Specific Data and AI Regulations

While overarching policies like the GDPR and CPRA provide a broad data privacy and security framework, industry-specific regulations address a particular sector's unique challenges and risks. These regulations ensure that organizations in specialized fields adhere to best practices in managing sensitive data and responsibly deploying AI technologies. While governments mandate these regulations, companies providing software in these industries work in a global market and thus need to have their software compliant for their major markets. Here are some examples of industry-specific rules and their implications for data and AI.

Health Care: HIPAA and FDA AI Guidance

The *Health Insurance Portability and Accountability Act (HIPAA)* mandates strict protections for

patient health information in the United States. This includes maintaining confidentiality, integrity, and availability across electronic systems. AI systems used in health care, such as those for medical imaging or personalized treatment recommendations, must comply with HIPAA's privacy and security rules. In addition, the FDA's *Software as a Medical Device (SaMD)* guidance sets specific criteria for the approval and monitoring of AI tools in medical contexts. These regulations emphasize transparency, accuracy, and patient safety, requiring continuous updates to AI models as new data becomes available.

Education: FERPA and Student Privacy

The *Family Educational Rights and Privacy Act (FERPA)* protects student records in educational institutions. AI systems for personalized learning platforms, student assessment tools, or administrative systems must comply with FERPA's data access, sharing, and security requirements. Violations can result in the loss of federal funding for institutions.

Finance: FINRA and PCI DSS

In the financial sector, the *Financial Industry Regulatory Authority (FINRA)* oversees the use of technology to ensure transparency and protect against fraud. AI applications like algorithmic trading and risk assessment must adhere to stringent reporting and oversight requirements. Additionally, the *Payment Card Industry Data Security Standard (PCI DSS)* governs the secure handling of payment data. AI systems involved in transaction monitoring and fraud detection must operate within these security standards to protect consumer data and prevent breaches.

Retail and E-Commerce: Customer Data Protection

Retailers and e-commerce platforms often utilize AI for recommendation engines, targeted advertising, and CRM. These activities must comply with data security standards like PCI DSS for payment processing. For instance, AI-driven marketing campaigns must respect consumer rights to opt out of data collection and avoid using personal data without explicit consent.

Manufacturing and IoT: Industrial Data Governance

Integrating IoT devices in manufacturing has created new challenges in data governance. AI systems used for predictive maintenance, supply chain optimization, and quality control rely on data generated by IoT sensors. These regulations align with good cybersecurity practices and also emphasize securing machine-to-machine communications.

Navigating Industry-Specific Compliance

Organizations must proactively align their AI practices with industry-specific regulations and broader legal frameworks. This includes regular audits, employee training, and adopting robust

data security measures tailored to the industry's requirements. By doing so, businesses can minimize risks and maintain compliance in an increasingly regulated environment.

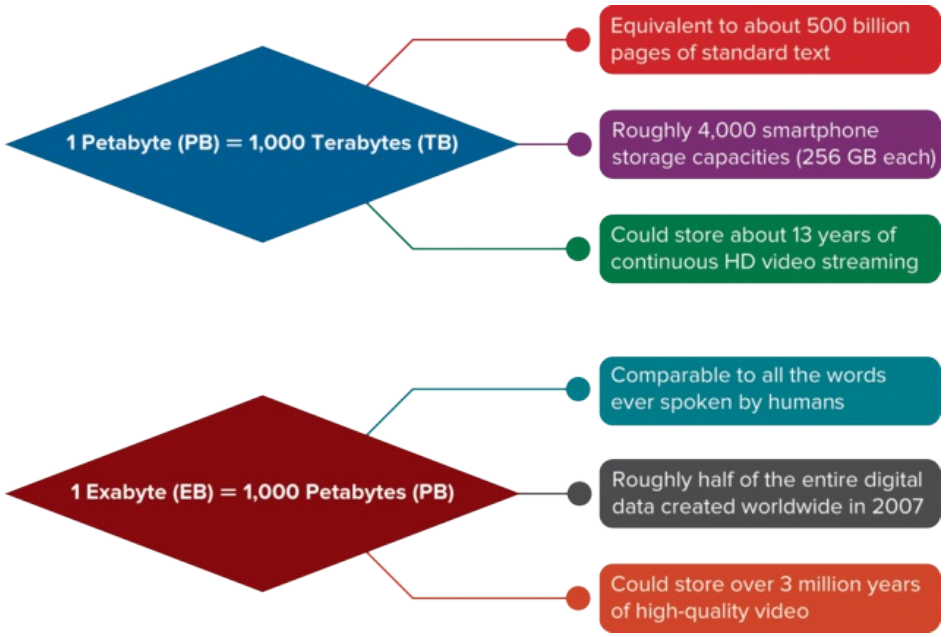
BIG DATA AND DATA MANAGEMENT (LO3-5)

The rapid growth of Big Data has revolutionized how organizations store and manage information, pushing the boundaries of traditional storage systems and necessitating innovative solutions to accommodate the increasing volume, velocity, and variety of data. One significant response to the demands of Big Data has been the expansion of data centers, which evolved to accommodate the unprecedented scale and complexity of modern data. This section explores key advancements in addressing the challenges posed by Big Data.

Larger Data Centers

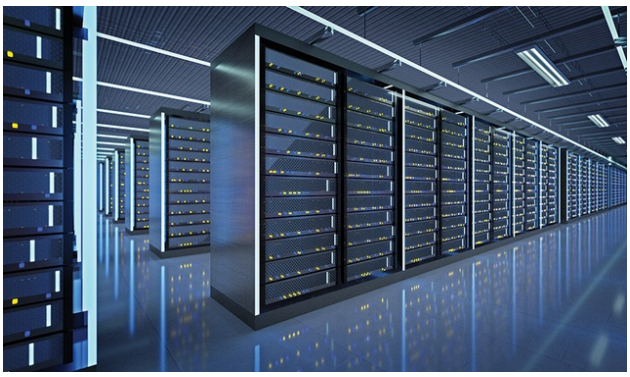
Many organizations have expanded the size of their data centers to handle the sheer volume of Big Data. Modern data centers are designed to store petabytes (PB) or even exabytes (EB) of data, but these measurements can be challenging to grasp. **Figure 3.3** puts these sizes into perspective. These massive storage capacities enable businesses, cloud providers, and AI systems to process, analyze, and store information, fueling modern data-driven decision making.

FIGURE 3.3 How Big Are a Petabyte and an Exabyte?



Data centers that hold an exabyte of data are rare and are primarily used for supercomputers.

Some examples are the CERN Data Center, whose data storage surpassed 1 exabyte, distributed across approximately 111,000 devices supporting its research into particle physics. The NASA Advanced Supercomputing Division's primary mission is to provide supercomputing resources and services to support various NASA missions and projects and has over an exabyte of data.[17] The U.S. National Security Agency (NSA) stores data estimated to be on the order of exabytes or larger, supporting national cybersecurity initiatives.



pp76/iStock/Getty Images

Data centers like this one can contain rows upon rows of servers.

Distributed Cloud-Based Storage Systems

While larger data centers meet the needs of centralized data storage, **distributed cloud-based storage systems** offer a flexible and scalable alternative that addresses storage capacity and more localized data access. Distributed cloud-based storage systems spread data across multiple servers and permit seamless scaling as data grows. Data is held in multiple locations, which reduces the risk of data loss and increases system reliability. Distributed storage can also bring data closer to its sources, such as IoT devices, reducing data transmission delays and enabling faster processing.

Third-party platforms like AWS, Azure, and Google Cloud allow organizations to scale storage capacity dynamically, paying only for the resources they use. Cloud storage as a service eliminates the need for large upfront investments in physical infrastructure, making it accessible to businesses of all sizes. However, the expanded scale of data storage raises concerns about security and compliance with regulations. Distributed and cloud-based systems require robust access controls to safeguard sensitive information. These systems also require that responsible companies encrypt data at rest and during transmission to protect it from unauthorized access and breaches.

Sustainability in Storage

As data storage strategies evolve to balance scalability, privacy, and performance, organizations are also grappling with the environmental impact of their infrastructure, emphasizing the need for sustainable storage practices. Data centers require vast amounts of electricity to power servers and maintain cooling systems, making sustainability a key consideration for businesses. Several drivers are propelling companies toward more sustainable practices in their storage solutions.

Rising Electricity Rates

Data centers' substantial energy consumption makes them particularly vulnerable to fluctuating and rising electricity costs. To reduce operating expenses, companies are exploring more energy-efficient hardware and innovative cooling techniques. Additionally, many organizations are investing in renewable energy sources to power their data centers, while others are evaluating the role of nuclear energy as a reliable, low-carbon power source for long-term sustainability.

Google's data center campus in Belgium features a 2.8-megawatt solar plant, comprising 10,665 ground-mounted photovoltaic panels. This installation reduces reliance on grid energy and decreases carbon dioxide emissions. In December 2024, Google announced a $20 billion partnership with Intersect Power and TPG Rise Climate to develop data centers alongside adjacent solar and wind farms in the United States.[18] This initiative aims to meet the increasing energy demands of AI technologies while ensuring a sustainable power supply. This is a strategy being pursued by numerous data centers around the world.

Companies also employ various strategies to power their data centers using hydroelectric and geothermal energy, including building dedicated facilities and contracting to purchase electricity. Here are some notable examples:

- ***Green Mountain Data Centers.*** This Norwegian company operates data centers powered entirely by onsite hydroelectric plants, ensuring a renewable energy supply with minimal carbon footprint.[19]
- ***Data Center Light.*** Located in Switzerland, Data Center Light utilizes an in-house hydroelectric plant, with servers running on 99.9 percent hydropower generated just meters away, supplemented by 0.1 percent solar power.[20]
- ***Google.*** In partnership with Fervo Energy, Google developed a geothermal system in Nevada. Operational since November 2023, this project supplies carbon-free electricity to the grid, supporting Google's data centers in the region.[21]
- ***Meta.*** Collaborating with Sage Geosystems, Meta is developing a project to provide up to 150 megawatts of geothermal energy for its data centers. It is expected to be operational by 2027.[22]

Companies are also exploring *small modular reactors (SMRs)* and revitalizing decommissioned nuclear plants to provide reliable, carbon-free energy for data centers. SMRs and type of nuclear power technology, similar to submarine and aircraft carrier reactors, that use *pressurized water reactor technology*, passive safety features, and long-duration fuel cycles to deliver stable, high-output power. While still in development, the first commercial SMRs could be operational by the late 2020s to early 2030s, with companies like NuScale, TerraPower, and X-energy leading the

charge. In the meantime, repurposing decommissioned reactors could accelerate nuclear energy adoption for data centers, though regulatory, financial, and public perception challenges remain. If successful, SMRs and atomic energy could complement renewables like wind and solar, ensuring continuous, scalable power for energy-intensive AI-driven operations.[23]

Cooling systems in data centers also often rely heavily on water, posing challenges in regions facing water scarcity or drought. Sustainable practices, such as air-based cooling and closed-loop water systems, help reduce dependency on water resources, ensuring long-term operational viability.

Regulatory Compliance and Corporate Social Responsibility

Governments worldwide are introducing stricter environmental regulations to curb carbon emissions and resource consumption. The European Union's Corporate Sustainability Reporting Directive (CSRD) mandates extensive sustainability reporting, including energy consumption metrics, to promote transparency and accountability.[24] Businesses must adapt their data center operations to comply with these laws, avoiding penalties while aligning with global sustainability goals.

But many businesses want to go beyond mere regulatory compliance. In an era where corporate social responsibility plays a key role in brand reputation, businesses are making public commitments to sustainability—particularly in data centers. One notable initiative is the *Climate Neutral Data Centre Pact*, a self-regulatory agreement launched in 2021 by data center operators and trade associations across Europe. The pact aligns with the European Green Deal's objective of achieving climate neutrality by 2030. Key commitments include:

- ***Energy efficiency.*** Meeting strict power usage effectiveness (PUE) targets by 2025 for new data centers and by 2030 for existing facilities.
- ***Renewable energy usage.*** By 2025, at least 75 percent of energy consumption will come from renewable sources, and by 2030, it will be 100 percent.
- ***Water conservation.*** Implementing measures to reduce water consumption in cooling systems.
- ***Circular economy initiatives.*** Ensuring 100 percent of used server equipment is assessed for reuse, repair, or recycling.
- ***Heat recycling.*** Where feasible, redirecting waste heat from data centers to district heating systems.

The more than 100 signatories, including major players like AWS, Google, Equinix, and Digital Realty, undergo independent third-party audits to verify compliance. If a member fails to meet the set targets, they risk expulsion from the pact.[25] By aligning with climate-focused policies, the companies enhance their corporate image and future-proof their infrastructure against evolving environmental regulations.

As organizations expand their data management capabilities, balancing scalability with

sustainability will remain a critical challenge. Innovations in energy-efficient storage systems, renewable energy integration, and resource optimization pave the way for greener data centers, ensuring that the digital economy thrives without compromising the planet's future.

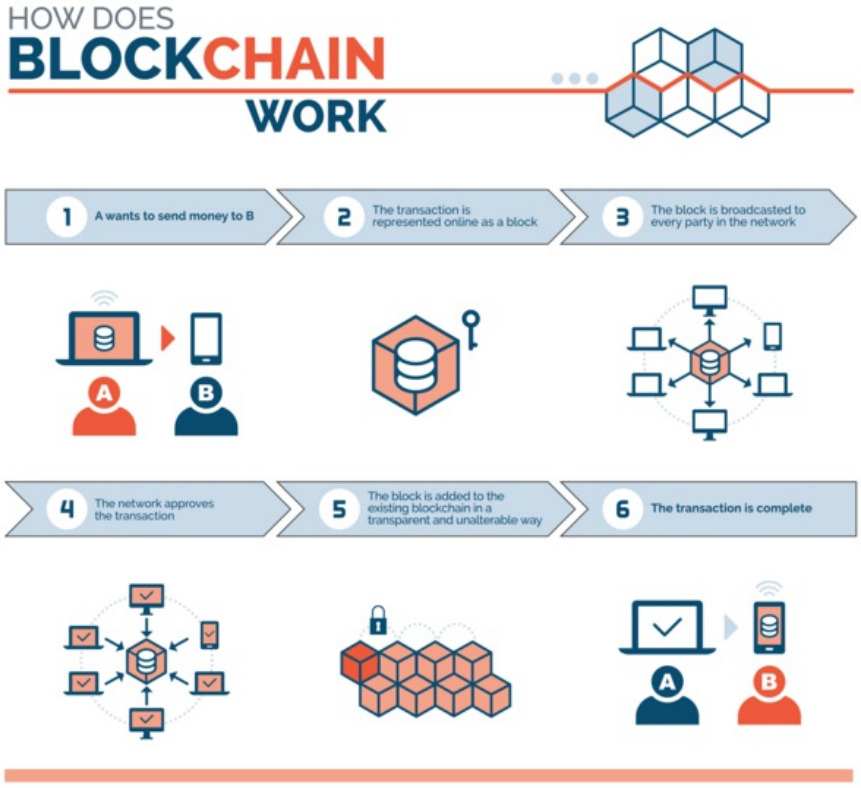
EMERGING TECHNOLOGIES IN AI DATA MANAGEMENT (LO3-6)

AI-driven businesses must handle increasing data complexity, security threats, and scalability challenges. Emerging technologies like blockchain, AI-powered cybersecurity, edge computing and storage, and federated learning are transforming AI data management, making it more secure, decentralized, and efficient.

Blockchain

Blockchain is a digital record-keeping technology (often associated with cryptocurrency) that securely and transparently records transactions in multiple locations (see **Figure 3.4**). Each transaction is entered as a cryptographically secured block and linked to the previous block, forming an unchangeable chain (hence, a blockchain.) Recording the data in multiple locations means the records can be compared against each other to ensure validity.

FIGURE 3.4 How Blockchain Works



elenabsl/Shutterstock

The following are some of the key features of blockchain:

- ***Transparency.*** All network participants have access to the same ledger, which contains a record of all transactions allowing for visible and auditable changes.
- ***Security.*** Once a block is added to the chain, it becomes nearly impossible to alter without modifying all subsequent blocks, ensuring data integrity.
- ***Decentralization.*** The network maintains the ledger collectively, eliminating reliance on a single authority.
- ***Efficiency.*** Blockchain operates continuously and autonomously, enabling secure and swift data exchanges.

Blockchains have been used to secure and validate various types of important documents, such as academic transcripts, supply chain data, and medical records. The data is immutable, traceable, and accessible only to authorized individuals, addressing critical privacy and security concerns. For example, Guardtime, an Estonian cybersecurity company, uses its Keyless Signature Infrastructure (KSI) blockchain to secure sensitive health care and pharmaceutical data, ensuring

unalterable patient records and privacy-compliant data sharing.[26]

While blockchain offers robust security and transparency, it has notable performance limitations that can make it less suited for high-volume transaction processing.

- ***Transaction speed constraints.*** Most blockchain networks require validation from multiple participants, slowing transaction processing times. Traditional centralized databases or payment networks, such as Visa, can handle tens of thousands of transactions per second, while blockchains Bitcoin and Ethereum process 3 to 7 transactions per second.[27]
- ***Scalability challenges.*** With the relatively slow speed to record a transaction, as the blockchain networks grows, increased transactions can lead to network congestion, slower speeds, and higher processing fees.

Blockchain's integration into AI-driven data management enhances trust and security across diverse sectors. Still, organizations must consider its scalability constraints when determining its applicability for large-scale data transactions.

AI-Powered Cybersecurity

AI plays a pivotal role in enhancing cybersecurity for data management. AI-driven systems can proactively predict, detect, and respond to emerging cyber threats in real time, offering an adaptive layer of protection that evolves alongside the threat landscape. These systems analyze huge amounts of data to identify patterns indicative of attacks, such as unusual network activity or unauthorized access attempts, and can implement immediate countermeasures to mitigate risks.

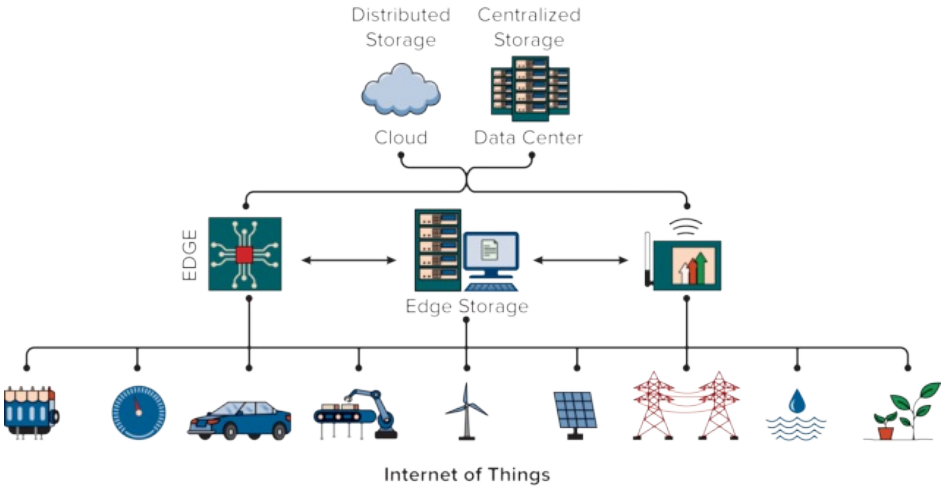
For example, AI anomaly detection tools identify deviations from typical user or system behavior, alerting security teams to potential breaches. Natural language processing of emails enables the analysis of phishing attempts, reducing response times and human error. By integrating AI into cybersecurity, organizations can better ensure the integrity, confidentiality, and availability of the data critical for AI applications.

Edge Computing and Storage

Big Data centers and centralized processing can be economical, but AI applications like autonomous vehicles sometimes need to make quick decisions. Edge computing and storage allows for more rapid decision making. As shown in **Figure 3.5**, edge computing processes data near the source of data generation, such as IoT devices or local servers, instead of sending it to centralized data centers. This distributed approach reduces data transmission and processing delays, conserves bandwidth, and supports real-time decision making. For example, in autonomous vehicles, edge computing enables real-time data processing directly within the vehicle, allowing immediate responses to dynamic driving conditions without relying on distant cloud servers.[28]

FIGURE 3.5 How Edge Computing Works

By processing data near its source, edge computing saves time, conserves bandwidth, and supports real-time decision making.



For data management, edge computing introduces the concept of **edge storage**, which involves smaller, distributed servers located near sensors and devices. This localized storage handles tasks such as data aggregation, filtering, and preprocessing before transmitting only relevant or summarized data to central systems. This reduces the load on centralized data centers, minimizes delays, and enhances efficiency, particularly for applications requiring instant responses. In practice, edge storage complements centralized and distributed storage systems, creating a layered architecture:

1. ***Centralized storage.*** Houses long-term, archival data and supports large-scale analytics.
2. ***Distributed storage.*** Balances data across cloud systems, improving redundancy and accessibility.
3. ***Edge storage.*** Manages real-time, local data near its source to enable immediate analysis and reduce data transfer demands.

This multi-tiered storage model ensures scalability and adaptability, aligning with the growing complexity of AI-driven ecosystems. **Table 3.3** compares and contrasts centralizes storage and processing with edge storage and processing.

Table 3.3 **Centralized Storage and Processing vs. Edge Storage and Processing**

Feature	Centralized Storage and Processing	Edge Storage and Computing
Data storage location	Stored in centralized data centers or cloud servers	Stored closer to the source (e.g., IoT devices, local servers, edge nodes)
Processing location	Processing occurs in centralized cloud servers or data centers	Processing occurs near the data source on local devices or edge servers
Latency	Greater delays due to data transmission over long distances	Faster as processing happens closer to the user or device
Bandwidth usage	High bandwidth usage as large volumes of data are transmitted to central servers	Reduces bandwidth usage by processing data locally before sending summaries to the cloud
Scalability	Highly scalable due to cloud resources, but may require significant infrastructure investment	Scalable by adding more edge nodes, but management can be complex
Security	Strong security measures, but centralized nature can make it a prime target for cyberattacks	Enhanced security as sensitive data can be processed locally, reducing exposure to centralized breaches
Cost considerations	Can be expensive due to infrastructure, storage, and data transfer costs	Lower data transfer costs, but requires investment in distributed edge infrastructure
Use cases	Cloud applications, enterprise databases, large-scale AI/machine language processing, financial transactions	IoT applications, autonomous vehicles, smart cities, real-time analytics, industrial automation

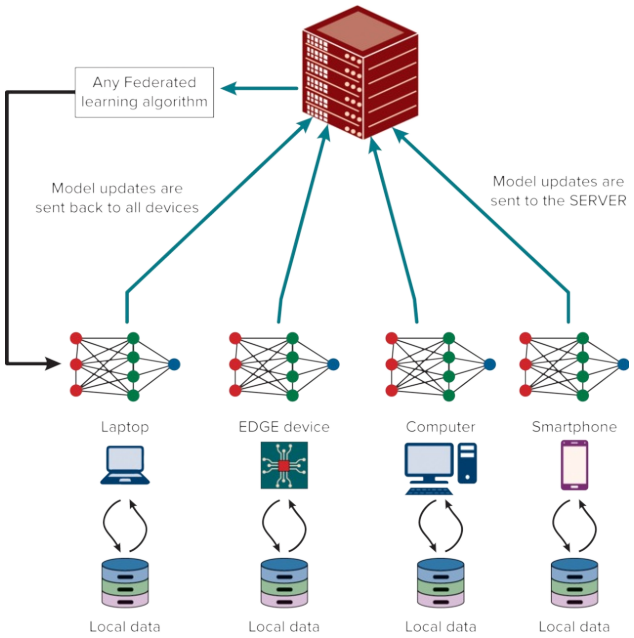
Federated Learning

While edge computing emphasizes real-time processing at the data source, **federated learning** is a complementary technology focused on privacy-preserving model training. Federated learning enables AI models to be trained across multiple decentralized devices or servers while keeping the data localized (see **Figure 3.6**). Instead of transferring sensitive datasets to a central repository, only the changes to the model are shared (not the data), preserving privacy. A prime example is in the healthcare sector where federated learning has been applied to predict clinical outcomes in COVID-19 patients. A collaboration of 20 institutions worldwide trained AI models on

decentralized data to predict oxygen needs in COVID-19 patients, enhancing model accuracy while preserving patient privacy.[29]

FIGURE 3.6 How Federated Learning Works

Decentralized devices train a model and share only model updates, not raw data.



This approach is useful in data-sensitive industries such as health care and finance, where regulations demand stringent privacy compliance. Federated learning keeps the data at the local computer which reduces data storage and processing at a central location and reduces the chance of a data breach during data transmission. **Table 3.4** compares and contrasts centralizes storage and processing with federated learning.

Table 3.4 Centralized Storage and Processing vs. Federated Learning

Feature	Centralized Storage and Processing	Federated Learning
Data storage location	Stored in centralized data centers or cloud servers	Data remains distributed across multiple devices or nodes
Processing location	Processing occurs in centralized cloud servers or data centers	Processing occurs locally on each device, with only model updates sent to a central server
Latency	Higher delays due to data transmission over long distances	Faster since computations happen on local devices
Bandwidth	High bandwidth usage as large	Minimizes bandwidth usage by

usage	volumes of data are transmitted to central servers	transmitting only aggregated model updates instead of raw data
Scalability	Highly scalable due to cloud resources, but may require significant infrastructure investment	Highly scalable as more devices participate in the learning process
Security and privacy	Strong security measures, but centralized nature makes it a prime target for cyberattacks	Enhanced privacy and security since raw data never leaves the local device, reducing breach risks
Computational resource requirement	Processing is handled by powerful cloud-based servers, reducing local device burden	Relies on local device processing power, which may be limited and vary across devices
Reliability and fault tolerance	High reliability, but service disruptions or cyberattacks on a single point of failure can impact operations	More fault-tolerant since local failures do not impact the entire system, but network delays can affect aggregation
Model aggregation	Not applicable, as raw data is stored and processed in a central location	Requires aggregating decentralized model updates into a global model
Regulatory compliance	May require compliance with GDPR, HIPAA, and other regulations, depending on data type	Well-suited for privacy-sensitive industries (e.g., health care, finance) to comply with GDPR, HIPAA
Cost considerations	Can be expensive due to infrastructure, storage, and data transfer costs	Reduces data transfer costs but requires strong coordination and device participation
Use cases	Cloud applications, enterprise databases, large-scale AI/ML processing, financial transactions	Personalized AI models, health care diagnostics, mobile keyboard prediction, smart assistants, IoT applications

ETHICAL CONSIDERATIONS IN AI DATA COLLECTION AND USE (LO3-7)



Ethical considerations are integral to using AI in business as a strategic priority. A 2024 study of tech companies found that businesses that adopt and consistently implement ethical practices are more likely to gain customer trust, foster long-term relationships, and enhance retention rates.[30] Ensuring that data is collected, used, and managed ethically also aligns with legal compliance. For multinational organizations, aligning ethical practices with global regulations like the GDPR, the EU AI Act, and CPRA discussed earlier in the chapter requires a proactive approach. This helps companies avoid legal risks and position themselves as ethical leaders. The following practices provide a foundation for ethical data management:

- ***Ensure transparency.*** Communicate what data is being collected, what its intended purpose is, and how it will be used.
- ***Ensure informed consent and user control.*** Individuals should be fully aware of how their data is used and have the right to opt-out or limit AI's access to their personal information. Data portability rights should allow users to transfer their data between platforms, as GDPR and similar regulations require. Organizations should establish a **Right to Be Forgotten policy**, allowing users to request data deletion.
- ***Ensure ethical data sourcing and provenance.*** AI systems should be trained only on legally and ethically acquired data. Avoid using "scraped" or copyrighted material without permission. Organizations must implement data provenance tracking to document where data originates and how it has been processed.
- ***Ensure fairness and nondiscrimination.*** As noted earlier in the chapter, AI must be designed to avoid reinforcing systemic biases in decision making. Diverse and unbiased data is required for impartial AI outputs. Companies should conduct regular **bias audits**, use fairness metrics, and ensure inclusive training datasets to promote equitable outcomes.
- ***Safeguard personal and sensitive data.*** Ethical obligations often overlap with legal requirements such as GDPR, the EU AI Act, and CPRA. AI data management and storage systems must ensure strong encryption, secure access controls, and robust data anonymization techniques. Regular security audits should be conducted to mitigate the risks of breaches and cyberattacks. Incident response plans must be formed for security attacks and natural disasters.
- ***Employ energy-efficient AI and data storage.*** AI models, especially large-scale neural networks, involve large data centers and consume vast amounts of energy. Data centers should use renewable energy when possible and adopt sustainable cooling systems. **AI lifecycle management** practices can ensure hardware and infrastructure are recycled or repurposed rather than

wasted.

As this discussion clarifies, compliance with diverse data privacy laws and ethical guidelines across jurisdictions is essential. Responsible data practices protect organizations from legal risks and strengthen their reputation, a key asset in the digital economy.

SUMMARY OF KEY POINTS AND QUICK WINS

1. **Data for AI.** Understanding structured and unstructured data is foundational for effective AI applications. Structured data, with its fixed formats, is easier to process, while unstructured data requires advanced techniques like natural language processing and computer vision. AI systems rely on diverse and complex data sources to function effectively. These sources can be broadly categorized into internal data sources, external data sources, IoT and sensor data, and online and social media data.
2. **Data security.** Robust data security measures include encryption, access control, backup and disaster recovery, and protecting sensitive AI data during transmission and storage. AI tools can enhance cybersecurity efforts by detecting and mitigating threats in real time.
3. **Data quality and integrity.** The quality and integrity of AI systems' underlying data are paramount for reliable insights. High-quality data is accurate and reliable, consistent across sources, free from bias, and complete. Ensuring high-quality data in AI systems begins with a structured data cleaning process. Data governance is the framework of policies, processes, and standards that ensure an organization's data quality, security, privacy, and compliance.
4. **Data regulations.** While robust data security measures protect sensitive information from breaches and misuse, database managers need to follow data regulations that govern how data is collected, stored, and used. Regulations like the GDPR and the CPRA mandate strict data privacy and security guidelines.
5. **Big data and data management.** The rise of big data has necessitated innovations in data storage and management, such as larger data centers, and distributed cloud-based storage systems. These technologies support scalable, secure, and efficient AI operations. As data storage strategies evolve to balance scalability, privacy, and performance, organizations are also grappling with the environmental impact of their infrastructure, emphasizing the need for sustainable storage practices.
6. **Emerging technologies in AI data management.** Blockchain provides transparency, security, and immutability in managing AI data. AI plays a pivotal role in enhancing cybersecurity for data management. Edge computing and storage allows for more rapid decision making. Federated learning enables AI models to be trained across multiple decentralized devices or servers while keeping the data localized.
7. **Ethical considerations in AI data collection and use.** Transparency, fairness, and accountability are key ethical principles in AI data collection and use. Businesses must address biases, ensure informed consent and user control, ensure ethical data sourcing and provenance, ensure fairness and nondiscrimination, safeguard personal and sensitive data, and employ energy-efficient AI and data storage.

QUICK WINS

1. **Understand data structures.** Recognize the difference between structured and unstructured data in AI systems. Apply appropriate techniques and tools to process and analyze different data types effectively.
2. **Keep data secure.** Implement robust data security measures, such as encryption and access controls, to protect sensitive AI data during storage and transmission. Utilize AI-powered tools to detect and mitigate cybersecurity threats.
3. **Comply with government regulations.** Ensure compliance with data regulations, such as GDPR and CPRA, safeguard privacy, avoid legal risks, and maintain stakeholder trust.
4. **AI starts with good data.** Prioritize data quality and integrity by investing in cleaning, standardization, and bias mitigation processes. Reliable AI outcomes depend on accurate, consistent, and complete data.
5. **Implement advantageous data management innovations.** Leverage innovations in big data management, such as distributed cloud storage, edge computing, and federated learning, to support scalable and efficient AI operations.
6. **Consider blockchain for secure record keeping.** Explore blockchain to enhance transparency, security, and trust in data management. Identify use cases where blockchain can improve data integrity and accountability.
7. **Practice ethical data management.** Promote ethical data practices by ensuring transparency, fairness, and accountability in AI data collection and use. Regularly audit for biases and maintain privacy safeguards to meet ethical and stakeholder expectations.

KEY TERMS

AI lifecycle management. Managing AI systems from development to decommissioning while ensuring sustainability and ethical considerations.

Bias audits. Assessments conducted to detect and mitigate biases in AI systems to promote fairness and inclusivity.

Blockchain. A secure digital ledger that records information in a way that is difficult to change, making it useful for tracking transactions and sharing data across multiple parties with trust and transparency.

California Privacy Rights Act (CPRA). A California law that strengthens consumer privacy rights by giving people more control over their personal data, including the right to know what data is collected, to correct it, to delete it, and to limit how businesses use or share it.

Data cleaning. The structured process of transforming raw, inconsistent, or incomplete data into high-quality inputs for AI systems.

Data governance. The framework of policies, processes, and standards that ensure an organization's data quality, security, privacy, and compliance.

Deduplication. The removal of redundant or conflicting records to ensure data accuracy and prevent distortion in AI models.

Distributed cloud-based storage systems. A method of storing data across multiple servers and locations in the cloud, allowing businesses to easily scale their storage needs, improve data access speeds, and reduce the risk of data loss due to server failures.

Edge computing. A method of processing data closer to where it is generated (such as a sensor or device) rather than sending it to a central server, which helps reduce delays and enables faster, more efficient real-time decision making.

Edge storage. A system that stores data locally on or near the device that collects it, such as a factory sensor or smart appliance, allowing information to be processed quickly without needing to send everything to a distant data center.

Encryption. A security technique that scrambles data into a coded format, allowing only individuals with the correct password or key to read it, thereby protecting information while it's being stored or shared.

European Union (EU) AI Act. A European Union law that classifies AI systems by risk and sets rules to make sure high-risk AI is safe, transparent, and respects people's rights, especially in areas like health care, hiring, or public safety.

Fairness metric. Quantitative measure used to assess whether an AI system treats different groups equitably.

Federated learning. A method of training AI models across many devices or servers without sharing raw data, which helps protect privacy by keeping personal data where it was created while

still improving the model.

General Data Protection Regulation (GDPR). A European law that gives people control over their personal data by requiring businesses to be clear about how they collect, use, and protect it; violations can lead to large fines.

Identity and Access Management (IAM). A system of policies and tools that enables organizations to control who can access what data and systems, ensuring that only authorized personnel have the necessary level of access.

Imputation. A method for filling in missing data using statistical estimates or predictive modeling to maintain dataset integrity.

Least privilege. A security principle that means users and systems should only have the minimum level of access necessary to do their job and nothing more to reduce the risk of misuse or data breaches.

Manipulative AI. AI systems designed to deceive, exploit, or unduly influence individuals.

Right to Be Forgotten policy. A policy that ensures the right of individuals to request the deletion of their data from databases.

Role-based access control. A method for managing access to resources by assigning permissions based on job roles ensuring that individuals only have access to the data and tools they need for their specific responsibilities.

Single source of truth approach. A strategy where all departments and systems rely on one trusted and up-to-date version of data, helping to avoid confusion, duplication, and errors in decision making.

Social scoring. AI systems that rate individuals based on their behaviors, activities, or characteristics.

Standardization. The process of ensuring data follows consistent formats, units, and naming conventions across different sources for better integration and analysis.

Structured data. Information that is organized in a fixed format—like spreadsheets or databases—making it easy for computers to search, sort, and analyze.

Unstructured data. Information that doesn't follow a consistent format, such as emails, videos, social media posts, or images; this type of data is harder to organize but can hold valuable insights when processed with AI.

DISCUSSION QUESTIONS

1. **Data security.** How can organizations balance the need for robust data security with the challenges posed by managing large, distributed datasets? What are the potential risks of relying on AI for threat detection and mitigation?
2. **Data quality and integrity.** Why is the adage "garbage in, garbage out" particularly relevant in AI applications? What strategies can organizations implement to ensure their data's accuracy, consistency, and completeness for AI systems?
3. **Regulatory compliance.** In a global business environment, how should companies approach compliance with diverse data privacy laws like GDPR and CPRA? What are the consequences of failing to adhere to data privacy regulations, and how can organizations proactively address these risks?
4. **Big Data and advanced technologies.** How have advancements in Big Data, edge computing, and federated learning transformed data storage and processing?
5. **Sustainability.** How can organizations address the environmental impact of large data centers while maintaining efficient data management practices?

EXERCISES

1. **Data structure and types.** Discuss the uses and challenges of managing unstructured data in AI applications.
2. **Data security.** Develop a data security plan for a hypothetical company, including data encryption, access control, and disaster recovery. Explain how these measures protect against cybersecurity threats.
3. **Data quality and integrity.** Create a checklist for data quality in AI projects, including steps for data cleaning, validation, and bias detection. Explain the importance of each step for reliable AI outcomes.
4. **Regulatory compliance.** Report on the impact of data regulations (e.g., GDPR) on AI data management practices, and propose strategies for ensuring compliance.
5. **Ethical considerations.** Discuss the ethical implications of AI data collection and usage in classrooms, covering informed consent, data privacy, and bias. Suggest ethical AI guidelines that meet legal and societal standards.

CASE STUDY 3.1: AMAZON—AI IN THE HIRING PROCESS

Incorporating AI into the hiring process offers several potential benefits that can streamline candidate screening and enhance the efficiency and effectiveness of the hiring process. By utilizing AI, companies can automate the initial screening of résumés, saving substantial time and resources that human resources (HR) departments typically invest in reviewing applications. This automation accelerates the hiring process and ensures a more objective evaluation of candidates by minimizing human biases, leading to a more diverse and inclusive workforce.

Furthermore, AI algorithms can analyze vast amounts of data to identify patterns and predict which candidates will likely best fit for a role based on their skills, experiences, and other relevant factors. This predictive capability can improve the quality of hires by matching candidates to positions where they are most likely to succeed and contribute to the organization's goals. Leveraging AI in candidate screening promises to make the recruitment process more efficient, equitable, and practical, allowing companies to attract and retain top talent more effectively.

The Start of the Journey

Amazon's journey with its AI-driven recruitment tool began in 2014. The project's goal was to automate hiring and improve talent acquisition efficiency by using AI to analyze résumés and identify top talent based on patterns in historical hiring data. This system was expected to enhance efficiency, reduce manual résumé reviews, and uncover potentially overlooked candidates. Over 150,000 people were employed at Amazon at that time, giving them a large database of past candidates, hires, application materials, and success at the company with which to train the AI. It did not work quite the way they had planned, though.

Clues of Bias

The first clue that the AI selection process at Amazon produced biased results came to light when the company discovered that its system was not rating candidates gender-neutrally. Specifically, the AI favored male candidates over female candidates for technical roles. This bias was traced back to the training data used to develop the AI system, which consisted of résumés submitted to the company over 10 years. Since men have historically dominated the tech industry, AI learned to replicate these existing biases, presuming that male candidates were preferable for technical positions. For instance, the system could downgrade résumés that included words like "women's," as in "women's chess club captain," or mentioned attendance at all-women's colleges.

This bias stemmed not from deliberate programming but from learning from skewed historical

data, primarily male applicant résumés. Algorithms that disproportionately weed out job candidates of a particular race, gender, or religion violate Title VII, the federal law prohibiting discrimination in employment. Intent does not matter. “Disparate impact discrimination” is enough to make such practices illegal. However, failure-to-hire lawsuits are very rare because it’s difficult for someone who never got an interview to identify the policy or practice that led to their rejection.

Repercussions and Repairs

The incident had significant repercussions for Amazon, leading to public scrutiny and internal reassessment of AI ethics and practices. Although it did not result in direct lawsuits, the case had reputational impacts and highlighted the need for greater transparency and responsibility in AI deployments.

Amazon’s engineers attempted to rectify this by programming the AI to ignore gendered words and balance the training dataset. Despite these efforts, the bias persisted. Superficial data adjustments are often insufficient, and a more profound reconsideration of AI training approaches is necessary to avoid inherent biases. A critical need exists for careful oversight, diverse training data, and regular auditing of AI systems to detect and correct biases that could lead to unfair or discriminatory outcomes.

Ultimately, Amazon discontinued the tool due to the ongoing bias issues and associated reputational and ethical risks. This decision underscored the complexities of developing unbiased AI systems, particularly in nuanced domains like human resource management. The Amazon case has reshaped the approach to AI in recruitment, emphasizing the development of technically proficient, ethically sound, and transparent systems. It has led to a focus on balanced training datasets, human oversight, and continuous bias monitoring in AI systems.

Furthermore, the Amazon case has influenced regulatory discussions in the United States and Europe, contributing to narratives around ethical AI frameworks and legislation like the GDPR. It spurred innovation in ethical AI and development of tools and methodologies to ensure fairness and unbiasedness in AI systems.

Discussion Questions

1. How could Amazon have better anticipated and mitigated the gender bias in its AI recruitment tool? Discuss the ethical implications of deploying AI tools without thorough bias testing.
2. Compare the Amazon AI recruitment case with another known instance of AI bias. What similarities and differences do you observe in terms of the root causes and impacts of the biases?
3. Discuss how the choice of data impacts the output of AI algorithms. What steps can be taken to ensure a diverse and unbiased dataset?

4. Should specific regulations govern AI in recruitment? How might these regulations have influenced the Amazon case?
5. Considering the Amazon case, how do you envision the future of AI in human resource management? What are the potential benefits and pitfalls?

Case References

Chang, Xinyu. (2023). “Gender Bias in Hiring: An Analysis of the Impact of Amazon’s Recruiting Algorithm.” *Advances in Economics, Management and Political Sciences*, 23, 134–140. 10.54254/2754-1169/23/20230367.

CASE STUDY 3.2: DARKTRACE—AI IN NETWORK SECURITY

Cyberthreats are more sophisticated and persistent than ever. Organizations face challenges such as advanced persistent threats (APTs), zero-day exploits, ransomware attacks, and insider threats. Traditional cybersecurity measures, often reliant on predefined rules and known threat signatures, struggle to keep pace with evolving tactics. Darktrace, an innovator in AI-powered network security, utilizes technologies like deep learning and generative AI to detect and mitigate threats in real time autonomously.

Founded in 2013 in Cambridge, England, Darktrace become a cutting-edge player in the cybersecurity industry. A collaboration of mathematicians and cybersecurity experts, including members of the UK's intelligence agencies, the company has been at the forefront of using AI for network threat detection and response. Darktrace serves clients across sectors, from health care and finance to government and manufacturing.

At the heart of Darktrace's technology is its ability to understand normal behavior within an organization's network. The system uses unsupervised deep learning models to create a unique "pattern of life" for each company environment, identifying deviations that may signal a threat. Unlike traditional systems, which rely on predefined rules, Darktrace's AI adapts to emerging threats without prior knowledge of specific attack vectors.

Darktrace integrates generative AI to simulate potential threat scenarios. This helps organizations identify and defend against vulnerabilities. The system can suggest adjustments to network configurations, reducing exposure to risks.

One of Darktrace's standout features is its autonomous response capability. When a threat is detected, it isolates compromised devices, blocks malicious traffic, and neutralizes attacks without human intervention. This speed is valuable combating fast-moving threats like ransomware.

Benefits and Challenges

Darktrace's AI-driven approach delivers tangible benefits to organizations, including:

- ***Improved threat detection.*** Identifies previously unseen threats that traditional systems miss.
- ***Cost savings.*** Reduces the need for extensive manual monitoring and response teams.
- ***Operational continuity.*** Minimizes downtime by responding to threats in real time.
- ***Regulatory compliance.*** Enhances data protection and ensures adherence to laws like GDPR and CPRA.

Despite its advantages, Darktrace faces several challenges:

- ***Explainability.*** The deep learning models that Darktrace employed operated as "black boxes," making it difficult for users to understand the rationale behind decisions.
- ***False positives.*** Autonomous responses risk incorrectly isolating legitimate activity, potentially disrupting operations.
- ***Integration complexity.*** Organizations with diverse IT systems may struggle to integrate Darktrace's solutions.
- ***Evolving threats.*** As cybercriminals develop more sophisticated tactics, Darktrace must adapt its models to stay ahead.

Broader Implications

Darktrace's innovative use of unsupervised and generative AI has applications beyond just cybersecurity. These technologies hold significant potential to enhance anomaly detection across industries, such as fraud prevention in finance, identifying abnormal patient data in health care, and detecting inefficiencies or disruptions in supply chain management. Darktrace demonstrates the transformative power of AI in industries that depend on complex data analysis. Despite ongoing challenges, its innovations are defining the landscape of digital defense and setting a standard for the future of AI-driven solutions.

Discussion Questions

1. How does Darktrace's reliance on unsupervised learning illustrate the importance of high-quality data in AI systems? What challenges might Darktrace face in ensuring the accuracy and integrity of the vast data streams it analyzes?
2. Cybersecurity is fundamentally about protecting data. How does Darktrace's approach differ from traditional cybersecurity measures in managing and safeguarding data? What lessons can other industries, such as finance or health care, learn from Darktrace's methods of securing and analyzing data?
3. Darktrace's autonomous response capability allows it to act without human intervention. What are the potential risks and benefits of allowing AI systems to make autonomous decisions in critical scenarios? How might the "black box" nature of Darktrace's deep learning models impact trust in its decisions? What steps could the company take to improve transparency?
4. What challenges might organizations face when integrating Darktrace's solutions into their existing IT infrastructure? How can these challenges be mitigated?
5. What additional features or innovations could Darktrace develop to strengthen its market position?

Case References

Darktrace. “Darktrace AI: Combining Supervised and Unsupervised Machine Learning.” Technical White Paper. Darktrace Resources. **https://cdn.prod.website-files.com/626ff4d25aca2edf4325ff97/62a292c20b7371fb41311930_wp-machine-learning.pdf**

Darktrace. “The State of AI in Cybersecurity: Understanding AI Technologies.” Darktrace Blog, July 24, 2024. **https://darktrace.com/blog/the-state-of-ai-in-cybersecurity-understanding-ai-technologies**.

Darktrace. “Why Artificial Intelligence is the Future of Cybersecurity.” Darktrace Blog. November 25, 2024. **https://darktrace.com/blog/why-artificial-intelligence-is-the-future-of-cybersecurity**.

CASE STUDY 3.3: TRADELENS—BLOCKCHAIN AS A SOLUTION, BUT WHY DID IT FAIL?

The global shipping industry is a complex web of stakeholders—shippers, carriers, freight forwarders, customs officials, and port operators—each generating and processing vast amounts of documentation. This reliance on paper-based systems and siloed digital platforms is inefficient and has a high risk of errors and fraud. There was a lack of transparency and coordination in the industry.

Instrumental to international shipping logistics is exchanging documents such as bills of lading, customs declarations, and invoices. These documents track shipments, confirm ownership, and ensure compliance with trade regulations. The traditional process involves intermediaries, redundant manual data entry, verification, and approval steps. With shipments passing through multiple jurisdictions, discrepancies, and errors can lead to lengthy delays. The need for a secure, verifiable, real-time information exchange system became increasingly evident as global trade expanded.



Sven Hansche/Shutterstock
On a loaded containership, each container has a variety of important shipping documents that need to be managed.

Blockchain as a Solution

Recognizing these inefficiencies, Maersk, the world's largest container shipping company, partnered with IBM in 2016 to develop TradeLens, a blockchain-based platform to digitize and streamline global trade logistics. Blockchain's decentralized and immutable ledger could provide a standard, trustworthy record, improving transparency, security, and efficiency. TradeLens aimed

to:

- ***Reduce paperwork.*** Digitalize shipping documentation to lower processing times and costs.
- ***Enhance security.*** Provide an immutable record of transactions, reducing fraud and errors.
- ***Increase transparency.*** Allow real-time tracking of shipments and secure data sharing among authorized stakeholders.
- ***Enable AI optimization.*** Leverage blockchain-stored data for intelligent contract automation, predictive analytics, and operational improvements using AI.

By offering these benefits, blockchain had the potential to revolutionize the shipping industry.

An Unexpected Failure

Despite early successes—processing over 70 million containers and 35 million documents—TradeLens shut down in January 2023. Why did a seemingly perfect solution fail?

1. ***Industry resistance and competitive barriers.*** The shipping industry is highly competitive and fragmented. While TradeLens required stakeholder collaboration, many companies were reluctant to join a platform controlled by Maersk, a major competitor. Competitors feared Maersk's dominance and perceived TradeLens as an attempt to monopolize trade data rather than a neutral industry standard.
2. ***Lack of regulatory support.*** Many countries still mandate physical documentation for trade transactions, limiting blockchain's ability to replace traditional processes. Without widespread regulatory acceptance, TradeLens struggled to gain full adoption across global trade networks.
3. ***Interoperability issues.*** Multiple competitor platforms emerged, such as CargoX, Blockshipping GSCP, CargoSmart, and GSBN. These platforms did not share information, leading to fragmentation rather than standardization. Few carriers participated in multiple platforms, reducing the network effects that blockchain technology relies on.
4. ***Integration challenges.*** TradeLens required seamless integration with existing IT systems used by ports, shippers, and customs agencies. However, many stakeholders operated on outdated legacy systems, making it costly and complex to transition to blockchain-based operations.
5. ***Commercial viability issues.*** While blockchain provided security and transparency, transaction costs and processing delays arose due to the need to replicate data across multiple computers and servers. The industry's slow adoption rates meant TradeLens failed to generate sufficient commercial interest to sustain operations.

Lessons from Failure

What lessons can be learned from the TradeLens failure?

1. ***Blockchain alone is not enough.*** While blockchain can enhance security and transparency, it

requires widespread industry buy-in to be effective. Without an industrywide commitment, its benefits are limited.

2. ***Industry governance matters.*** A neutral, industrywide consortium—rather than a single company's initiative—might have gained broader acceptance. Maersk's control over TradeLens deterred competitors from participating.
3. ***Regulatory alignment is key.*** For blockchain to succeed in global trade, regulatory frameworks must evolve to support digital documentation and smart contracts.
4. ***Interoperability must be addressed.*** Competing blockchain platforms need standardization and interoperability to maximize network effects and adoption.

Conclusion

TradeLens demonstrated that blockchain could solve key inefficiencies in global trade but failed due to industry resistance, regulatory barriers, lack of interoperability, and integration challenges. Future efforts in digitizing global trade must focus on technological innovation, fostering collaborative, neutral governance models, and addressing regulatory and interoperability concerns.

Discussion Questions

1. What factors prevented the widespread adoption of TradeLens, and how could a blockchain-based system gain industrywide acceptance in a competitive market?
2. How do existing trade regulations hinder the adoption of digital solutions like blockchain, and what steps could be taken to align regulatory frameworks with technological advancements?
3. How did the emergence of multiple blockchain platforms contribute to TradeLens' failure, and what strategies could improve interoperability across competing solutions in global trade?
4. How did Maersk's involvement affect TradeLens' credibility among industry players, and what alternative governance models might have encouraged broader participation?
5. While blockchain improved security and efficiency, what commercial and operational challenges made TradeLens unsustainable, and how could future initiatives address these issues to create a viable business model?

Case References

Gbi, Ivan. "Why IBM and the Maersk Group Venture Was Doomed to Fail." *LTO Network Blog*, October 30, 2018. **https://medium.com/ltonetwork/why-ibm-and-the-maersk-group-venture-was-doomed-to-fail-the-state-of-b2b-blockchain-adoption-dfff01e4b6fc**

"Tradelens Discontinues Operations. Why You Should Care." *Forbes*, December 5, 2022.

https://www.forbes.com/sites/loracecere/2022/12/05/tradelens-discontinues-operations-why-you-should-care/?sh=7b631524cec5

ENDNOTES

[2] Alder, Steve. “UHG Increases Change Healthcare Data Breach Victim Count to 190 Million.” *HIPAA Journal*, January 25, 2025. **https://www.hipaajournal.com/change-healthcare-responding-to-cyberattack/**

[3] Amazon Web Services. "IAM Roles." *AWS Identity and Access Management User Guide*. Last modified January 15, 2025. **https://docs.aws.amazon.com/IAM/latest/UserGuide/id_roles.html**

[4] Fink, Camille. “Hurricane Ida Response and Recovery in Louisiana.” *Health Facilities Management*, September 17, 2021. **https://www.hfmmagazine.com/articles/4284-hurricane-ida-response-and-recovery-in-louisiana**

[5] Office of Cybersecurity, Energy Security, and Emergency Response. “Colonial Pipeline Cyber Incident.” *U.S. Department of Energy*. **https://www.energy.gov/ceser/colonial-pipeline-cyber-incident**

[6] National Institute of Standards and Technology. “The NIST Cybersecurity Framework (CSF) 2.0.” Last modified February 26, 2024. **https://doi.org/10.6028/NIST.CSWP.29**

[7] Lohr, Steve. “For Big-Data Scientists, ‘Janitor Work’ Is Key Hurdle to Insights.” *The New York Times*, August 17, 2014. https://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-janitor-work-is-key-hurdle-to-insights.html

[8] Editverse. *Data Cleaning Techniques: Ensuring uality in your 2024–2025 Research*, August 2024. **https://editverse.com/data-cleaning-techniques-ensuring-quality-in-your-2024-2025-research/**

[9] Thota, Prahalad. “Inside Wells Fargo’s Data Management System.” *ET CIO*, October 29, 2019. https://cio.economictimes.indiatimes.com/news/strategy-and-management/inside-wells-fargos-data-management-system/71799797

[10] **https://gdpr.eu/fines/_**; Wolford, Ben. “What are the GDPR Fines?” *GDPR.eu*; Data Privacy Manager. “20 Biggest GDPR Fines So Far [2025].” *Data Privacy Manager*, February 12, 2024. **https://dataprivacymanager.net/5-biggest-gdpr-fines-so-far-2020/**

[11] California Privacy Protection Agency. (2024, January 2). *California Consumer Privacy Act Regulations*, January 2, 2024. **https://cppa.ca.gov/regulations/pdf/cppa_regs.pdf**

[12] California Privacy Protection Agency. *California Consumer Privacy Act Regulations*, January 2, 2024. **https://cppa.ca.gov/regulations/pdf/cppa_regs.pdf**

[13] California Privacy Protection Agency. *California Consumer Privacy Act Regulations*, January 2, 2024. **https://cppa.ca.gov/regulations/pdf/cppa_regs.pdf**

[14] California Department of Justice. “Attorney General Bonta Announces Settlement with Sephora as Part of Ongoing Enforcement of California Consumer Privacy Act.” *State of California Department of Justice*, August 24, 2022. **https://oag.ca.gov/news/press-releases/attorney-general-bon**

ta-announces-settlement-sephora-part-ongoing-enforcement

[15] European Parliament. "EU AI Act: First Regulation on Artificial Intelligence." *European Parliament*, June 8, 2023. **https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence**

[16] Californians for Consumer Privacy. "Text of the California Privacy Rights Act (CPRA) of 2020." **https://www.caprivacy.org/cpra-text/**

[17] NASA. *Advanced Computing: Powering NASA's Science and Engineering Discoveries*, March 2021. **https://www.nas.nasa.gov/SC21/assets/pdf/NASA_AdvancedComputing_2021-03.pdf**

[18] Google. "Positive Energy: Belgian Site Becomes First Google Data Center to Add On-Site Solar." *Google Sustainability*, September 2018. **https://sustainability.google/operating-sustainably/stories/belgium-solar/**; St. John, Jeff. "Google Plans to Build Gigawatts of Clean Power and Data Centers Together." *Canary Media*, December 10, 2024. https://www.canarymedia.com/articles/clean-energy/google-has-a-20b-plan-to-build-data-centers-and-clean-power-together

[19] Green Mountain. "Sustainability." **https://greenmountain.no/why-green-mountain/renewable-power/**

[20] Ungleich Glarus AG. "Data Center Light: 99.9% Hydropower." **https://datacenterlight.ch/en-us/cms/hydropower/**

[21] Terrell, Michael. "A First-of-Its-Kind Geothermal Project Is Now Operational." *The Keyword*, November 28, 2023. **https://blog.google/outreach-initiatives/sustainability/google-fervo-geothermal-energy-partnership/**

[22] Meta. "New Geothermal Energy Project to Support Our Data Centers." *Meta Newsroom*, August 26, 2024. https://about.fb.com/news/2024/08/new-geothermal-energy-project-to-support-our-data-centers/

[23] "Big Tech Pours $1.5bn into Nuclear Reactors to Power AI Boom." *Financial Times*, February 19, 2025. https://www.ft.com/content/2d84198e-7eeb-4154-bbf2-9a469b0cc700; "Nuclear-Powered Data Centers: When Will SMRs Finally Take Off?" *Data Center Knowledge*, January 22, 2025. **https://www.datacenterknowledge.com/energy-power-supply/nuclear-powered-data-centers-when-will-smrs-finally-take-off-**

[24] European Parliament and Council. "Directive (EU) 2022/2464 of the European Parliament and of the Council of 14 December 2022 Amending Regulation (EU) No 537/2014, Directive 2004/109/EC, Directive 2006/43/EC and Directive 2013/34/EU, as Regards Corporate Sustainability Reporting." *Official Journal of the European Union*, December 16, 2022. **https://eur-lex.europa.eu/eli/dir/2022/2464/oj/eng**

[25] Climate Neutral Data Centre Pact. *Climate Neutral Data Centre Pact*. (n.d.). **https://www.climateneutraldatacentre.net/**

[26]

[26] Allison, Ian. “Guardtime Secures Over a Million Estonian Healthcare Records on the Blockchain.” *International Business Times*, March 3, 2016. https://www.ibtimes.co.uk/guardtime-secures-over-million-estonian-healthcare-records-blockchain-1547367

[27] de Best, Raynor. “Transaction Speed Ranking of 45 Crypto—Including DeFi and Metaverse—in 2025.” *Statista*, February 5, 2025. https://www.statista.com/statistics/944355/cryptocurrency-transaction-speed/

[28] Arrow Electronics. “Sensors and Edge AI: How They Work Together to Provide Autonomous Driving.” Last modified May 17, 2022. https://www.arrow.com/en/research-and-events/articles/sensors-and-edge-ai-how-they-work-together-to-provide-autonomous-driving

[29] Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H. R., Albarqouni, S., Bakas, S., et al. (2020). “The Future of Digital Health with Federated Learning.” *npj Digital Medicine* 3(1), 19.

[30] Daniel, R. M., & Shevina, V. (2024). “Relationship between Ethical Business and Customer Retention at Tech Industry.” *International Research Journal of Modernization in Engineering Technology and Science*, 6(10), 1501–1505.

CHAPTER 4
Machine Learning



wrightstudio/123RF

Learning Objectives

- **LO4-1** Describe the three basic ways machines learn, and distinguish how they differ in their use of data.
- **LO4-2** Explain how machine learning is applied in forecasting, classification, segmentation, automation, and other ways across key business functions.
- **LO4-3** Describe ethical concerns posed by machine learning.

INTRODUCTION

This chapter introduces standard machine learning techniques used in business, specifically those that do not rely on neural networks. We start by looking at the three basic ways machines learn and distinguish how they differ in their use of data. We then examine how machine learning is applied to improve forecasting, assess risk, segment customers, and streamline operations. We also examine the concerns associated with machine learning, particularly the risks of embedded bias, as seen in several high-profile industry cases. A **Technical Supplement** at the end of the chapter provides more detailed explanations for interested readers. Machine learning algorithms associated with neural networks and generative AI are covered in **Chapter 5** and **Chapter 6**, respectively.

HOW MACHINE LEARNING MODELS LEARN (LO4-1)

Machine learning began as a tool for basic data analysis and pattern recognition. Today, it drives strategic planning and operational efficiency across industries—from finance and health care to retail and manufacturing. Machine learning algorithms learn from data in different ways, depending on the type of problem being solved and the data available:

- **Supervised learning** is used when the data includes known outcomes (e.g., past sales, product returns). It's useful for forecasting, risk scoring, and classification tasks.
- **Unsupervised learning** is used when we don't know the outcomes in advance. These algorithms group or structure data without guidance, making them ideal for customer segmentation or identifying unusual patterns.
- **Reinforcement learning** involves learning through trial and error. While less common in standard business analytics, it is used in areas like dynamic pricing, robotics, and supply chain optimization.

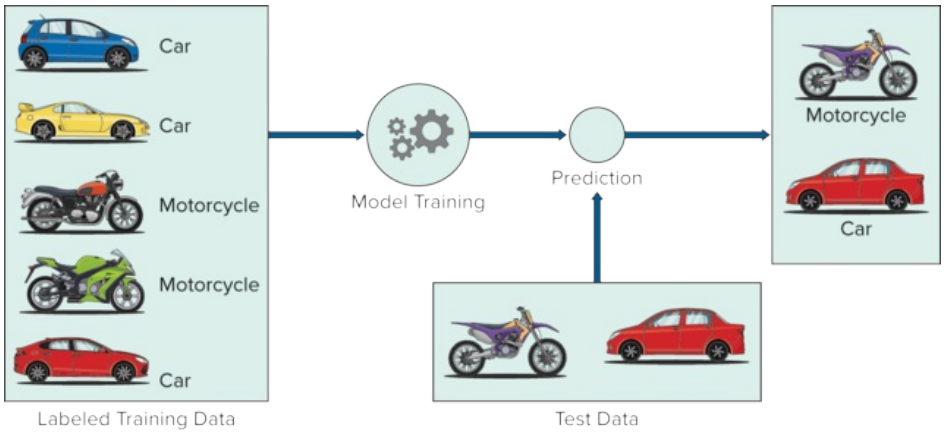
Understanding these categories helps organizations pick the right tool for the problem—and know what kind of data is needed.

Supervised Learning

Supervised learning is similar to a student learning under the guidance of a teacher. This type of machine learning relies on a training dataset with labeled data that "teaches" the algorithm to make predictions or classify data. The "supervisor" provides the algorithm with input-label pairs, and the machine learns to map the inputs to the correct labels. **Figure 4.1** shows cars and motorcycles as the input-label pairs. In business, labels might be "Spam" and "Not Spam" for emails or "Good Credit Risk" or "Bad Credit Risk" for loan applications.

FIGURE 4.1 Supervised Machine Learning

In supervised learning, the algorithm is trained using data tagged with the correct labels. After the model is trained, new test data is fed into the model to validate its accuracy.



Businesses often turn to supervised learning for tasks such as credit scoring, where historical data can train an algorithm to predict a customer's creditworthiness, or for customer churn analysis, where patterns from previous customer behaviors inform predictions on current customer retention.

In 2012, Target gained attention for its advanced use of supervised learning in marketing analytics.[1] It sought to grow its customer base, working on the theory that significant life events like marriage, pregnancy, or moving provided an opportunity to shift brand loyalty. Marital status, births, and changes in residence are available through public records. The company decided to look at the past buying habits of women who had given birth to see if it could identify expectant mothers so it could market to them during their pregnancy. Target labeled these women in its database and then used supervised learning to identify purchasing patterns. It identified 25 products that expectant mothers changed their buying patterns on during pregnancy. The changes happened at specific times during their pregnancy, giving Target good insights into their pregnancy stage and what they might be looking to purchase next. These included:

- Switching from scented to unscented lotions and personal care products early in the pregnancy.
- No longer buying alcohol.
- Reducing purchases of caffeinated beverages.
- Changing pain relief medications to acetaminophen-based ones.
- Reducing purchases of processed and deli meats.

This allowed Target to build a model to predict customer pregnancies and estimate the due dates. The company used this information to send timely, stage-specific coupons for baby products, aiming to secure immediate sales and hopefully generate long-term customer loyalty.

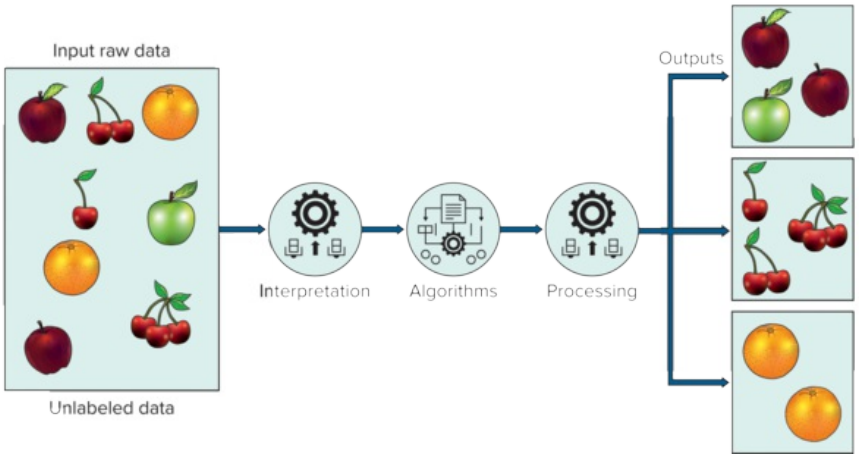
Unsupervised Learning

In contrast to supervised learning, unsupervised learning is like a student learning without explicit

instruction, finding patterns and structure in the data independently. This type of learning deals with unlabeled data and is adept at identifying hidden patterns or intrinsic structures within the data set and separating them into groups. It looks for patterns and groupings in data on its own—without being told the correct answers. It's like sorting puzzle pieces into piles based on shape or color, even if you don't know what the finished picture is supposed to look like. Unsupervised learning algorithms are handy for segmenting customers for targeted marketing campaigns or in detecting behavior that could indicate fraud or system failures. **Figure 4.2** illustrates the process.

FIGURE 4.2 Unsupervised Machine Learning

In unsupervised learning, the algorithm looks at a mix of data—like apples, cherries, and oranges—and figures out how to group similar items together on its own. It's not told what the items are in advance; it just learns to spot patterns and similarities.



Walmart looked at sales at its locations in the path of Hurricane Charley in 2004 using an unsupervised machine learning technique to develop a strategy for stocking and promoting items when hurricanes were announced.[2] As you might predict, items included bottled water, batteries, duct tape, generators, beer, and flashlights. Surprisingly, it also found that stores sold out of Strawberry Pop-Tarts. Walmart now strategically places Strawberry Pop-Tarts on endcaps and at checkout counters in anticipation of hurricanes, increasing sales seven times higher than usual. Its analysis of buying patterns allows it to avoid stockouts and maximize its profits in natural disasters.

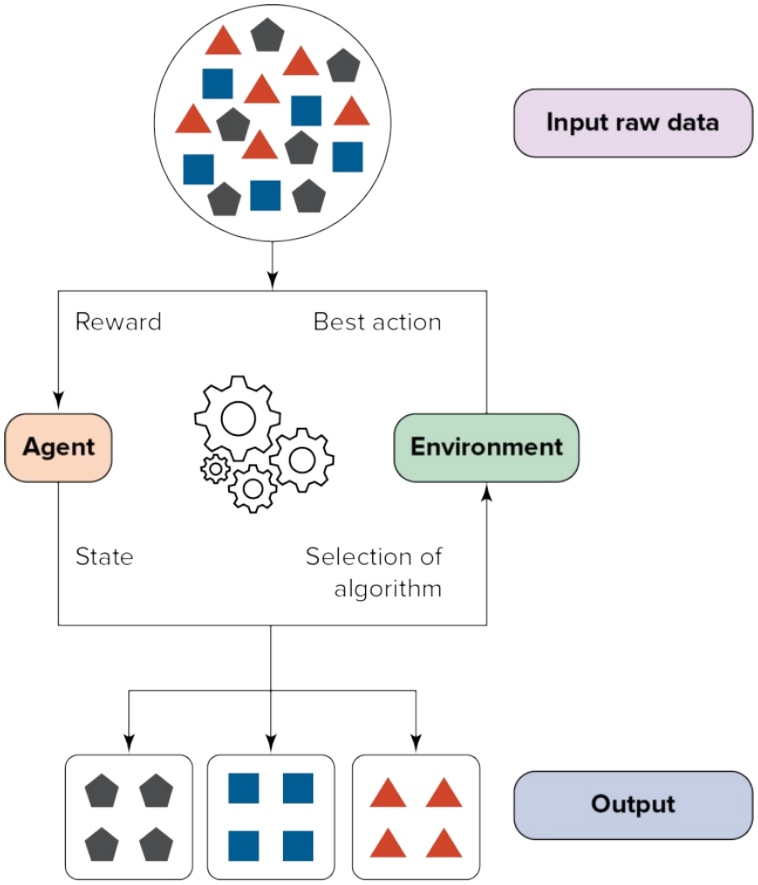
Reinforcement Learning

Reinforcement learning is a behavioral learning model. Reinforcement learning teaches a system (called an *agent*) to make better decisions through trial and error. The agent takes actions, gets feedback (like rewards or penalties), and learns from its experiences to improve over time. This type of learning is often associated with robotics and gaming, where an algorithm learns to make a sequence of decisions that will maximize a cumulative reward, like a high score. **Figure 4.3**

illustrates the process.

FIGURE 4.3 Reinforcement Learning with a Classification Task

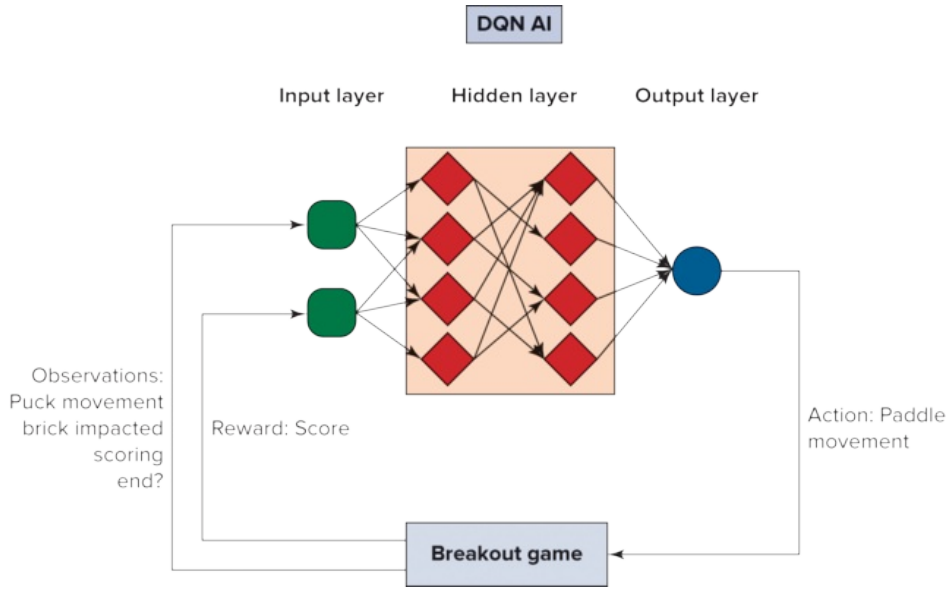
The agent receives raw data containing objects of different shapes and colors. Through trial and error, it selects actions to classify them correctly. With each attempt, it receives feedback (rewards or penalties) based on its degree of success in the task and refines its decision-making process, gradually improving its ability to separate the objects efficiently.



Reinforcement learning does not require a neural network to function, but neural networks can enhance the performance of a reinforcement learning algorithm in solving complex problems. For example, the AI research lab DeepMind trained an algorithm to master the classic Atari game Breakout.[3] Using reinforcement learning with a neural network, its AI, DQN (Deep Q-Network), learned to play Breakout solely by observing the screen pixels and scoring (see **Figure 4.4**). The AI was given no prior information or instructions about the game mechanics, only the goal to maximize its score. Over time, learning from successes and failures, DQN learned to play at the level of a professional human game tester and developed an unknown strategy of tunneling the ball above the bricks. Overall, DQN was tried on 49 Atari games, and on over half of them, it achieved scores of more than 75 percent of the professional human game testers.

FIGURE 4.4 Reinforcement Learning in Action

This diagram illustrates how DeepMind's AI mastered the game of Breakout by interacting with the game environment, receiving rewards in the form of higher scores and penalties by missing the puck, and continuously improved its strategy through trial and error.



In the business context, reinforcement learning can optimize inventory levels by discovering the best stocking strategies to minimize stockouts or enhance personalization to user interactions on a website by seeking to maximize click-throughs.

MACHINE LEARNING IN BUSINESS FUNCTIONS (LO4-2)

Businesses adopt machine learning techniques because they have proven effective in solving challenges from generating more accurate predictions to making better decisions, operating more efficiently, and understanding their customers better. This section explores how organizations apply machine learning algorithms, organized by the specific outcomes they aim to achieve:

- ***Data management and preprocessing.*** High-quality data is foundational for the effective use of AI. Organizations use machine learning algorithms to engage in **data preprocessing** to clean and prepare data for further analysis.
- ***Prediction and forecasting.*** Organizations use machine learning to engage in **prediction and forecasting** to anticipate demand, sales, or the levels of a metric of interest.
- ***Classification and risk assessment.*** Organizations use machine learning to conduct **classification and risk assessment** to identify the most likely categorical outcomes, such as defaulting, changing service providers, or having a medical condition.
- ***Customer segmentation and personalization.*** Organizations use machine learning to engage in **customer segmentation and personalization**, dividing customers into distinct groups to effectively tailor products, services, and marketing efforts.
- ***Process optimization and automation.*** In **process optimization and automation**, organizations use machine learning to streamline processes and automate routine tasks.

This section introduces various machine learning techniques that do not require neural networks and briefly describes their applications in business. A Technical Supplement is available at the end of this chapter for readers interested in a more in-depth understanding of these algorithms. **Table 4.1** lists standard machine learning algorithms and their primary business functions. Each entry in the table includes the algorithm's core purpose and typical business applications, helping readers identify tools that align with specific business challenges. Note that algorithms may be used for more than one business function.

Table 4.1 **Standard Machine Learning Algorithms and Their Primary Business Functions**

Algorithm	Primary Business Functions	Primary Purpose	Common Business Applications
k-means clustering	Data management and preprocessing; customer segmentation and personalization	Group data based on similarity	Customer segmentation, market research, inventory categorization
k-nearest neighbors (*k*-NN)	Data management and preprocessing	Classify based on similarity to nearest data points	Customer classification, recommendation systems, imputing missing data
Principal component analysis (PCA)	Data management and preprocessing	Reduce data complexity and identify patterns	Feature reduction, exploratory data analysis, market research
Linear regression	Prediction and forecasting	Predict continuous outcomes	Sales forecasting, pricing models, risk assessment
Decision trees	Prediction and forecasting; classification and risk assessment	Classify or predict using decision rules	Credit scoring, customer segmentation, operational decisions
Random forests	Prediction and forecasting; classification and risk assessment	Ensemble predictions from multiple decision trees	Fraud detection, customer behavior prediction, forecasting
Logistic regression	Classification and risk assessment	Predict binary outcomes	Churn prediction, fraud detection, marketing response likelihood
Support vector machines (SVMs)	Classification and risk assessment	Find optimal boundaries for classification	Text classification, image recognition, customer sentiment analysis
Collaborative filtering	Customer segmentation and personalization	Recommend items based on user similarity	Personalized product/content recommendations
Time series analysis	Process optimization and automation	Forecast values over time	Inventory management, economic forecasting, maintenance scheduling

Data Management and Preprocessing

Proper data management and preprocessing are critical for the most effective use of machine learning. They ensure raw data is transformed into a suitable format for analysis, improving the quality and performance of machine learning models by addressing inconsistencies and standardizing data structures. This process includes:

- ***Merging databases.*** Combining data from multiple sources into a single, unified repository is a key data transformation component. This process often involves data cleaning, deduplication, and standardizing database formats to ensure the merged data is accurate and coherent.
- ***Data transformation.*** Converting data into appropriate formats or structures through **data transformation** facilitates further analysis. Categorical variables may be encoded into numerical formats. Numerical variables may be rescaled using normalization or standardization. **Normalization**, also known as min-max scaling, rescales data to a specific range, typically between 0 and 1. This process adjusts each data point proportionally within this range, ensuring that features contribute equally to the analysis. It is useful when the data does not follow a normal distribution and is beneficial for algorithms that do not assume any data distribution. **Standardization**, also referred to as Z-score normalization, transforms data to a normal distribution with a mean of 0 and a standard deviation of 1. Each data point's value reflects how many standard deviations it is from the mean. Many machine learning techniques perform best when input variables are on similar scales.
- ***Dimensionality reduction.*** Corporate databases can be huge, with many correlated variables. Using **dimensionality reduction**, machine learning algorithms can consolidate and reduce the number of variables under consideration, simplifying models and increasing the speed of analysis.
- ***Feature engineering.*** **Feature engineering** involves creating new features or modifying existing ones to better represent the underlying problem to the predictive models.

Common Machine Learning Algorithms

The following are some of the common machine learning algorithms used in data management and preprocessing:

- ***k-means clustering.*** While primarily a clustering algorithm, **k-means clustering** can be used in preprocessing to identify and remove outliers or to discover inherent groupings in the data, which can inform feature engineering and selection. The variable, k, represents the number of clusters to create and is chosen by the user.
- ***Imputation methods.*** **Imputation method** techniques such as *k-nearest neighbors (k-NN)* imputation or regression imputation are employed to handle missing data by estimating and filling in missing values based on the similarity to other data points. Here, the variable, k, represents the number of nearby data points to use to fill in the value and is set by the user.

- ***Principal component analysis.*** **Principal component analysis (PCA)** is a technique for dimensionality reduction that transforms correlated variables into a smaller number of uncorrelated variables called *principal components*. This simplification helps reduce a dataset's complexity while retaining most of the original variance.

Incorporating these algorithms and techniques during the data preprocessing phase is essential for building robust and efficient machine learning models.

Example: Spotify and PCA

The exact techniques that streaming services use in their recommendation systems are often proprietary. However, based on a 2022 paper by Aditya Raman, we can explore how Spotify could use PCA as part of its music recommendation system.[4]

Each song in Spotify's catalog has data on its audio features like tempo, rhythm, harmony, danceability, and energy. It also has song information regarding the artists, release date, genre, and record label. PCA combines this data with user engagement data (e.g., skip rate, play count, liked songs, etc.) into a smaller set of principal components. By condensing many variables into a few uncorrelated components, PCA helps reveal patterns in user preferences more clearly. This allows Spotify to identify similar songs and listener profiles more easily, improving the accuracy of personalized playlists and music recommendations.



Eugenio Marongiu/Shutterstock

Spotify uses AI to recommend music based on what you've listened to before. Its algorithms help you discover new songs and artists that match your tastes.

Prediction and Forecasting

Managing a business would be easy if we knew what was going to happen. We would choose the decision alternatives that give the best outcome. While no algorithm can predict the future with certainty, machine learning has become an essential tool for improving the accuracy of forecasts involving numerical data. These models analyze patterns in historical data to anticipate future trends, helping businesses make better, data-driven decisions. Here are some of its business applications:

- ***Demand and sales forecasting.*** Retailers, manufacturers, and service providers use machine learning to predict demand, regulate inventory levels, reduce storage costs, and staff appropriately.
- ***Financial forecasting.*** Financial institutions use machine learning to project revenue streams, identify risks, and identify investment opportunities. Organizations in all sectors analyze historical financial data and economic indicators for budgeting and strategic planning.
- ***Quality control.*** By understanding the expected range of values for a product or process, companies can swiftly identify deviations from these norms. This prompt detection enables them to halt operations, determine the cause, and implement corrective actions to maintain quality standards.

Common Machine Learning Techniques

The following are some of the common machine learning techniques used in prediction and forecasting:

- ***Linear regression.*** **Linear regression** models the relationship between a dependent variable and one or more independent variables. It is useful for forecasting continuous outcomes like demand, revenue, or financial valuation and for identifying and understanding variables that affect those outcomes.
- ***Decision trees and random forests.*** **Decision trees** split data based on feature values to make predictions. **Random forests** combine multiple decision trees to enhance accuracy and avoid overfitting. These methods are well-suited for complex datasets with numerous variables.

Example: Zillow and Linear Regression

Online real estate marketplace Zillow employs linear regression with other machine learning models in its property valuation tool, famously known as Zestimate.[5] It estimates home market values using variables like square footage, location, and the number of bedrooms and bathrooms. By correlating these factors, Zestimate provides homeowners and potential buyers with a quick and reasonably accurate estimate of a property's market value.

Example: Heart Attack Prediction and Decision Trees

Decision trees are crucial in quickly and accurately assessing the probability of a heart attack in

emergency rooms.[6] Clinicians can rapidly categorize patients based on symptoms, medical history, and vital signs to identify high-risk patients. This helps provide prompt, appropriate care for critical cases, streamlining the decision-making process during these time-sensitive situations.

Classification and Risk Assessment

Beyond numerical forecasting, businesses also use machine learning to classify outcomes and assess risks. Classification models help companies predict which outcome is most likely to occur, such as whether a customer will make a purchase, switch service providers, or default on a loan. Accurately identifying categories and assessing associated risks are critical for decision making. Machine learning is a powerful tool in this domain, enabling organizations to analyze complex datasets and predict categorical outcomes with high precision. Here are some of its business applications:

- ***Credit risk assessment.*** Financial institutions evaluate the likelihood of loan defaults by analyzing applicants' credit histories, transaction behaviors, and other relevant data. This increases the accuracy of credit scoring models, leading to better lending decisions.
- ***Fraud detection.*** Machine learning models detect fraudulent activities in real time by identifying unusual transaction patterns in banking and credit card operations. The cardholder is automatically contacted to verify flagged transactions.
- ***Customer churn prediction.*** Machine learning predicts which customers will likely discontinue services. Companies can implement targeted retention strategies by analyzing customer behavior and engagement metrics.

Common Machine Learning Techniques

The following are some of the common machine learning techniques used in classification and risk assessment:

- ***Logistic regression.*** **Logistic regression** is similar to linear regression except it predicts the probability of a binary outcome, making it suitable for scenarios like default prediction (yes/no) or fraud detection (fraudulent/non-fraudulent).
- ***Decision trees and random forests.*** Decision trees and random forests can also be used with categorical data, classifying by splitting the data based on feature values.
- ***Support vector machines.*** **Support vector machines (SVMs)** are supervised learning models that analyze data for classification and regression analysis. They work best in scenarios where the data has distinct separations between classes.

By integrating machine learning into classification and risk assessment processes, businesses can enhance their decision-making capabilities, proactively manage potential risks, and maintain a competitive edge in their respective markets.

Example: Booking.com and Logistic Regression

The COVID-19 pandemic led to a surge in travelers needing to modify their reservations on **Booking.com**. To address this, **Booking.com** incorporated logistic regression into its predictive model to estimate the probability of a user changing their travel dates.[7] The model analyzed user demographics, travel preferences, and interaction history to predict booking modifications. **Booking.com** adjusted its interface and tailored recommendations for users most likely to modify their reservations, enhancing user experience and engagement. This approach improved the platform's efficiency and profitability by reducing cancellations and increasing customer satisfaction.

Example: KFC and SVM

KFC wanted to measure customer response to its new Salted Egg Chicken menu, promoted by celebrity Raditya Dika in Indonesia. The company collected and classified comments from YouTube and Twitter using SVM algorithms.[8] KFC found that celebrity endorsements got more people to try the menu and increased public awareness. However, it also showed that the customer perception of the quality of the new menu items did not change. SVM demonstrated its effectiveness in *sentiment analysis*, providing KFC with valuable insights into customer reactions and the overall impact of their marketing strategies.

Example: NEC Labs America and Random Forests

NEC Labs America utilized random forest techniques in its bioinformatics research to classify disease samples, identify disease-associated genes, and predict protein-protein interactions.[9] This aided in the understanding of disease mechanisms and developing diagnostic tools. Machine learning was also used to screen drugs as possible treatments for diseases.

Customer Segmentation and Personalization

Machine learning enables businesses to segment their customer base effectively and deliver crafted messages and experiences that resonate with each customer segment. For example, machine learning algorithms analyze purchasing behavior, demographics, and engagement patterns to identify distinct customer groups. This segmentation allows businesses to strategically offer products to meet the needs and desires of each group. Companies also use machine learning algorithms to deliver personalized content and product recommendations to each customer to enhance customer engagement and increase conversion rates. For example, if a person sees a product they might buy, a show that interests them, or a song by an artist they like when they log on a website, its placement is most likely due to machine learning.

Common Machine Learning Techniques

The following are some of the common machine learning techniques used in customer segmentation and personalization:

- ***Clustering algorithms (e.g., K-means).*** These algorithms group customers based on similarities in their data, such as purchasing habits, preferences, or demographics, permitting targeted marketing strategies.
- ***Collaborative filtering.*** Often used in recommendation systems, **collaborative filtering** predicts a customer's interests by analyzing preferences from similar users, enabling personalized product or content suggestions.
- ***Decision trees.*** Decision trees assist in segmenting customers by making sequential decisions based on their attributes, aiding in the development of tailored marketing campaigns.

Example: Resource Allocations for Emergency Services Using K-Means Clustering

Emergency services in Yvelines, a department in the western part of the Île-de-France region in northern France, performed a K-means clustering analysis to improve its fire and rescue services.[10] It grouped fire operations data clusters to better understand patterns and trends in emergency calls. With this knowledge, it allocated resources to fire stations better, reducing response times and costs.

Process Optimization and Automation

Machine learning is employed by businesses to make inventory and supply chain decisions and automate routine tasks. This leads to increased efficiency, better service levels, and cost reduction. For example, machine learning is used in the following:

- ***Supply chain optimization.*** Machine learning models predict demand changes, optimize inventory levels, and streamline logistics, resulting in a more responsive and efficient supply chain.
- ***Predictive maintenance.*** Machine learning algorithms can predict potential failures before they occur using machine sensor data, allowing for proactive maintenance scheduling and reducing downtime.
- ***Robotic process automation (RPA).*** Combining machine learning with automation enables the computerization of tasks that involve decision making, such as processing invoices or managing customer inquiries, thereby improving accuracy and speed.

Common Machine Learning Techniques

The following are some of the common machine learning techniques used in process optimization and automation:

- ***Regression analysis.*** Regression analysis is used to identify relationships between variables, aiding in forecasting and decision-making processes.
- ***Time series analysis*****.** **Time series analysis** algorithms are essential for modeling and forecasting with temporal data. They are used in applications like demand forecasting for inventory management and maintenance scheduling.

Ford's Use of AI for Predictive Maintenance

In its manufacturing facilities, Ford utilizes AI algorithms to analyze sensor data from machinery, enabling the prediction of equipment failures before they occur. This proactive approach minimizes unplanned downtime and enhances production efficiency.[11]

Comparing Standard Machine Learning Algorithms

Table 4.2 compares standard machine learning algorithms, highlighting their strengths, limitations, applicable data types, and typical data volume requirements. It is intended to help assess the practicality of each method based on available data and analytical goals. For a more in-depth discussion of each, go to the end-of-chapter technical supplement.

Table 4.2 **Standard Machine Learning Algorithms: Strengths, Limitations, Data Types, and Typical Data Needs**

Algorithm	Key Strengths	Limitations	Data Type	Typical Data Needs
k-means clustering	Simple and efficient; good for well-separated clusters	Requires predefined cluster count; sensitive to scale and outliers	Numerical	Moderate
k-nearest neighbors (k-NN)	Simple and intuitive; effective with enough data	Slow with large data; sensitive to irrelevant features	Numerical, categorical	Low
Principal component analysis (PCA)	Reduces data dimensions, improves visualization	May lose interpretability; assumes linearity	Numerical	Moderate
Linear regression	Easy to implement and interpret	Assumes linearity; sensitive to outliers	Numerical	Low
Decision trees	Easy to understand; handles mixed data	Prone to overfitting; unstable with data changes	Numerical, categorical	Low
Random forests	High accuracy; reduces overfitting	Less interpretable; computationally intensive	Numerical, categorical	Moderate to high
Logistic regression	Fast and interpretable; good for binary classification	Assumes linear boundaries; not good for complex relationships	Numerical, categorical	Low
Support vector machines (SVMs)	Effective in high-dimensional spaces	Requires careful parameter tuning; slow with large datasets	Numerical	High
Collaborative filtering	Highly personalized; improves user engagement	Needs lots of data; cold start problem	Numerical, categorical	High
Time series analysis	Captures trends over time; good for forecasting	Assumes regular patterns; struggles with missing data	Time series	Moderate to high

Machine learning can affect almost every business function. Through predictive analytics, businesses anticipate customer needs, forecast demand, and manage risk. Through automation, they streamline operations, improve supply chains, and respond to changing conditions in real

time. By integrating machine learning into decision-making processes, companies become more data-driven, proactive, and responsive. But as the influence of machine learning grows, so do concerns, which we look at next.

ETHICAL RISKS OF MACHINE LEARNING IN BUSINESS (LO4-3)



As machine learning becomes more embedded in business operations, so do fairness, transparency, and responsible use concerns. These technologies offer value, but they also carry risks—especially when models are trained on biased data or applied without clear oversight.

Algorithmic Bias in Machine Learning

Machine learning systems can unintentionally reflect or amplify historical biases in the data they are trained on. **Chapter 3** introduced Amazon's résumé screening algorithm, which penalized résumés from women based on patterns in past hiring data. This creates a business risk with ethical, legal, and reputational implications. Biased algorithms can result in discriminatory hiring, lending, policing, and customer service practices. These outcomes can violate civil rights laws, invite regulatory scrutiny, and erode trust among employees, customers, and the public.[12]

Strategies for Reducing Bias in Machine Learning Systems

Businesses have a clear responsibility to identify and reduce bias in machine learning models. In practice, this includes:

- Using representative data to reflect the diversity of the populations affected.
- Conducting **bias audits** to test for disparate outcomes across groups.
- Applying **fairness-aware modeling**—algorithmic fairness techniques that adjust model training or outputs to produce more equitable results.
- Generating and carefully using **synthetic data**—artificially generated data used to supplement or balance real datasets, especially when data is limited or imbalanced—to supplement underrepresented groups where appropriate.
- Employing explainable AI (XAI) to understand how decisions are made.
- Engaging stakeholders—from legal and HR to affected communities—in the design process.
- Monitoring models continuously as data and business contexts evolve.

Bias in machine learning reflects societal and organizational patterns. Businesses that treat it

seriously are more likely to build systems that are effective, equitable, and aligned with long-term values.

An Algorithmic Bias Audit Law in New York City

In 2023, New York City implemented Local Law 144, one of the first laws in the U.S. requiring companies to conduct independent bias audits of automated hiring tools.[13] The goal was to improve fairness in employment decisions made by algorithms by requiring annual audits and public disclosure of results. The law marked an important step toward accountability, but its impact has been limited. Narrow definitions and exemptions have meant that machine learning hiring tools remain unaudited. Regulating fairness in AI systems is complex, and developing standards and enforcement mechanisms is ongoing.

Other Concerns

Machine learning poses a number of other concerns. For example, the impacts of workforce displacement due to AI, such as job automation, retraining, and organizational adaptation, will be explored further in **Chapter 7**. Machine learning systems also often rely on large volumes of personal data, which pose privacy, security, and regulatory concerns. We discussed such concerns in **Chapter 3**.

In sum, machine learning can significantly enhance business performance, but fairness, transparency, and accountability principles must guide its use. As adoption grows, so does the responsibility to ensure that AI works for everyone, not at someone's expense.

SUMMARY OF KEY POINTS AND QUICK WINS

1. **How machines learn.** Machine learning algorithms learn from data in different ways. Supervised learning uses labeled data to make predictions; unsupervised learning uncovers patterns in unlabeled data; and reinforcement learning learns through feedback and rewards. Understanding these methods helps match business problems to the right tools.
2. **Applications of machine learning in business.** Machine learning techniques such as regression, classification, and clustering have been around for decades and are widely used across business functions. These methods are used in forecasting, risk assessment, customer segmentation, and operations.
3. **Responsible use of machine learning.** As businesses adopt machine learning, they must address ethical concerns. Bias in data can lead to discriminatory outcomes, and regulations like New York City's Local Law 144 now require accountability. Responsible practices include using representative data, conducting audits, and building explainable models.

QUICK WINS

- **Match learning type to business need.** Understand the differences among supervised, unsupervised, and reinforcement learning. Use this knowledge to select machine learning approaches that align with the nature of your business data and goals.
- **Start with practical machine learning tools.** Focus on machine learning methods that do not require neural networks—such as regression, classification, and clustering—for accessible, proven ways to apply AI in forecasting, segmentation, and process improvement.
- **Build fair, accountable systems.** Audit your data and models for bias. Engage diverse stakeholders, document your AI decision-making processes, and comply with emerging laws. This will protect your brand and build trust.

KEY TERMS

Algorithmic bias. Systematic and unfair outcomes in machine learning caused by biased training data or model design.

Bias audits. Assessments conducted to detect and mitigate biases in AI systems to promote fairness and inclusivity.

Classification and risk assessment. Applying machine learning to categorize data into discrete groups or predict the likelihood of certain outcomes, such as loan defaults or customer churn.

Collaborative filtering. A recommendation method that predicts user preferences based on patterns from similar users.

Customer segmentation and personalization. Grouping customers based on behavior or preferences to tailor marketing, product recommendations, or services.

Data preprocessing. The process of cleaning, transforming, and organizing raw data to make it suitable for machine learning analysis.

Data transformation. Changing the format or structure of data to make it compatible with machine learning models, including encoding or rescaling.

Decision trees. A model that uses branching rules to split data into subsets for prediction or classification.

Dimensionality reduction. Techniques that reduce the number of variables in a dataset while retaining important information.

Fairness-aware modeling. Techniques that adjust models to produce more equitable outcomes for diverse groups.

Feature engineering. Creating or modifying input variables to improve model performance.

Imputation method. Techniques for filling in missing data using estimates based on other available information.

K-means clustering. An unsupervised algorithm that groups data points into a specified number of clusters based on similarity.

Linear regression. A supervised learning method that models the relationship between one or more independent variables and a continuous outcome.

Logistic regression. A supervised algorithm used to predict binary outcomes, such as yes/no or true/false decisions.

Machine learning. A subset of AI where algorithms learn from data to make predictions or decisions without being explicitly programmed.

Normalization (min-max scaling). A technique that scales data to a specific range, typically 0 to 1, to ensure consistency across features.

Prediction and forecasting. The use of machine learning to estimate future outcomes based on

historical data, such as sales or demand.

Principal component analysis (PCA). A dimensionality reduction method that transforms correlated variables into smaller uncorrelated components.

Process optimization and automation. Using machine learning to streamline operations, reduce costs, and improve efficiency through automated decision-making.

Random forests. An ensemble learning method that combines multiple decision trees to improve accuracy and reduce overfitting.

Reinforcement learning. A learning method where an algorithm improves performance by interacting with an environment and receiving feedback in the form of rewards or penalties.

Standardization (Z-score normalization). A scaling method that transforms data to have a mean of 0 and a standard deviation of 1.

Supervised learning. A type of machine learning where the algorithm learns from labeled data to make predictions or classifications.

Support vector machines (SVMs). A supervised learning algorithm that finds the best boundary to separate different categories of data.

Synthetic data. Artificially generated data used to supplement or balance real datasets, especially when data is limited or imbalanced.

Time series analysis. A technique for analyzing data points collected over time to identify trends and make future predictions.

Unsupervised learning. A machine learning approach that identifies hidden patterns or groupings in data without labeled outcomes.

DISCUSSION QUESTIONS

1. **Machine learning types and business use.** How do supervised, unsupervised, and reinforcement learning apply to business problems?
2. **Applying machine learning to business functions.** How can traditional machine learning techniques like regression, classification, and clustering improve decision making in marketing, operations, or finance?
3. **Responsible use of machine learning.** What are businesses' responsibilities when using machine learning systems? Discuss how they can reduce bias and increase transparency in their AI tools.
4. **Algorithmic accountability.** Why is explainability important in machine learning applications? In what kinds of business situations is it most critical?
5. **Regulating fairness in machine learning.** New York City's Local Law 144 aimed to improve fairness in algorithmic hiring by requiring bias audits, yet many tools remain unaudited. What challenges do governments face when trying to regulate fairness in AI systems, and what can businesses do to prepare for future regulations?

EXERCISES

1. **Machine learning types and business use.** Provide examples of when supervised, unsupervised, and reinforcement learning might be used.
2. **Choosing the right machine learning tool.** Consider the following scenarios:
   - Predicting if a customer is going to switch mobile plans.
   - Grouping grocery store customers for better marketing.
   - Estimating demand for new storage units in cities.

   What machine learning techniques might best suit each? What data would you want to answer each question successfully. Explain your reasoning.
3. **Algorithm evaluation practice.** Select two common machine learning algorithms (e.g., decision trees and logistic regression). Compare them in terms of interpretability, accuracy, and ease of implementation for a business-facing team. In what situations would one be preferred over the other?
4. **Process optimization scenario.** A logistics company wants to reduce delivery delays during winter months. Describe how it could apply time series analysis and regression to anticipate issues and optimize delivery routes. What data would it need?
5. **Interpreting real-world use.** Read a recent article or case study about a company using machine learning in operations, marketing, or finance. Summarize what technique was used and why. Reflect on whether the approach was appropriate and what ethical considerations may apply.

CASE STUDY 4.1: NETFLIX RECOMMENDATION SYSTEM

Steve Krug's first law of web usability, "Don't Make Me Think,"[14] sums up Netflix's recommendation system: Users don't want to spend time searching for something to watch. Netflix uses machine learning algorithms to eliminate decision fatigue by creating a personalized content grid that presents viewers with titles they will most likely enjoy upon login. By making content discovery effortless, Netflix increases engagement, reduces churn, and strengthens customer loyalty.

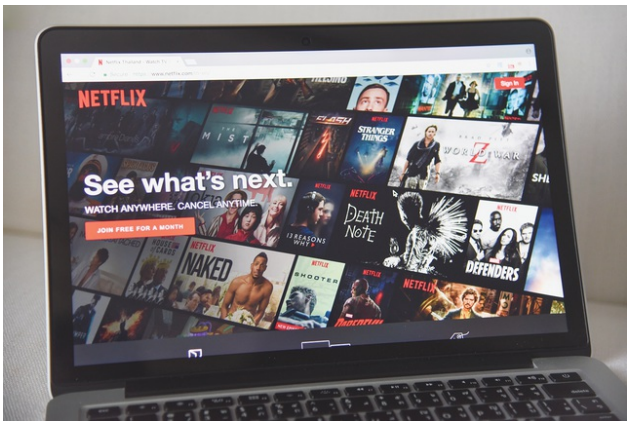
Netflix's History

Founded in 1997 as a DVD rental service, Netflix adapted to changing technology and consumer demand, shifting to streaming in 2007. As competition grew and licensing costs soared, Netflix began producing original content in 2013, starting with *House of Cards*, to strengthen its competitive edge.

Central to its success is machine learning–powered recommendations, which tailor content to each viewer's preferences. The personalized content grid strategically arranges titles, ensuring users quickly find something to watch, minimizing frustration, and reducing the temptation to switch platforms.

The Heart of Netflix: Its Recommender Grid

When users log into Netflix, they see a grid of shows selected just for them. Machine learning algorithms analyze user data and behavior and content metadata to organize rows like "Because You Watched [Title]" or "Top Picks for You." This grid is continuously updated. Viewing history, time of day, recent behavior, and diversity can modify the order of the rows to keep recommendations fresh and engaging. Netflix even selects a title's cover art to increase the chance of selection.



Jesse33/Shutterstock

Netflix's AI recommendation system generates a dynamic selection of shows and movies, making content discovery effortless and engaging.

Ordering titles within a row is primarily based on the predicted probability that a viewer will watch each title. Other factors can influence ordering, including variety to avoid repetition, contextual relevance (e.g., time of day or recent viewing behavior), and the featuring of new or trending titles. Netflix strategically promotes its original content, balancing user preferences with business objectives. The goal is to ensure that rows are both personalized and engaging.

Netflix's recommendation system employs three main approaches in generating the grid:

1. ***Collaborative filtering.*** Netflix uses data from a user to identify other users with similar characteristics and viewing patterns using techniques like K-nearest neighbors. It then uses this group's viewing history to select shows to place on the grid.
2. ***Content-based filtering.*** Netflix recommends items with similar attributes (e.g., genres, actors) to those a user has watched or rated highly.
3. ***Deep learning.*** Netflix enhances recommendations by analyzing complex patterns, considering factors like time of day and device used. Deep learning adds a contextual dimension to recommendations.

The Data Driving the System

Two main types of data are driving the system:

1. ***User data.*** Netflix's recommendation system depends on extensive user data, including viewing history, genres, titles, ratings, searches, pauses, and rewatches, to model preferences. Contextual information, like the time of day, device used, and binge-watching patterns, adjust recommendations dynamically. For example, a viewer watching on a phone at 8 a.m. may see shorter, mobile-friendly content, while a viewer watching on a smart TV in the evening is likelier to indulge in long-form dramas or binge-worthy series.
2. ***Show data.*** Netflix analyzes extensive metadata for each title in its library to create content-

based recommendations. This metadata includes details like genres, cast, directors, release year, and plot summaries. Combined with user data, this analysis enables Netflix to recommend content that aligns with a viewer's tastes while ensuring some diversity in the selection.

Challenges and Ethical Considerations

Netflix must balance personalization with ethical concerns:

- ***Data privacy.*** Like any web-based company that collects sensitive customer data, Netflix incorporates robust protections to prevent breaches and misuse in compliance with data privacy and security regulations.
- ***Algorithmic bias.*** Personalization improves engagement, but it can reinforce biases. Recommender systems favor popular content, making it harder for niche or diverse stories to reach broad audiences. Netflix conducts audits to measure content diversity and avoid generating viewing bubbles.

Balancing personalization and ethical responsibility will remain central to sustaining trust and user satisfaction.

Business Impact

Netflix's recommendation system is responsible for 80 percent of Netflix's viewing hours. Making selection easy has increased viewing hours and guided its original content production. By "not making them think," Netflix has a low churn rate and has cemented its position as a leader in the streaming market, with 301 million paid memberships in over 190 countries.

Discussion Questions

1. How does Netflix's recommendation system contribute to reducing churn and increasing viewer engagement?
2. How does Netflix use machine learning to organize and display content rows in a way that enhances viewer engagement? What role do visual elements like personalized cover art play in this process?
3. What potential advancements in machine learning could further enhance Netflix's recommendation system? Consider areas like mood-based recommendations, social recommendations, or interactive content.
4. Netflix uses extensive metadata for its content, such as genres, cast, and themes. How does this data enhance the accuracy of recommendations, and are there additional types of metadata that could improve the system?

5. With growing expectations for transparency in AI systems, how might Netflix need to adapt its recommendation algorithms to address regulatory and user demands?

Case References

Barwal, D., et al. (2023). "The Impact of Netflix Recommendation Engine on Customer Experience." *AIP Conference Proceedings*, 2736, 060005. **https://doi.org/10.1063/5.0170916**

Gomez-Uribe, C. A., & Hunt, N. (2016). "The Netflix Recommender System: Algorithms, Business Value, and Innovation." *ACM Transactions on Management Information Systems (TMIS)*, *6*(4), 1–19. **https://doi.org/10.1145/2843948**

Manfredi, L. (January 21, 2025). "Netflix Tops 300 Million Subscribers Globally, Adds Record-Breaking 19 Million in Q4." *TheWrap*. **https://www.thewrap.com/netflix-earnings-q4-2024/**

Romero Meza, L., & D'Urso, G. (2024). "User's Dilemma: A Qualitative Study on the Influence of Netflix's Recommender System on Choice Overload." *Psychological Studies*, 69, 349–367. **https://doi.org/10.1007/s12646-024-00807-0**

CASE STUDY 4.2: FLIGHT DELAYS AT JFK

Flight delays are frustrating for passengers and airlines. Research shows that passengers begin feeling aggravated after a 64-minute delay, particularly when they miss connections or face unexpected accommodations. Delays cause headaches for airline operations, leading to cascading delays, increased costs, and logistical challenges.

To protect passengers' rights, the European Union and the United States have regulations requiring airlines to provide assistance or compensation under certain conditions when flights are delayed or canceled. While the EU mandates compensation based on delay duration and distance, U.S. airlines are not required to compensate passengers for weather-related cancellations but must offer refunds for unused tickets. The intent is to hold airlines accountable and ensure fair treatment for passengers during disruptions.

A delay at one airport can affect subsequent flights and increase operational costs. Airlines may need to adjust schedules, reroute crews, or rent replacement aircraft to mitigate disruptions. Predicting and managing delays efficiently is crucial to minimizing financial and logistical impacts.

JFK Airport is one of the world's busiest travel hubs, handling domestic and international flights. Due to its high traffic volume and exposure to adverse weather conditions, the ability to predict flight delays would enhance its ability to handle them effectively. Machine learning algorithms were implemented to predict flight delays. The implementation involved:

1. ***Data collection.*** Researchers collected a year's worth of data, including:
   - Flight schedules and delays
   - Weather conditions
   - Airport congestion levels
   - Air traffic control patterns

   This dataset included information on 458,897 total flights, of which 173,881 were international. While JFK is primarily a passenger hub, almost 13,000 cargo planes landed there.
2. ***Machine learning algorithms.*** Seven algorithms were trained and tested to complete the binary classification of flight delays. These included logistic regression, k-nearest neighbor (k-NN), Gaussian Naïve Bayes, decision tree, support vector machine (SVM), random forest, and gradient boosted tree. These models were selected to compare performance across different machine learning techniques—from simpler, interpretable models (logistic regression, decision trees) to more complex ensemble methods (random forest, gradient boosted trees) known for handling nonlinear relationships in data.
3. ***Analysis.*** Accuracy, precision, and recall were used to evaluate each algorithm's performance. The decision tree algorithm performed the best, with an accuracy of 0.9778.

Implementation

While machine learning models have been tested using JFK flight data, there are no published records of their official implementation at the airport yet. If deployed, such a system could allow airlines and passengers to manage the delay better.

Making predictions in real time poses a series of additional challenges. The systems must promptly accept data updates from many sources (e.g., weather reports and forecasts, flight updates, HR for staffing changes) and process updates fast enough to allow interpretation and action. Machine learning's ability to anticipate delays has demonstrated promising results in research settings. Now, the challenge is to incorporate it in operations effectively. Improvements in handling delays at major hubs would improve travelers' satisfaction levels and allow the airlines to operate more efficiently.

Discussion Questions

1. The study found that decision trees outperformed other models in predicting flight delays at JFK. What characteristics of flight delay data might make decision trees more effective than models like logistic regression or K-nearest neighbors? Under what conditions might other models perform better?
2. Implementing real-time machine learning systems in airport operations presents several logistical and technical challenges. What are the biggest hurdles airlines and airports would face in adopting such a system? How might they address these challenges?
3. How could a real-time flight delay prediction system impact an airline's profitability, cost management, and resource allocation? What financial benefits might airlines see from successfully implementing predictive models for delay management?
4. From a customer perspective, how would real-time delay predictions and proactive communication change the passenger experience? Would knowing about a delay earlier significantly improve passenger satisfaction, or do other factors matter more?
5. If machine learning-based flight delay predictions prove successful at JFK, what would be the key challenges in scaling this technology to other airports? How might predictions differ at smaller regional airports compared to large international hubs?

Case References

IBS Software. "It Takes a Flight Delay of Just 64 Minutes for British Holidaymakers to See Red, Research Finds." August 24, 2023. **https://www.ibsplc.com/news/it-takes-a-flight-delay-of-just-64-minutes-for-british-holidaymakers-to-see-red-research-finds#:~:text=Of%20the%20respondents%2C%2061%20per%20cent%20don%27t,to%20passengers%20%E2%80%93%20with%2045%20per%20cent**

Khanmohammadi, S., Tutun, S., & Kucuk, Y. (2016). “A New Multilevel Input Layer Artificial Neural Network for Predicting Flight Delays at JFK Airport.” *Procedia Computer Science*, 95, 237–244. **https://doi.org/10.1016/j.procs.2016.09.321**

Leach, B. (2023, August 25). “Study Reveals Scale of Flight Delays and Impact on Travellers.” *The Independent*. **https://www.independent.co.uk/travel/flight-delays-travelling-passengers-compensation-b2397885.html**

Port Authority of New York and New Jersey. *2015 Annual Traffic Report*. April 1, 2016.

Tang, Y. (2021). “Airline Flight Delay Prediction Using Machine Learning Models.” In 2021 *5th International Conference on E-Business and Internet (ICEBI 2021)*, October 15–17, 2021, Singapore. ACM, New York, NY, USA, 7 pages. **https://doi.org/10.1145/3497701.3497725**

CASE STUDY 4.3: UPS VEHICLE ROUTING

Amazon Prime and similar services have normalized same-day and next-day delivery, leaving logistics companies scrambling to keep pace. Fast delivery is now a critical decision factor for online shoppers, creating opportunities and significant challenges for logistics providers.

Challenges for Logistics Companies

Meeting rising consumer demands requires sophisticated logistics solutions, but comes with challenges such as:

- ***Operational demands.*** Delivering packages fast and reliably requires inventory management and warehouse systems that can keep the right products in stock and rapidly process orders for delivery.
- ***Route optimization.*** Once the trucks are loaded, AI-driven routing algorithms must adjust dynamically to traffic, weather, and unexpected delays. Routing also needs to consider safety by avoiding high-risk intersections and road conditions.
- ***Scalability.*** Expanding rapid delivery services to rural and underserved areas requires a robust, adaptable distribution network.
- ***Cost and sustainability concerns.*** Faster deliveries often mean more trips with fewer items, increasing cost, fuel consumption, and carbon emissions.
- ***Package tracking.*** Customers expect to track their package throughout the delivery process and receive a notification when delivered on an app.

Logistics firms must incorporate technology, automation, and data analytics to meet these challenges. UPS is tackling them with its AI-powered ORION system.

UPS and ORION: AI for Efficiency

Founded in 1907, UPS has grown into the world's largest package delivery company, operating in more than 220 countries. UPS has heavily invested in technology to maintain its competitive edge, using AI for route optimization to streamline operations, reduce costs, and improve sustainability.

ORION (On-Road Integrated Optimization and Navigation) is a machine learning system that dynamically optimizes driver routes. It processes real-time data on traffic, weather, vehicle load, and driver experience to adjust delivery routes while packages are out for delivery.

The Left-Turn Problem

One of ORION's most notable innovations is its minimization of left turns (or right turns in left-driving countries). Traffic safety research showed that:

- Thirty-six percent of vehicle accidents occur at intersections.
- Left turns are three times more likely to result in fatal crashes involving pedestrians.
- Vehicles making left turns idle longer, increasing delivery time, fuel waste, and CO_2 emissions.

The ORION AI reduced left turns to 10 percent of total turns on average. The result is UPS achieves longer routes in distance but shorter routes in time, improving both safety and fuel efficiency.

Real-Time Optimization and Cost Savings

The initial routing with minimal left turns is a good starting point. UPS has about 100,000 delivery vehicles on the road. For them, ORION's AI processes 30,000 route optimizations per minute, utilizing fleet telematics and big data analytics to reduce miles driven. The results are impressive:

- ***Fuel savings.*** 10 million gallons annually
- ***Carbon reduction.*** 20,000 fewer tons of CO_2 emissions annually
- ***Operational efficiency.*** 350,000 more packages delivered each year
- ***Fleet reduction.*** 1,100 fewer vehicles needed each year
- ***Total miles saved.*** 28.5 million miles annually

Conclusion

When the tire hits the road (pun intended), UPS shines. It is known for its on-time delivery rate of 97.5 percent overall, with 96.5 percent on-time delivery during the busy December delivery period in 2024. The University of Michigan's American Customer Satisfaction Index (ACSI) ranked it number one in its industry for customer satisfaction, and it has been ranked on *Fortune's* World's Most Admired Companies list. UPS's ORION system is an important part of its success.

Discussion Questions

1. How does UPS's use of machine learning for route optimization reflect its strategic approach to innovation? Consider how this technology aligns with the company's broader business objectives.
2. Discuss how ORION's focus on minimizing left-hand turns contributes to operational efficiency and safety. What are the implications of this approach for urban and suburban delivery routes?
3. How does improving operational efficiency through route optimization indirectly enhance

customer satisfaction? Consider the role of delivery speed and reliability.

4. Identify potential challenges UPS might face in further scaling or enhancing its machine learning capabilities for route optimization. How could these challenges be addressed?
5. Speculate on future trends in logistics and delivery services related to machine learning and AI. What other innovations might companies like UPS explore to maintain their competitive edge?

Case References

McFarland, Matt. "The Case for Almost Never Turning Left While Driving." *The Washington Post*. Last modified April 9, 2014. **https://www.washingtonpost.com/news/innovations/wp/2014/04/09/the-case-for-almost-never-turning-left-while-driving/**.

UPS. "UPS Beats Competitors in On-Time Holiday Deliveries for Seven Years Straight." UPS. Last modified January 27, 2025. **https://about.ups.com/us/en/our-stories/customer-first/ups-beats-competitors-in-on-time-holiday-deliveries-for-seven-ye.html**

ENDNOTES

[1] Duhigg, Charles. “How Companies Learn Your Secrets.” *New York Times Magazine*, February 16, 2012. https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html.

[2] Hays, Constance L. “What Wal-Mart Knows About Customers’ Habits.” *The New York Times*, November 14, 2004. https://www.nytimes.com/2004/11/14/business/yourmoney/what-walmart-knows-about-customers-habits.html

[3] Mnih, V., Kavukcuoglu, K., Silver D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., et al. (2015). “Human-Level Control through Deep Reinforcement Learning.” *Nature* 518(7540), 529–533. **https://doi.org/10.1038/nature14236**

[4] Raman, Aditya. “Explaining Spotify Song Classification and Principal Components Analysis using Python.” *Medium*, August 26, 2022. https://medium.com/@adityaraman/explaining-spotify-song-classification-and-principal-components-analysis-using-python

[5] Torlay, Lucas. “Creating a Zillow Zestimate Using Machine Learning.” *Medium*, May 26, 2022. **https://medium.com/@ltorlay/creating-a-zillow-zestimate-using-machine-learning-a4785b9e541**

[6] Xue, D., Frisch, A., and He, D. (2019). Differential Diagnosis of Heart Disease in Emergency Departments Using Decision Tree and Medical Knowledge. In: Gadepally, V., et al. *Heterogeneous Data Management, Polystores, and Analytics for Healthcare*. DMAH Poly 2019. Lecture Notes in Computer Science, vol 11721. Springer, Cham. https://doi.org/10.1007/978-3-030-33752-0_16; Palaniappan, S., & Awang, R., “Intelligent Heart Disease Prediction System Using Data Mining Techniques.” *IEEE/ACS International Conference on Computer Systems and Applications AICCSA 2008*, pp. 108– 115, March 31, 2008–April 14, 2008.

[7] ActiveWizards Group LLC. “5 Real-World Examples of Logistic Regression Application.” *ActiveWizards Blog*. **https://activewizards.com/blog/5-real-world-examples-of-logistic-regression-application**; Lynn, Jacob. “Predicting Cancellations with Survival Modeling.” *Booking.com Data Science Blog*. November 11, 2024. **https://booking.ai/predicting-cancellations-with-survival-modeling-a299af54249b**

[8] Pinem, F. J., Andreswari, R., & Hasibuan, M. A.. “Sentiment Analysis to Measure Celebrity Endorsement’s Effect using Support Vector Machine Algorithm.” In *Proceedings of the 2018 International Conference on Electrical Engineering and Computer Science and Informatics (EECSI)*, pp. 690–695. Malang, Indonesia: IEEE, 2018. https://core.ac.uk/download/pdf/296976312.pdf

[9] Qi, Y. (2019). “Random Forest for Bioinformatics.” In *Machine Learning Techniques for Bioinformatics* (pp. 1–18). Springer. **https://www.cs.cmu.edu/~qyj/papersA08/11-rfbook.pdf**

[10] Mallouhy, R. E., Guyeux, C., Jaoude, C. A., & Makhoul, A. “K-Mean Clustering: A Case Study in Yvelines, Île-de-France.” In *2022 14th International Conference on Computational Intelligence*

and Communication Networks (CICN), pp. 1–6. Al-Khobar, Saudi Arabia: IEEE, 2022. **https://doi.org/10.1109/CICN56167.2022.10008344**.

[11] *AI Tech Daily.* "AI in Automotive Manufacturing: Driving the Future of Vehicle Production." August 5, 2024. https://medium.com/ai-tech-daily/ai-in-automotive-manufacturing-driving-the-future-of-vehicle-production

[12] Chopra, R., Clarke, K., Burrows, C. A., & Khan, L. M. "Joint Statement on Enforcement Efforts against Discrimination and Bias in Automated Systems." *Consumer Financial Protection Bureau*, May 2, 2025. https://www.consumerfinance.gov/about-us/newsroom/joint-statement-on-enforcement-efforts-against-discrimination-and-bias-in-automated-systems/

[13] Groves, L., Metcalf, J., Kennedy, A., Vecchione, B., & Strait, A. "Auditing Work: Exploring the New York City Algorithmic Bias Audit Regime." arXiv, February 2024. **https://arxiv.org/html/2402.08101v1**

[14] Krug, Steve. *Don't Make Me Think: A Common Sense Approach to Web Usability.* 2nd ed. Berkeley: New Riders, 2005.

CHAPTER 4S
Machine Learning Technical Supplement

This supplement provides a more detailed explanation of the machine learning algorithms introduced in **Chapter 4** and introduces metrics commonly used to assess their performance. It is intended for those who want a more detailed understanding of these tools.

It is beyond the scope of this book to cover each of these methods to the degree that you could perform them on data. However, we will focus on providing a managerial-level understanding of each method, how it works, its applications, and its strengths and weaknesses.

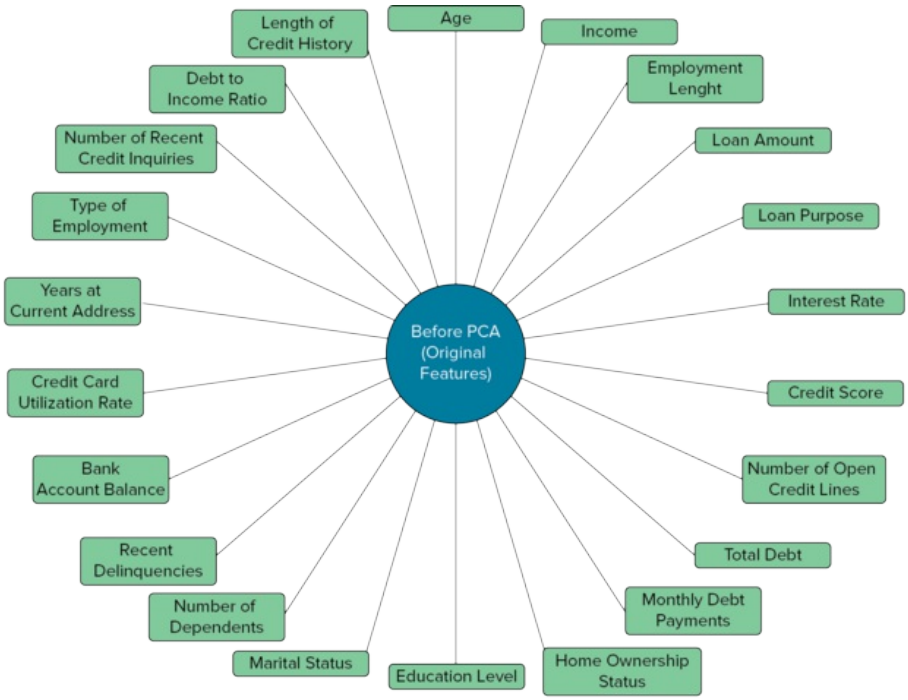
PRINCIPAL COMPONENT ANALYSIS

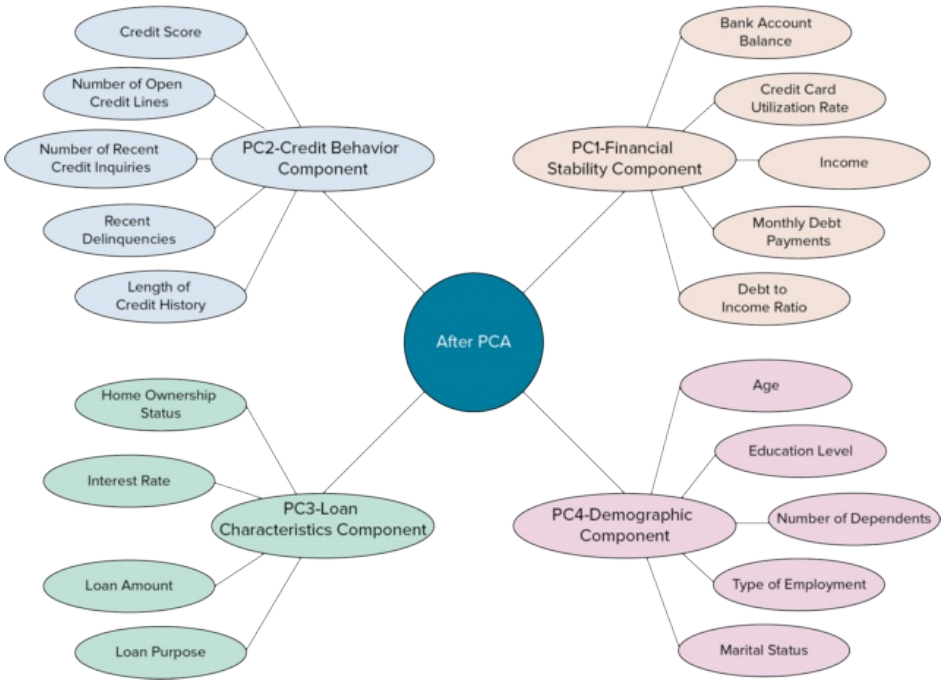
- **Machine learning type:** Unsupervised
- **Primary use:** Data Preprocessing

Imagine you have a massive dataset with many variables. While this data is rich in information, the number of variables can be overwhelming and challenging to analyze effectively. When a data set contains many variables, groups of variables will be related to each other. Principal component analysis (PCA) can reduce a large dataset's dimensionality (number of variables) while retaining as much of the original information as possible. For example, consider **Figure 4S.1**. As you can see, there are 22 original features before PCA. After PCA, there are four principal components (PC1—Financial Stability, PC2—Credit Behavior, PC3—Loan Characteristics, and PC4—Demographic) and three features (Employment Length, Total Debt, and Years at Current Address).

FIGURE 4S.1 Before and After PCA

PCA has drastically reduced the number of features in this example from 22 to 7. The color coding shows which variables were combined into each principal component.





How PCA Works

PCA involves the following:

1. ***Identifying correlated variables.*** PCA first identifies sets of correlated variables. For instance, in a dataset containing both "annual income" and "luxury car ownership," these two variables might be closely related. PCA will identify groups of these relationships to consolidate the information they contain.
2. ***Creating principal components.*** Instead of selecting one variable from each group to represent the group, PCA combines each group of correlated variables into a new principal component (PC) variable. Each new PC variable is a weighted combination of the variables that formed it. The resulting set of PC variables is uncorrelated.
3. ***Ranking principal components.*** The PC variables are then ranked so that the first PC contains the most variation in the dataset, the second includes the second most variation, and so on.

Business Applications of PCA

PCA has numerous business applications, including:

- ***Market research.*** It can identify patterns in customer preferences and behaviors by distilling large customer survey datasets into the most influential factors.
- ***Risk management.*** It can identify the key factors that drive portfolio risk and returns in finance.
- ***Feature reduction in machine learning.*** It can simplify datasets before applying machine learning algorithms. This can improve performance and accuracy.

Advantages of PCA

The following are some advantages of PCA:

- ***Fewer variables.*** The key benefit of PCA is that it reduces the number of variables while preserving much of the original information. Frequently, the first few principal components can explain most of the patterns in the data. With fewer variables, it is easier to visualize and interpret patterns in the data. Reducing the number of variables also speeds computations. In data-heavy industries like finance, health care, and marketing, this can give faster insights at a lower cost.
- ***Fewer redundant data points.*** PCA also filters out irrelevant or redundant data points leading to cleaner insights. It is like summarizing a long book into a few key themes without losing the main message.
- ***Revealing hidden data structures.*** PCA is also advantageous in exploratory data analysis to help reveal hidden structures that were not apparent initially.
- ***Segmentation.*** When working with high-dimensional datasets, PCA can map them into two or three dimensions, making complex relationships more straightforward. This is particularly useful in customer segmentation or medical diagnostics, where patterns might not appear at

first glance

Limitations of PCA

The following are some limitations of PCA:

- ***Loss of intuition.*** The new principal components don't always have obvious meanings. For instance, in a retail dataset, a principal component might combine "total purchases" and "customer age," which doesn't directly translate into a human-understandable concept.
- ***Assumes linearity.*** PCA works best when relationships are linear—if your data contains complex, nonlinear patterns (e.g., predicting customer churn based on behavior), other techniques might work better.

PCA is a powerful tool for reducing the number of variables to consider in a dataset, allowing for more straightforward analysis, visualization, and interpretation. Its ability to distill related variables into the most informative components makes it valuable in analytical applications.

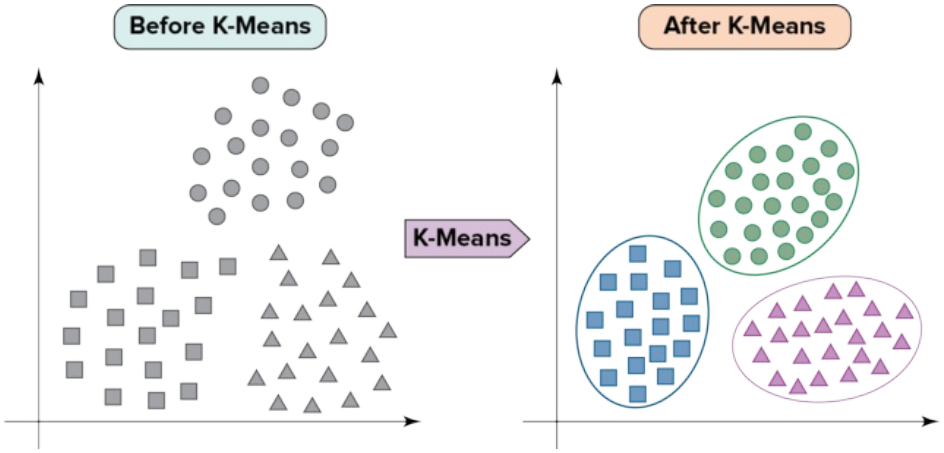
K-MEANS CLUSTERING

- **Machine learning type:** Unsupervised
- **Primary use:** Customer segmentation and personalization

K-means sorts observations into groups (or clusters) based on similarities (see **Figure 4S.2**). It is especially useful when there are no labels or predefined categories. For example, in a dataset containing various fruits, *k*-means can group the data based on features such as sweetness, shape, and color—organizing similar fruits into the same cluster, even if the specific fruit types are not known in advance. This makes *k*-means valuable for uncovering hidden patterns in unlabeled data.

FIGURE 4S.2 K-Means Clustering

Grouping by similarity: Points share the color of their closest center as k-means partitions the space into clusters.



How *k*-Means Clustering Works

K-means involves the following steps:

1. ***Choose a value for k.*** Selecting the number of clusters (k) to build can be based on domain knowledge (like knowing there are six types of fruits in your mix) or guided by techniques like the *elbow method*. In the elbow method, try increasing the value of k and evaluate the improvement gained with an additional cluster. Stop when adding more clusters does not significantly improve results.
2. ***Initialize.*** Select k starting points (called centroids) to cluster around. Each centroid is a point in the data space that contains a value for each variable being considered—similar to how a row of data includes values for all features. These are often chosen randomly, though other formal techniques for initialization exist.

3. ***Form groups.*** Each piece of data (each fruit, in our analogy) is assigned to the nearest centroid based on similarity across all variables. This means comparing the data point's values to each centroid's and placing it into the group with the closest match.
4. ***Refine.*** For each group, calculate a new centroid by taking the average of all the data points in that group. This means computing each variable's mean value across the cluster's points. The new centroid becomes the updated center of that group. If the centroids no longer change between iterations, the algorithm has converged—meaning the grouping is stable and the clustering process can stop.
5. ***Reform groups.*** Repeat steps 3 and 4, assigning points to groups and updating the centers until the centroids stop changing, meaning the groups are as accurate as possible.
6. ***Using the model.*** Assign new observations to the nearest cluster.

Business Applications of *k*-Means Clustering

K-means has numerous business applications, including:

- ***Market segmentation.*** In marketing, clustering helps segment customers based on buying patterns, preferences, or demographics, enabling more targeted marketing strategies.
- ***Inventory categorization.*** Retailers can use clustering to organize their inventory, grouping similar items to optimize their supply chain and sales strategies.
- ***Anomaly detection.*** Clustering can detect unusual patterns or outliers in data, useful in fraud detection or system health monitoring.
- ***Outlier identification.*** *K*-means clustering can also be used in identifying outliers in the dataset.

Advantages of *k*-Means Clustering

The following are some advantages of *k*-means clustering:

- ***Discovery of natural groupings.*** Clustering is excellent for exploratory data analysis, allowing businesses to discover natural groupings or patterns in the data.
- ***Simplicity and flexibility.*** Algorithms like *k*-means are simple to understand and implement. They can be applied to a wide range of problems.

Limitations of *k*-Means Clustering

The following are some limitations of *k*-means clustering:

- ***Sensitivity to initialization and noise.*** *k*-means' results can be sensitive to the initial selection of centroids and can be influenced by outliers or noise in the data.
- ***Assumption of spherical clusters.*** *K*-means assumes that clusters are spherical and of similar size,

which might not always be the case in real-world data.

Clustering algorithms like *k*-means play a crucial role in data analysis, particularly in scenarios where the data lacks labels or predefined categories. They are key in uncovering hidden patterns and insights, which can be pivotal for strategic decision making in various business contexts.

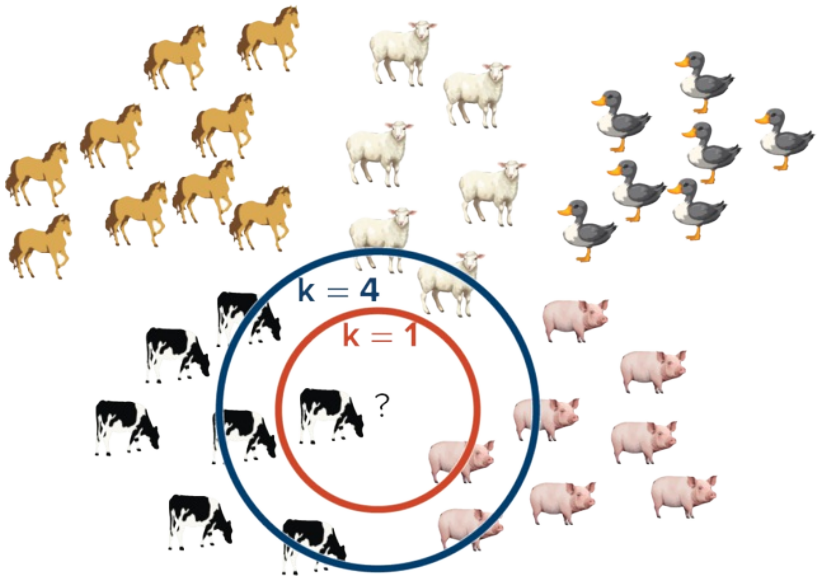
K-NEAREST NEIGHBORS

- **Machine learning type:** Supervised
- **Primary use:** Classification and risk assessment

K-nearest neighbors (*k*-NN) is a simple and intuitive machine learning method that makes predictions by looking at the "neighbors" closest to a new data point (see **Figure 4S.3**). It works on the idea that similar things exist near each other, so if you want to predict something about a new customer, for example, *k*-NN finds the most similar past customers and bases its decision on them. It is commonly used for both classification and prediction of values.

FIGURE 4S.3 *K*-Nearest Neighbor

Classification by proximity: The query point (?) inherits the label shared by the majority of the nearest neighbors, showing the algorithm in action.



How *k*-NN Works

K-NN involves the following steps:

1. ***Measure distance.*** When a new data point needs to be classified, *k*-NN looks at the "*k*" closest data points in the training set, "closest" meaning most similar based on distance using variable values (usually Euclidean distance).
2. ***Vote on outcome.*** The algorithm checks the labels of the *k* neighbors and assigns the most common label to the new point.

K-NN doesn't build a model in advance; it stores the training data and makes decisions only when a prediction is needed. This is called a "lazy learning" approach.

Choosing the right k is essential. A small k may be too sensitive to noise, while a large k may blur class boundaries. Typically, values like 3, 5, or 7 are tested. Odd values for k are used so there are no ties in the voting. Cross-validation or visual inspection can help determine the best value for a given problem.

Business Applications of *k*-NN

K-NN has numerous business applications, including:

- ***Customer classification.*** Businesses can use *k*-NN to assign new customers to existing segments based on the behaviors and characteristics of similar customers.
- ***Recommendation systems.*** *k*-NN can suggest products or content by analyzing users' preferences with similar tastes.
- ***Anomaly detection.*** *k*-NN can help detect unusual behavior, such as fraud, by comparing new activity to typical patterns found in past data.

Advantages of *k*-NN

The following are some advantages of *k*-NN:

- ***Simple and intuitive.*** It is easy to understand and implement.
- ***No training time.*** The algorithm requires no training, which reduces costs and allows quick deployment. This is also valuable with frequently changing datasets.
- ***Flexible.*** It can be used for classification or assigning a numerical value by averaging the values of the nearest neighbors.

Limitations of *k*-NN

The following are some limitations of *k*-NN:

- ***Slow prediction.*** It slows with large datasets because each new input is compared to all stored data points.
- ***Sensitive to irrelevant features.*** Features should be carefully selected and scaled for high-quality predictions.
- ***Doesn't scale well.*** Predictions take longer to complete as the dataset size grows.

K-NN is a powerful tool where speed and simplicity are needed, especially for businesses with small- to medium-sized datasets. Its ability to adapt and make decisions based on similarity makes

it effective in customer-focused applications, provided the data is clean and well-prepared.

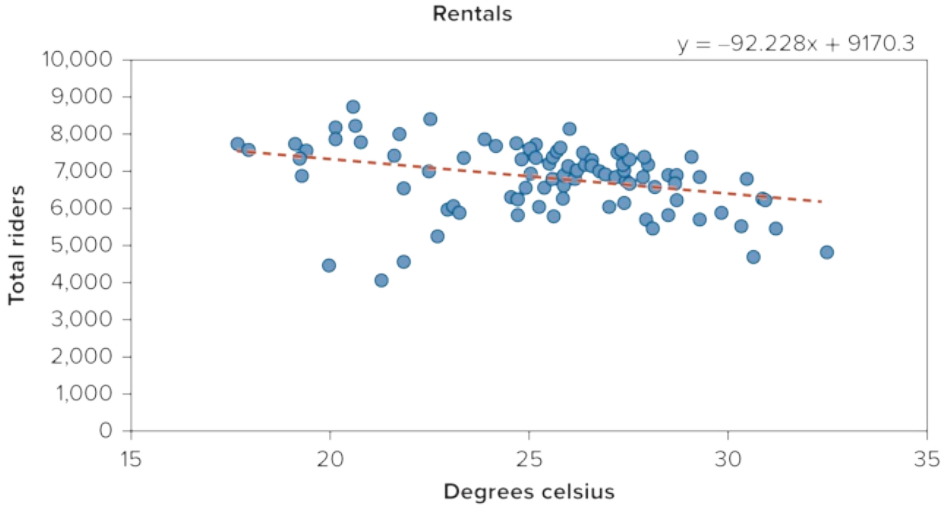
LINEAR REGRESSION

- **Machine learning type:** Supervised
- **Primary use:** Prediction

Linear regression has a long history of being a valuable tool for exploring the relationships between a set of variables *(independent variables)* with a numerical variable of interest called a *dependent variable*. Linear regression looks at collected data to identify linear patterns between each independent variable and the dependent variable, allowing for predictions based on those historical relationships. **Figure 4S.4** shows how the summer heat meant fewer bicycle rentals for DC BikeShare.

FIGURE 4S.4 Linear Regression of Bicycle Rentals

This graph shows that as the temperature increased during the summer, the number of bicycle rentals decreased. An increase of 1 degree Celsius meant a decrease of 92.2 rentals (1 degree F = −51.2 rentals), all other things being equal.



JMP software

For example, a zoo might want to predict the number of visitors it will receive on a given day. Using linear regression, we could build a model with factors such as temperature, day of the week, and advertising spending as independent variables plus daily visitor count as the dependent variable. The model might reveal that warmer temperatures increase attendance, weekends attract more visitors, and higher ad spending boosts turnout. Each of those relationships would be quantified by a coefficient indicating its impact. By understanding these relationships, the zoo can better allocate advertising dollars, prepare for high-traffic days, and make data-driven operational decisions.

How Linear Regression Works

Linear regression involves the following:

1. ***Finding the best fit line.*** The simplest form (simple linear regression) involves fitting a straight line to the pairs of data points between a single independent variable and the dependent variable. This line represents the best estimate of the relationship between an independent variable (like time or investment) and a dependent variable (like sales or returns). The algorithm calculates the line that minimizes the sum of the squared differences (or the errors) between the actual data points and the predicted values by the line. This process is known as *least squares regression*.
2. ***Interpreting the model.*** The output of a linear regression model is an equation that describes the relationship between the dependent variable and independent variables. For example, suppose we want to predict monthly electricity usage (in kilowatt-hours) for a household based on the number of occupants. The resulting equation might look like:

   $$Electricity\,Usage = 200 + 150 \times Occupants$$

   Here, 200 is the intercept, representing the baseline electricity usage for an unoccupied house (such as appliances that remain plugged in). The 150 represents the slope, meaning that for every additional occupant, the electricity usage is expected to increase by 150 kilowatt-hours per month. This equation allows us to estimate how much energy a household will use based on the number of people living there.

This method extends to multiple linear regression when there are multiple independent variables. This allows for more complex models that can capture the impact of several variables on the dependent variable of interest.

Business Applications of Linear Regression

Linear regression has numerous business applications, including:

- ***Sales forecasting.*** Businesses use linear regression to forecast future sales based on historical sales data and relevant variables, such as advertising spend or market trends. Accurate sales forecasts are used in budgeting, staffing, and inventory management.
- ***Market analysis.*** Car dealerships and automotive pricing platforms (Kelley Blue Book, Edmunds, and Carvana) use linear regression to estimate vehicle resale values based on factors like make, model, mileage, year, and condition. Auto sales are a matter of public record and are used in the dataset to predict used car value, helping buyers and sellers make informed pricing decisions.

Advantages of Linear Regression

The following are some advantages of *k*-NN:

- ***Simplicity and interpretability.*** Linear regression models are straightforward to understand and interpret, making them a popular choice for many business applications.
- ***Predictive power.*** When the relationship between variables is linear, these models can be quite powerful in prediction.

Limitation of Linear Regression

There is one main limitation to linear regression:

- ***Assumes a linear relationship.*** Linear regression assumes a linear relationship between the variables. While linear relationships are common, nonlinear relationships do occur.

Linear regression is a powerful, easy-to-understand tool for modeling and predicting quantitative data. Its ability to provide clear insights into the relationships between variables makes it a staple in business analytics' toolbox.

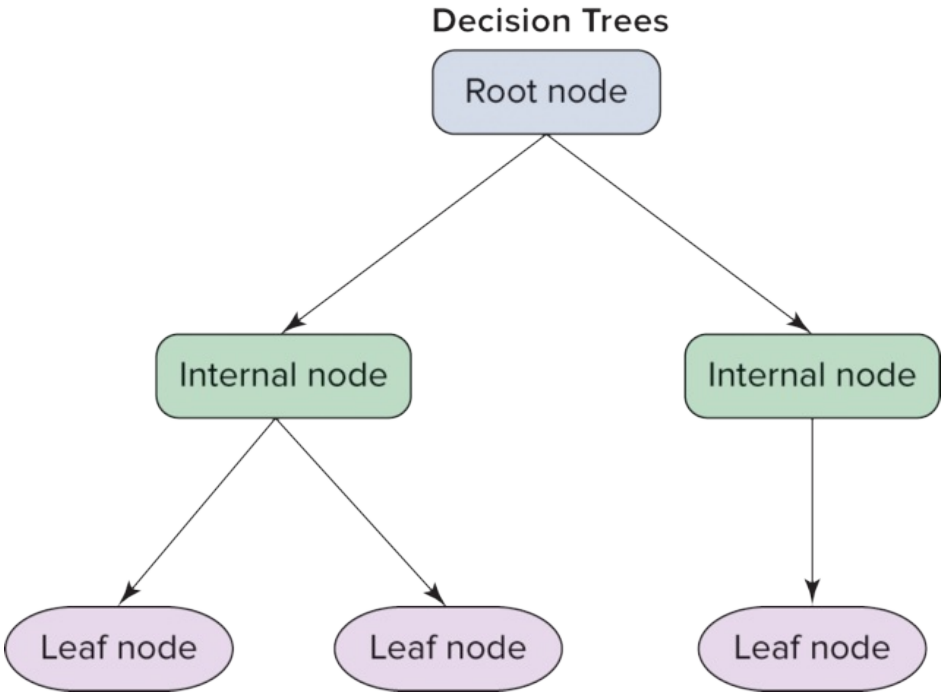
DECISION TREES

- **Machine learning type:** Supervised
- **Primary use:** Classification

Decision trees are supervised learning algorithms primarily used for classification, but they can also predict quantitative values. Decision trees are called trees because they branch out in a structure resembling a flowchart, making step-by-step branch decisions based on specific criteria. This branching approach allows decision trees to classify data into multiple categories or predict continuous values in regression tasks. A decision tree splits the data into branches, leading to different outcomes. At each node (or branching point), the tree asks a question about one of the data features, and based on the answer, the data is split accordingly. The decision tree structure is show in **Figure 4S.5**.

FIGURE 4S.5 Structure of a Decision Tree

A decision tree uses a series of questions (nodes) to split data into smaller groups based on different conditions. It starts at the root node, branches through internal nodes, and ends at leaf nodes, which represent the final decision or outcome for each path.



How Decision Trees Work

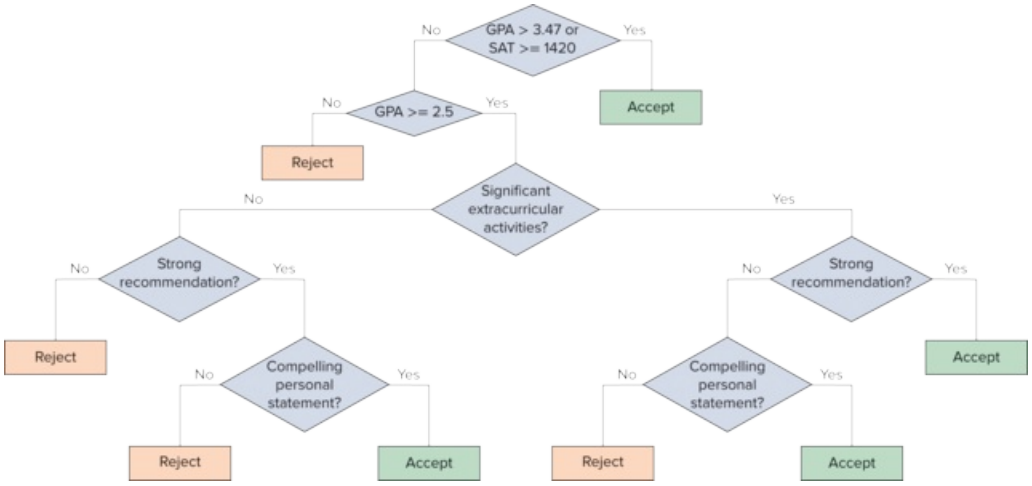
Using a decision tree involves the following steps:

1. ***Start with the root node.*** A decision tree starts at a single node, called the *root*.
2. ***Add branches to the node.*** The algorithm selects the feature that best splits the data. Branches are added to the node based on a decision rule based on the chosen feature.
3. ***Repeat the splitting process.*** Nodes are added to the end of each branch, and features are selected for each node that splits the data best at each step. Each split separates the data into more explicit categories, making predictions easier. This process continues until the tree reaches its final points, called *leaves*, where a classification or decision is made.
4. ***Prune the tree.*** Branches that add complexity but little value are removed, helping the model generalize better.

Figure 4S.6 presents a sample of a decision tree.

FIGURE 4S.6 Sample Decision Tree

This decision tree illustrates the process a university might use to determine whether to admit an applicant based on their academic scores, extracurricular activities, recommendation letters, and personal statement. The tree was formed using a training data set targeting students who successfully graduated from college in five years or less, with cutoff points for each decision determined by the algorithm to access the desired class size with a high likelihood of graduating.



Business Applications of Decision Trees

Decision trees have numerous business applications, including:

- ***Credit risk assessment.*** Financial institutions use decision trees to assess the risk of lending based on factors like credit history, income, and employment status.
- ***Customer segmentation.*** Marketing teams apply decision trees to segment customers based on behavior, preferences, or demographics for targeted marketing efforts.

- ***Medical diagnosis.*** In health care, decision trees can help diagnose conditions like breast cancer and COVID mortality by analyzing patients' symptoms and test results.

Advantages of Decision Trees

The following are some advantages of decision trees:

- ***High explainability.*** A decision tree's logic is easy to follow and interpret, making it a suitable tool for business decision making.
- ***Versatility.*** Decision trees can be used with both numerical and categorical data and applied to classification and numerical value estimation tasks.
- ***Handling nonlinear relationships.*** Decision trees can model complex, nonlinear relationships.

Limitations of Decision Trees

The following are some limitations of decision trees:

- ***Risk of overfitting.*** Without proper pruning, decision trees can create overly complex trees that do not generalize from the training data to unseen data.
- ***Instability.*** Small changes in the data can lead to different splitting, making some trees less reliable.

Decision trees are practical and versatile tools, especially for classification tasks. Their explainability makes them appealing for business applications, where that is as important as making accurate predictions.

RANDOM FORESTS

- **Machine learning type:** Supervised
- **Primary use:** Classification

Random forests represent an advancement over single decision trees, offering a more robust and accurate model by combining the predictions from multiple decision trees. This is called *ensemble learning*, where various models are used to solve the same problem, and the results are combined into a single prediction.

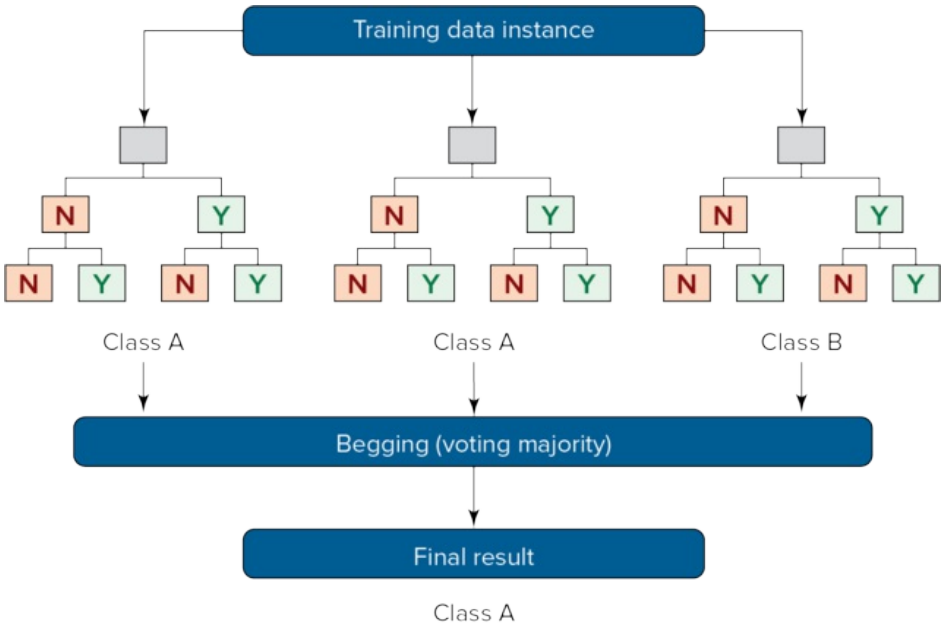
How Random Forests Work

Using a random forest involves the following (see **Figure 4S.7**):

1. ***Building multiple decision trees.*** A random forest algorithm creates a "forest" of decision trees. Each tree is built from a random sample of the data. This sampling method is known as *bootstrap aggregating* or *bagging*. Bagging creates some diversity among the trees and reduces the problem of overfitting.
2. ***Selecting random features.*** Random forests randomly select a subset of features for each tree's decision splits. This adds to the tree diversity in the model and helps reduce overfitting.
3. ***Combining predictions.*** Once all the trees in the forest are built, the random forest algorithm makes predictions by aggregating the predictions of the individual trees. Classification tasks are typically done by majority voting, and numerical prediction tasks usually involve averaging the projections.
4. ***Performing error estimation.*** An interesting aspect of random forest is its method of internal error estimation. It uses *out-of-bag (OOB) samples* (data not included in the training sample for a tree) to test each tree's performance on data it has not seen before to identify cases of overfitting.

FIGURE 4S.7 Random Forest Combining Decision Tree Results for Prediction

This figure illustrates how a random forest algorithm combines multiple decision trees to improve prediction accuracy. Each tree in the forest is trained on a different subset of the data, and their predictions are aggregated to produce a final, more reliable outcome.



Business Applications of Random Forests

Random forests have numerous business applications, including:

- ***Fraud detection.*** Random forests can effectively identify patterns indicative of fraudulent activity in financial transactions.
- ***Customer behavior prediction.*** In retail, they predict customer behavior, such as the likelihood of a customer purchasing a product or responding to a marketing campaign.
- ***Supply chain optimization.*** In manufacturing and logistics, random forests help predict demand, manage inventory levels, and mitigate supply chain risks.

Advantages of Random Forests

The following are some advantages of random forests:

- ***High accuracy.*** Random forests are generally more accurate than single decision trees due to their ensemble approach.
- ***Prevent overfitting.*** Using multiple trees formed from different subsets and implementing different variables reduces the risk of overfitting and makes the model more robust.
- ***Versatility.*** Random forests are effective for classification and quantitative estimation tasks and work well with categorical and numerical data.

Limitations of Random Forests

The following are some limitations of random forests:

- ***Interpretability.*** While individual trees are easy to interpret, a random forest is more complex and less intuitive.
- ***Computational intensity.*** Building and using many trees can require significant computational resources, especially with very large datasets.

Random forests provide a powerful tool for predictive analytics in various business applications. Their ability to combine the simplicity of decision trees with the power of an ensemble approach results in a versatile and robust model suitable for a wide range of tasks.

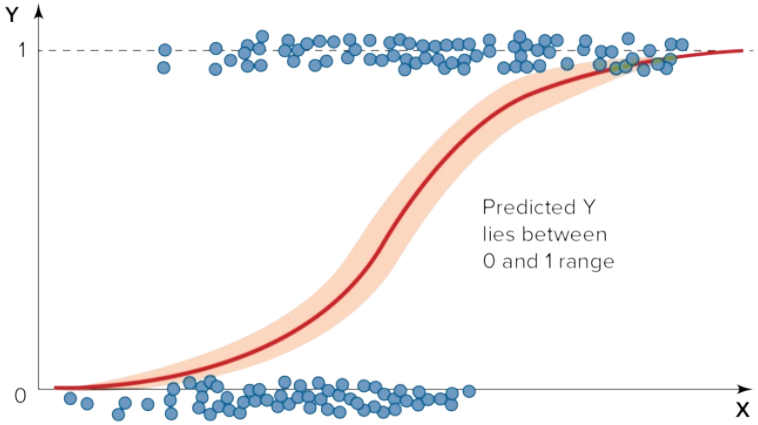
LOGISTIC REGRESSION

- **Machine learning type:** Supervised
- **Primary use:** Classification

Logistic regression is like linear regression, but it is used when the outcome of the variable of interest is either yes or no (binary) instead of a numerical value. If we used linear regression for this task, it could generate probabilities higher than 1 and lower than 0. Logistic regression estimates the probability of the target outcome occurring using historical data. For example, is an email spam or not, or will a loan be paid back or not? Logistic regression models these probabilities using an *S-shaped curve* (called a sigmoid function), which flattens out near 0 and 1 (see **Figure 4S.8**). This curve ensures that predictions always fall between 0 and 1, making it well-suited for representing the likelihood of binary outcomes.

FIGURE 4S.8 The S-Shaped Curve in Logistic Regression

Logistic regression uses a sigmoid (S-shaped) curve to map input values to probabilities between 0 and 1. This curve helps estimate the likelihood of a binary outcome—such as whether a customer will churn or an email is spam. A threshold, typically set at 0.5, is used to classify predictions as "yes" or "no" based on the estimated probability.



How Logistic Regression Works

Logistic regression involves the following:

1. ***Choosing a target outcome.*** First, choose a target outcome (Yes); all other possibilities are grouped as non-target (No). For example, a fitness club offers Basic, Standard, Premium, and VIP memberships, but it wants to predict which members will most likely upgrade to VIP. VIP membership becomes the target outcome in this case, and all other membership levels are classified as non-target.

2. ***Estimating probabilities.*** Logistic regression estimates the probability of the target outcome using historical data. For example, customer demographics, along with historical data on visits, spending, referrals, engagement with promotions, and purchase behavior, could be used to predict the probability of a member upgrading to VIP. The model might calculate a member has a 75 percent chance of upgrading to VIP. This probability helps the gym make a targeted promotion to encourage an upgrade.
3. ***The logistic function.*** Logistic regression transforms input from one or more independent variables (such as membership type, number of gym visits, or spending on extra services) into a probability score using the *logistic* (or *logit*) *function*. This function ensures outputs always have a value between 0 and 1. Because of this, logistic regression is widely used for binary classification problems, where the goal is to predict one of two possible outcomes.

Business Applications of Logistic Regression

Logistic regression has numerous business applications, including:

- ***Credit scoring.*** Banks use logistic regression to predict the likelihood of a customer defaulting on a loan based on their credit history, income, debt, and other factors.
- ***Marketing and sales.*** In marketing, logistic regression is used to predict the likelihood of a customer purchasing a product, which can help in targeted marketing strategies.
- ***Customer retention.*** Companies use logistic regression to predict the probability of customers switching service providers, enabling them to implement retention strategies.

Advantages of Logistic Regression

The following are some advantages of logistic regression:

- ***Simplicity and interpretability.*** Logistic regression is easy to implement and understand. Its mathematical foundation is understandable even for teams with limited machine learning expertise.
- ***Computational efficiency.*** Logistic regression can process large datasets quickly while using minimal memory and processing power. This makes it an excellent choice for organizations starting with machine learning because it allows quick insights without requiring extensive computing resources.
- ***Versatility in classification.*** Logistic regression is effective for binary classification problems (e.g., predicting whether a transaction is fraudulent) and can also be extended to multi-class classification (e.g., determining whether a customer subscribes to Basic, Standard, or VIP membership).
- ***Transparency and explainability.*** Unlike black-box models like neural networks, logistic regression models have high explainability. The model gives the strength and direction of

relationships between input variables and the target outcome, making troubleshooting errors and validating results easier.

Limitations of Logistic Regression

The following are some limitations of logistic regression:

- ***Limited to linear decision boundaries.*** Logistic regression assumes a linear relationship between input variables and the target outcome. Other machine learning models may perform better if the relationship is highly nonlinear.
- ***Sensitive to outliers and irrelevant features.*** Since logistic regression relies on numerical weights assigned to each predictor, extreme outliers or irrelevant variables can skew results. Proper feature selection and data preprocessing (e.g., normalization) are recommended to increase accuracy.
- ***May struggle with highly complex data.*** Logistic regression is a good starting method for classification tasks. However, other machine learning algorithms can better capture complex relationships in large, high-dimensional datasets.

Logistic regression is a robust and straightforward tool for binary classification problems, making it a popular choice in various business applications.

SUPPORT VECTOR MACHINES

- **Machine learning type:** Supervised
- **Primary use:** Classification

Support vector machines (SVMs) are supervised learning methods for classification, regression, and outlier detection. Renowned for their accuracy in high-dimensional spaces, SVMs are particularly useful when dealing with complex datasets that are not easily separable.

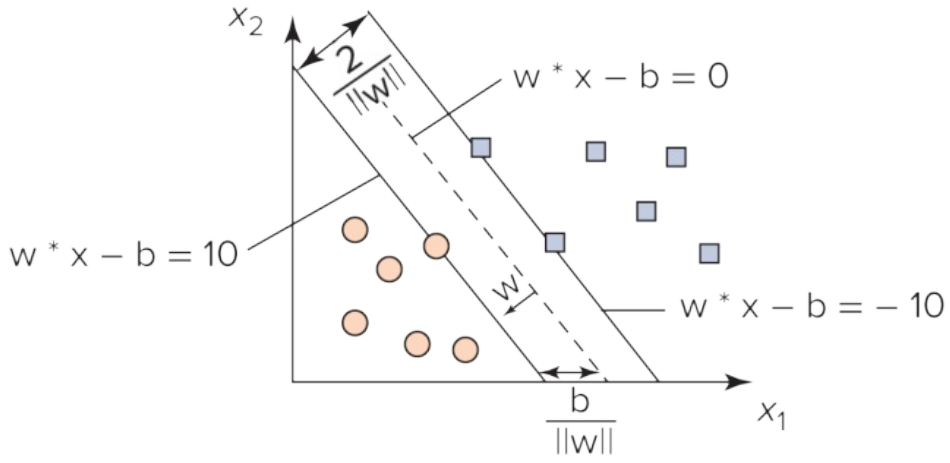
How SVM Works

Using SVM involves the following:

1. ***Separating data.*** SVM draws a line (or surface) that divides data into distinct categories.
2. ***Maximizing the margin.*** It chooses the boundary that leaves the widest gap between the two groups, helping the model make better predictions on new data (see **Figure 4S.9**).
3. ***Support vectors.*** Only a few key data points near the boundary, called *support vectors*, determine where the line is placed. These points are critical in defining the model.

FIGURE 4S.9 Support Vector Machine (SVM) Classification

SVM finds the best boundary (called a hyperplane) that separates two groups of data. It chooses the line that maximizes the distance, or margin, between itself and the nearest points from each group—helping ensure accurate classification of new data.



Business Applications of SVMs

SVMs have numerous business applications, including:

- ***Customer classification.*** Based on historical behavior, companies use SVMs to identify whether a new customer will likely convert or seek another service provider.
- ***Text categorization.*** SVMs are widely used to classify documents, emails, or reviews as spam or sentiment.
- ***Fraud detection.*** In finance, SVMs help flag suspicious transactions based on patterns of legitimate and illegitimate activity.

Advantages of SVMs

The following are some advantages of SVMs:

- ***Effective in high-dimensional spaces.*** SVMs perform well even when the dataset has many features or variables.
- ***Work well with clear class separation.*** SVMs excel when data has a well-defined boundary between categories.
- ***Robust against overfitting.*** Especially when used with regularization, SVMs are less prone to overfitting than some other methods.

Limitations of SVMs

The following are some limitations of SVMs:

- ***Not ideal for large datasets,*** SVMs slow down significantly as the dataset gets larger.
- ***Less effective when classes overlap.*** When the boundary between categories is fuzzy, simpler models may perform just as well with less complexity.
- ***Limited interpretability.*** The model is more difficult to explain than decision trees or regression.

SVMs are a powerful tool for classification problems, especially when the goal is to make clear-cut decisions from complex data. They can be very effective when precision matters and the dataset is well structured.

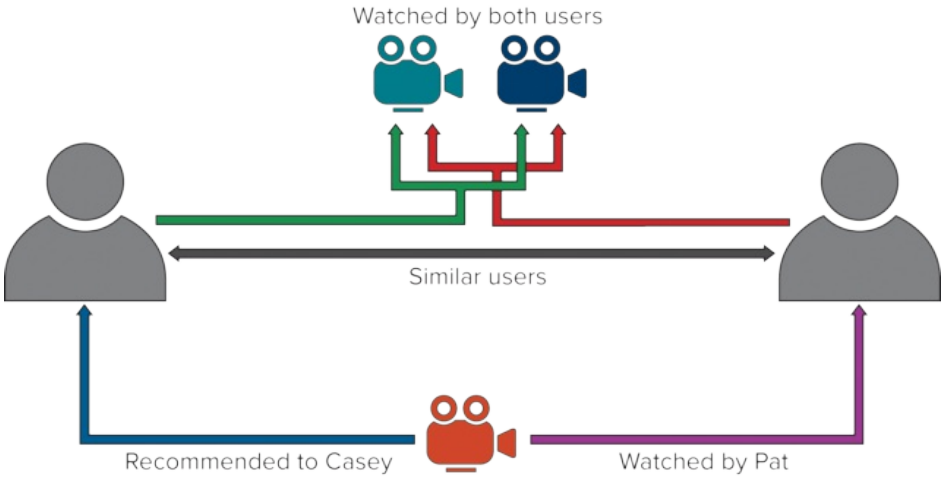
COLLABORATIVE FILTERING

- **Machine learning type:** Unsupervised
- **Primary use:** Recommendation systems

Collaborative filtering predicts a user's preferences based on similar users' preferences. It operates on the principle that users who have agreed in the past will likely agree again in the future (see Figure 4S.10).

FIGURE 4S.10 Collaborative Filtering for Movie Recommendations

Collaborative filtering recommends movies by finding patterns in what similar users have liked. Instead of analyzing the movie itself, the system compares your viewing history to others' and suggests titles enjoyed by users with similar tastes.



How Collaborative Filtering Works

There are two main approaches:

1. ***User-based collaborative filtering.*** This method identifies users similar to the target user based on past behavior (e.g., product ratings, purchases, or clicks). The system then recommends items that similar users liked, but the target user hasn't interacted with yet.
2. ***Item-based collaborative filtering.*** This approach does not find similar users but identifies similar items based on user behavior. For example, if most people who watch *Show A* also watch *Show B*, then *Show B* might be recommended to someone who liked *Show A*.

Both methods typically use similarity metrics to compare users or items. The most common is the correlation coefficient,

Collaborative filtering uses a user-item interaction matrix as its foundation, where each cell

represents an action such as a rating or purchase. Patterns within this matrix are analyzed to make predictions.

Business Applications of Collaborative Filtering

Collaborative filtering has numerous business applications, including:

- ***Product recommendations.*** E-commerce platforms like Amazon use collaborative filtering to suggest products based on customer behavior.
- ***Streaming services.*** Platforms like Netflix or Spotify recommend shows or songs by comparing users with similar listening or viewing patterns.
- ***Online retail.*** Apparel or grocery services can suggest items frequently bought together or preferred by users with similar profiles.

Advantages of Collaborative Filtering

The following are some advantages of collaborative filtering:

- ***Highly personalized.*** Recommendations are based on actual user behavior.
- ***Detailed product metadata is not required.*** The system only needs user-item interactions for calculations.

Limitations of Collaborative Filtering

- ***Cold start problem.*** It struggles to make recommendations for new users or new items.
- ***Scalability challenges.*** As the number of users and items grows, it processes slower.
- ***Data sparsity.*** Many users only interact with a small percentage of available items, limiting the system's effectiveness.

Collaborative filtering provides an effective way to analyze user behavior and make recommendations the user will like. It increases customer engagement and sales making it a popular marketing tool.

TIME SERIES ANALYSIS

- **Machine learning type:** Statistical/supervised (depending on the model)
- **Primary use:** Forecasting

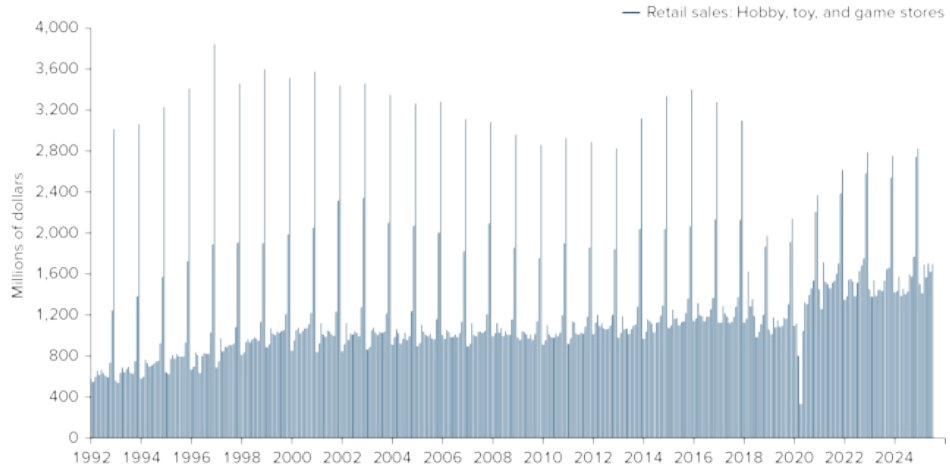
Time series analysis involves analyzing data points collected or recorded at regular time intervals, such as daily sales, monthly revenue, or hourly web traffic. The goal is to identify trends, cycles, and seasonal patterns to make accurate forecasts (see **Figure 4S.11**). Common techniques include:

- ***Trend analysis.*** Observing whether the data shows a long-term increase or decrease.
- ***Seasonality detection.*** Identifying repeating weekly, monthly, or annual patterns (e.g., increased toy sales in December).
- ***Autoregressive models (e.g., ARIMA).*** Using past values of the variable itself to predict future outcomes.

FIGURE 4S.11 Retail Sales—Hobby, Toy, and Game Stores (1992–2025)

This graph shows monthly retail sales in hobby, toy, and game stores with a strong seasonal pattern—peaking each year during the holiday season—and a steady upward trend over time. The sharp dip in 2020 highlights the impact of the COVID-19 pandemic on consumer spending.

Source: U.S. Census Bureau, Retail Sales: Hobby, Toy, and Game Stores [MRTSSM45112USN], retrieved from Federal Reserve Economic Data (FRED), Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/MRTSSM45112USN, May 4, 2025.



How Time Series Analysis Works

Time series analysis is used to model and forecast data that is collected over time. It identifies patterns such as trends and seasonality and uses these patterns to predict future values. The primary assumptions are that patterns and trends from the past will continue to the future. The

process typically involves the following steps:

1. ***Collect and visualize the data.*** Gather data collected at regular intervals (e.g., monthly sales, daily website visits) and plot it to observe any patterns—such as upward trends, repeating seasonal cycles, or sudden shifts.
2. ***Decompose the series.*** Separate the time series into its components:
   - ***Trend.*** Long-term upward or downward movement in the data.
   - ***Seasonality.*** Regular patterns that repeat over fixed periods, such as higher toy sales in December.
   - ***Residual/noise.*** Irregular fluctuations not explained by trend or seasonality.
3. ***Choose a forecasting model.*** Select a suitable model based on the patterns observed. Common time series models include:
   - ***Moving averages or exponential smoothing.*** Use weighted past values to smooth out noise and predict short-term changes.
   - ***ARIMA (autoregressive integrated moving average).*** Captures trends and patterns by combining autoregression, differencing, and moving averages.
   - ***Seasonal ARIMA (SARIMA):*** Extends ARIMA by modeling seasonal cycles explicitly.
4. ***Fit the model.*** Train the model using historical data. The model learns the underlying structure of the time series, including how previous values influence future outcomes.
5. ***Forecast future values.*** Use the trained model to generate forecasts. These predictions are typically visualized alongside historical data to assess how well the model captures trends and cycles.
6. ***Evaluate accuracy.*** Compare forecasted values to actual outcomes using metrics like mean absolute error (MAE) or root mean squared error (RMSE). Adjust the model if necessary to improve accuracy.

Business Applications of Time Series Analysis

Time series analysis has numerous business applications, including:

- ***Sales and revenue forecasting.*** Companies can project future revenue to guide budgeting and production planning.
- ***Inventory management.*** Time series forecasts help ensure stock levels match expected demand.
- ***Website and app traffic.*** Marketers monitor traffic over time to optimize campaigns and manage server capacity.

Advantages of Time Series Analysis

The following are some advantages of time series analysis:

- ***Captures time-dependent patterns.*** It excels at identifying seasonality and trends in business data.
- ***Improves planning accuracy.*** It supports more reliable forecasting, which helps reduce waste and improve efficiency.

Limitations of Time Series Analysis

The following are some limitations of time series analysis:

- ***Assumes historical patterns will continue.*** Sudden changes (e.g., market disruptions) may not be captured.
- ***Sensitive to missing data and outliers.*** Gaps or irregularities in time-based data can distort results.
- ***Minimum data requirements.*** Three years of data is recommended to reliably identify and predict seasonal patterns.

Time series analysis is essential for businesses that rely on forecasting. Whether anticipating product demand or budgeting for future quarters, understanding trends over time leads to more confident, data-driven decisions.

METRICS FOR THE QUALITY OF ALGORITHMS

Several algorithms can analyze the same problem, but how do we determine which performs best? This section covers key metrics for evaluating machine learning models, including when to use them, how they are calculated, and their strengths and weaknesses. Since many classification metrics rely on values from the confusion matrix, we start there before discussing accuracy, precision, recall, and other key metrics.

Confusion Matrix

A *confusion matrix* is a table used to evaluate the performance of a classification algorithm by comparing its predictions to actual outcomes (see Table 4S.1). It categorizes results into four groups:

- **True positives (TP)**, where the model correctly predicts a positive outcome.
- **True negatives (TN)**, where it correctly predicts an adverse outcome.
- **False positives (FP)**, where it incorrectly predicts a positive outcome.
- **False negatives (FN)**, where it misses a positive case.

Table 4S.1 **Confusion Matrix Structure**

Actual / Predicted	Predicted Positive	Predicted Negative
Actual positive	✅ **True positive (TP)** (Correctly identified positive cases)	✖ **False negative (FN)** (Missed positive cases)
Actual negative	✅ **False positive (FP)** (Incorrectly predicted as positive)	✖ **True negative (TN)** (Correctly identified negative cases)

This breakdown helps identify how often a model is correct and the types of errors it makes. For example, in fraud detection, a false positive might flag a legitimate transaction as fraudulent, while a false negative could let actual fraud go undetected. Key metrics, such as accuracy, precision, and recall, are calculated using values from the confusion matrix.

The confusion matrix provides a detailed breakdown of a model's performance. We can use these outcomes to calculate various measures of a model's quality.

Confusion Matrix Example: Spam Classification

Imagine we are evaluating a spam detection model that classifies emails as spam or not spam. Here is how its predictions compare to actual outcomes:

Actual / Predicted	Predicted Spam	Predicted Not Spam	Total
Actual spam	50 (TP) ✅	10 (FN) ✖	60
Actual notsSpam	5 (FP) ✖	35 (TN) ✅	40
Total	55	45	100

Key values from this confusion matrix:

- **True positives (TP) = 50** (Spam emails correctly identified as spam)
- **False positives (FP) = 5** (Non-spam emails incorrectly marked as spam)
- **False negatives (FN) = 10** (Spam emails incorrectly marked as not spam)
- **True negatives (TN) = 35** (Non-spam emails correctly classified)
- **Total cases = 100** (Total number of emails evaluated)

Now that we have established the confusion matrix and its key components, we can use it to calculate essential metrics that measure a model's performance. To illustrate these calculations, we will use this spam email classification example. Each metric will be applied to this confusion matrix to demonstrate how it is computed and interpreted.

Metrics

No single metric tells the whole story. The best metric to use depends on the business problem. For example, fraud detection systems might prioritize recall (catching as much fraud as possible), while spam filters emphasize precision (avoiding spam flags on desired emails). Understanding when to prioritize each metric is as important as knowing how they are calculated.

1. Accuracy

The most commonly used metric is accuracy, which tells us the proportion of correctly predicted cases out of all cases. While accuracy provides a simple way to evaluate a model, it does not always tell the whole story—especially when dealing with imbalanced datasets.

- **Use.** General performance evaluation of classification models.
- **Formula.** $\text{Accuracy} = \dfrac{TP + TN}{TP + TN + FP + FN}$
- **Applicable algorithms.** Logistic regression, decision trees, random forests, SVM, neural networks.

- **Business application.** Provides a quick overall assessment of model performance in tasks like customer segmentation, fraud detection, and product categorization.
- **Scale:** 0 to 1 (or 0 to 100 percent).
- **Desirable score.** Closer to 1 is better, but accuracy can be misleading in imbalanced datasets—for example, a model predicting fraud in only 1 percent of transactions might be 99 percent accurate by always predicting "no fraud," yet still be completely useless.
- **Example calculation using the spam classification confusion matrix.** From our spam email classification example, we use the following confusion matrix values:

  True positives (TP) = 50

  False positives (FP) = 5

  False negatives (FN) = 10

  True negatives (TN) = 35

  Total cases = 100

  Applying the formula:

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \\ &= \frac{50 + 35}{50 + 35 + 5 + 10} \\ &= \frac{85}{100} = 0.85\,(85\%) \end{aligned}$$

- **Interpretation.** The model correctly classifies 85 percent of emails. While this may seem high, we must consider whether false positives (flagging legitimate emails as spam) or false negatives (missing actual spam emails) pose a bigger problem. Accuracy alone does not tell us which type of error is more frequent or important, so additional metrics like precision and recall are needed.

2. Precision (also known as positive predictive value or PPV)

Precision focuses on the quality of positive predictions by checking how many of the predicted positive cases were actually correct. Precision is emphasized in situations where false positives are costly, like spam filtering where it could cause a person to miss an important email that was put in the spam folder.

- **Use.** Measures how often the model correctly identifies positive cases out of all predicted positives
- **Formula.**

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Applicable algorithms.** Logistic regression, decision trees, random forests, SVM, neural networks.
- **Business application.** Precision is critical in fraud detection, where false positives block legitimate transactions, or spam filtering, where false positives might send important emails to the spam folder. In medical testing, while false positives can lead to unnecessary follow-up tests, false negatives (missing a real condition) are typically more concerning.
- **Scale.** 0 to 1 (or 0 to 100 percent).
- **Desirable score.** Higher is better, but a high precision may come at the cost of lower recall, meaning more actual positive cases could be missed.
- **Example calculation using the spam classification confusion matrix.** From our spam email classification example, we use the following confusion matrix values:
  - True positives (TP) = 50
  - False positives (FP) = 5

  Applying the formula:

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP + FP} \\ &= \frac{50}{50 + 5} \\ &= \frac{50}{55} \approx 0.91 \, (91\%) \end{aligned}$$

- **Interpretation.** The model correctly identifies 91 percent of the emails it classifies as spam. However, 5 out of 55 emails were incorrectly flagged as spam (false positives), which could inconvenience users by sending important emails to the spam folder.

3. Recall (also known as sensitivity or true positive rate)

Recall focuses on catching as many positive cases as possible. This metric is stressed when false negatives are costly.

- **Use.** Measures how well the model identifies positive cases out of all actual positives.
- **Formula.**

$$\text{Recall} = \frac{TP}{TP + FN}$$

- **Applicable algorithms.** Logistic regression, decision trees, random forests, SVM, neural networks.
- **Business application.** Recall is critical in medical testing, where missing an actual illness (false negative) so the patient does not get treatment could be life-threatening, and in fraud detection, where failing to flag fraudulent activity could cause significant financial losses.
- **Scale.** 0 to 1 (or 0 to 100 percent).
- **Desirable score.** Higher is better, but a high recall may come at the cost of lower precision, meaning more false positives could occur.
- **Example calculation using the spam classification confusion matrix.** From our spam email classification example, we use the following confusion matrix values:
  - True positives (TP) = 50
  - False negatives (FN) = 10

  Applying the formula:

$$\begin{aligned} \text{Recall} &= \frac{TP}{TP + FN} \\ &= \frac{50}{50 + 10} \\ &= \frac{50}{60} \approx 0.83\,(83\%) \end{aligned}$$

- **Interpretation.** The model successfully identifies 83 percent of actual spam emails. However, 10 spam emails were incorrectly classified as not spam (false negatives), meaning some junk emails could still end up in the inbox. In a high-stakes scenario like disease detection, missing 17 percent of real cases could have serious consequences, so recall is often prioritized in such applications.

4. F1 score (harmonic mean of precision and recall)

The F1 score combines precision and recall into a single number, providing a more balanced measure of a model's performance—especially when the data is imbalanced or when both false positives and false negatives carry costs. It is the *harmonic mean*, not the average, which means it gives more weight to lower values. A model only achieves a high F1 score if both precision and recall are high.

- **Use.** Measures a model's accuracy when both precision and recall are important and

need to be balanced.

- **Formula.**

$$F1 = 2 \times (Precision \times\ Recall) / (Precision\ +\ Recall)$$

- **Applicable algorithms.** Logistic regression, decision trees, random forests, SVM, neural networks.
- **Business application.** F1 is useful in email spam detection, credit card fraud detection, medical diagnostics, and other situations where both types of classification errors (false positives and false negatives) matter.
- **Scale.** 0 to 1 (or 0 to 100 percent).
- **Desirable score.** Closer to 1 is better. A low F1 score usually signals that either precision or recall (or both) are poor.
- **Example calculation using the spam classification confusion matrix.** From our spam email classification example, we use previously calculated values:
  - Precision = 0.91
  - Recall = 0.83

  Applying the formula:

$$F1 = 2 \times (0.91\ \times\ 0.83) / (0.91\ +\ 0.83) \approx \mathbf{0.87}$$

- **Interpretation.** The F1 score of 0.87 shows that the model performs well when both false positives and false negatives matter. It balances the tradeoff between incorrectly flagging legitimate emails as spam (precision) and missing actual spam (recall).

5. Area under the receiver operating characteristic curve (AUC-ROC)

The receiver operating characteristic (ROC) curve originated during World War II when it was developed to assess radar signal detection. During the war, radar operators needed a method to distinguish between actual enemy aircraft and false signals. The ROC curve provided a way to visualize the tradeoffs between sensitivity (recall)—the ability to correctly detect enemy aircraft—and false positive rate—the likelihood of false alarms. This graphical tool helped optimize radar performance. Post-war, the ROC curve was adapted for medical decision making to evaluate diagnostic tests, and today, it is widely used in machine learning to assess classification model performance. The ROC curve plots sensitivity (recall) against false positive rate (1 – specificity) at various classification thresholds. The AUC (area under the curve) summarizes the model's overall ability to separate the two classes. The score quantifies how well a model distinguishes between classes, with higher values indicating better overall classification ability.

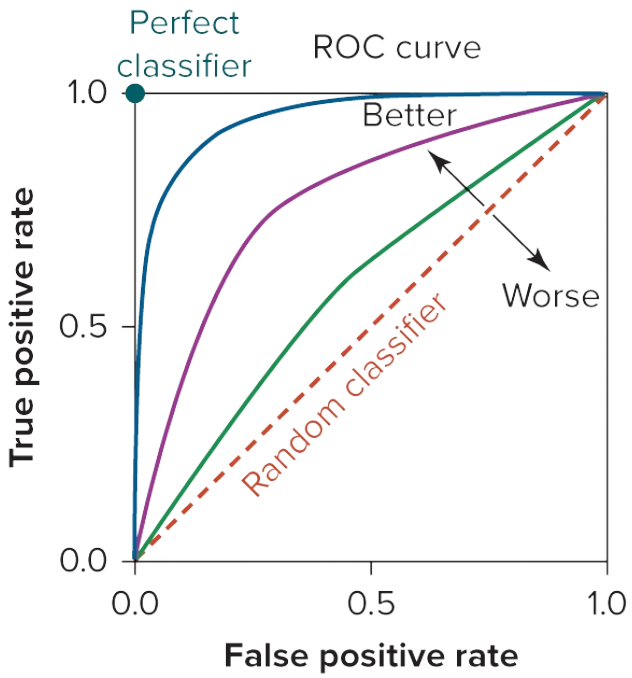
- **Use.** Measures how well a model distinguishes between positive and negative cases.
- **Formula.** AUC-ROC is computed by integrating the area under the ROC curve, typically using

numerical methods.

- **Applicable algorithms.** Logistic regression, decision trees, random forests, SVM, neural networks.
- **Business application.** Useful in fraud detection, medical diagnostics, and credit risk modeling, where distinguishing between classes is critical.
- **Scale.** 0 to 1 (or 0 to 100 percent).
- **Desirable score.** Higher is better.
  - 0.5: No discrimination (random classifier)
  - 0.5–0.7: Poor
  - 0.7–0.8: Acceptable
  - 0.8–0.9: Excellent
  - >0.9: Outstanding
- **AUC-ROC calculation.** While accuracy, precision, and recall can be calculated directly from a confusion matrix, the AUC-ROC requires a more complex process. It involves plotting true positive rates against false positive rates across many threshold values and calculating the area under that curve—an operation that typically requires software and a full set of prediction probabilities, not just classification outcomes. Because this is best handled by analytical tools, we focus here on interpreting the AUC-ROC score as reported by a model, rather than walking through a manual calculation (see **Figure 4S.12**).

FIGURE 4S.12 Interpreting the AUC-ROC

The area under the receiver operating characteristic curve (AUC-ROC) visualizes the performance of a binary classification model by plotting the true positive rate (sensitivity) against the false positive rate at various threshold levels. A model with an AUC close to 1.0 indicates excellent discriminatory power, while an AUC of 0.5 suggests no better performance than random guessing. This curve helps evaluate how well a model distinguishes between classes regardless of the classification threshold.



6. Mean absolute error (MAE) and mean squared error (MSE)

MAE is the average of absolute differences; MSE is the average of squared differences. These metrics help us understand how far off our predictions are in regression models, where we try to predict outcomes based on certain inputs. Simply averaging the errors (differences between predicted and actual values) could be misleading because errors in opposite directions (too high or too low) could cancel each other out, making it seem like our model is more accurate than it is. To solve this, we look at the errors in a way that treats all errors equally important, regardless of direction, by taking the absolute value of each error (making all errors positive) or squaring them.

- **Use.** This approach gives us a clearer picture of a model's accuracy.
- **Formulas.** MAE = $\Sigma\,|\text{Actual} - \text{Predicted}|\ /\ n$; MSE = $\Sigma(\text{Actual} - \text{Predicted})^2\ /\ n$, where n is the sample size.
- **Applicable algorithms.** Linear regression, polynomial regression, regression trees.
- **Business application.** Common in forecasting models, like sales or inventory demand forecasting.
- **Scale.** No fixed upper limit. Starts from 0.
- **Desirable score.** Closer to 0 indicates better performance. The acceptable threshold depends on the specific application and the range of the target variable.

7. Root mean squared error (RMSE)

RMSE is the square root of MSE, indicating error magnitude. It has an advantage over the MSE

in that it is in the same units as the modeled variable.

- **Use.** Magnitude of prediction error.
- **Formula.** RMSE = $\sqrt{\text{MSE}}$
- **Applicable algorithms.** Linear regression, random forest regression, gradient boosting regression.
- **Business application.** Relevant in any regression model where understanding the error magnitude is crucial, like financial forecasting.
- **Scale.** Starts from 0, with no fixed upper limit.
- **Desirable score.** Lower values are better, indicating smaller errors. Like MAE and MSE, the acceptable level depends on the context.
- **Example.** Compare MAE, MSE, and RMSE in a revenue forecasting model.

Month	Actual ($000)	Predicted ($000)	Error	Absolute Error	Squared Error
1	105	100	−5	5	25
2	98	95	−3	3	9
3	102	104	2	2	4
4	110	108	−2	2	4

- **MAE (mean absolute error).** (5 + 3 + 2 + 2) / 4 = **3** → On average, predictions are off by $3,000.
- **MSE (mean squared error).** (25 + 9 + 4 + 4) / 4 = **10.5**
- **RMSE (root mean squared error).** $\sqrt{10.5} \approx$ **3.24** → Slightly higher than MAE, due to penalizing larger errors.

Summary

Each metric provides a unique perspective on a machine learning model's performance and suitability for specific business tasks. Choosing which metrics to utilized is important to correctly assessing your models. Table 4S.2 summarizes the situations where each metric can best be used. By integrating these metrics into the evaluation process, managers can ensure that their AI-driven initiatives are technologically sound and aligned with their business goals and deliver tangible value.

Table 4S.2 **Metric Summaries**

Metric	Use Case	Best When You Want to...	Watch Out for...
Accuracy	General performance	Get a quick overview	Misleading with imbalanced classes
Precision	Spam filtering,	Avoid false positives	May miss true positives

	fraud		
Recall	Medical, fraud	Catch all true positives	May raise false positives
F1 score	Spam filtering, fraud, diagnostics	Balance precision and recall in imbalanced data	Doesn't show which type of error is driving the score
AUC-ROC	Overall classifier ranking	Compare classifiers	May hide class imbalance issues
MAE	Sales/forecasting	Know average error	Treats all errors equally
MSE	Forecasting	Penalize large errors	Sensitive to outliers
RMSE	Forecasting	Interpret error in original units	Sensitive to outliers

Note that there are many other metrics, but these are the most common ones for machine learning algorithms. The scales and desirable scores provide a guideline for interpreting the effectiveness of machine learning models. However, it is crucial to consider them in the context of the specific business problem, as the tradeoffs between different metrics can vary significantly based on the application and the domain. For instance, in medical diagnostics, a higher recall might be prioritized over precision to ensure all potential cases are identified, even at the cost of some false positives.

Algorithm Selection Questions

While evaluating algorithms through metrics is key, the underlying data's quality and quantity play a pivotal role in any machine learning endeavor. Let's discuss why data handling is critical for successful ML applications. The following questions encourage you to apply your understanding of different machine-learning techniques in business scenarios.

1. **Customer churn prediction.** A subscription-based streaming service wants to identify customers likely to cancel in the next month. What machine learning algorithm(s) might be most appropriate for this task? What data would be most useful?
2. **Warehouse inventory grouping.** A logistics firm wants to group items in a warehouse to improve layout efficiency. No predefined categories exist. Which algorithm(s) would you recommend and why?
3. **Predicting housing prices.** A real estate agency wants to predict housing prices based on size, location, and number of bedrooms. What algorithm would be appropriate, and how would you assess its performance?
4. **Product recommendations on a shopping app.** An e-commerce app wants to recommend products to users based on previous shopping behavior. What technique would you suggest and what data would power it?
5. **Analyzing high-dimensional customer data.** A company has hundreds of variables in its customer survey dataset. They want to reduce noise and discover key patterns. What algorithm(s) would help and why?

Metrics and Evaluation Questions

The following questions encourage critical thinking about how to assess model performance and which metrics to prioritize.

1. **Balancing precision and recall.** In a fraud detection system, why might a company prioritize recall over precision? Can you think of a business example where precision would be more important?
2. **Misleading accuracy.** You build a classification model that predicts loan default with 95 percent accuracy—but only 5 percent of applicants default. Why might this model be misleading? What other metrics would help?
3. **Comparing two classification models.** Two models have similar accuracy, but one has higher AUC-ROC. What does this tell you about their performance, and when would that matter?
4. **Model evaluation in imbalanced datasets.** In a health screening app, only 1 percent of users are expected to have a serious condition. What evaluation metrics should you use, and which ones might give a false sense of model quality?
5. **Metric calculation.** Logistic regression was applied to tax returns to identify which ones were fraudulent. Those who were identified as fraudulent would undergo an audit. Calculate the following metrics for the confusion matrix below.
   a. Accuracy
   b. Precision
   c. Recall

Actual / Predicted	Predicted Fraud	Predicted No Fraud	Total
Actual fraud	9	2	11
Actual no fraud	5	134	139
Total	14	136	150

CHAPTER 5
Neural Networks and Deep Learning



Andrey Suslov/Shutterstock

Learning Objectives

- **LO5-1** Compare neural networks and deep learning.
- **LO5-2** Explain the importance of neural networks in modern business applications.
- **LO5-3** Analyze how businesses use deep learning to enhance operations, decision making, and customer engagement.
- **LO5-4** Evaluate key business applications of deep learning, particularly in natural language processing.
- **LO5-5** Describe emerging capabilities in deep learning and their potential impact on business innovation.

INTRODUCTION

Neural networks are the workhorses of modern artificial intelligence. They process information in a manner that mimics the functioning of biological neurons. They have become essential for businesses, enabling the discovery of patterns and insights in large datasets. As companies increasingly embrace complex, data-driven strategies, deep learning—an advanced form of neural networks—has expanded the possibilities, allowing organizations to tackle challenges in image recognition, natural language understanding, and autonomous systems. This chapter explores how businesses use neural networks and deep learning, their challenges and limitations, and the emerging trends shaping the future of AI-driven innovation.  **A Technical Supplement at the end of the chapter provides more detailed explanations of types of neural networks for interested readers.**

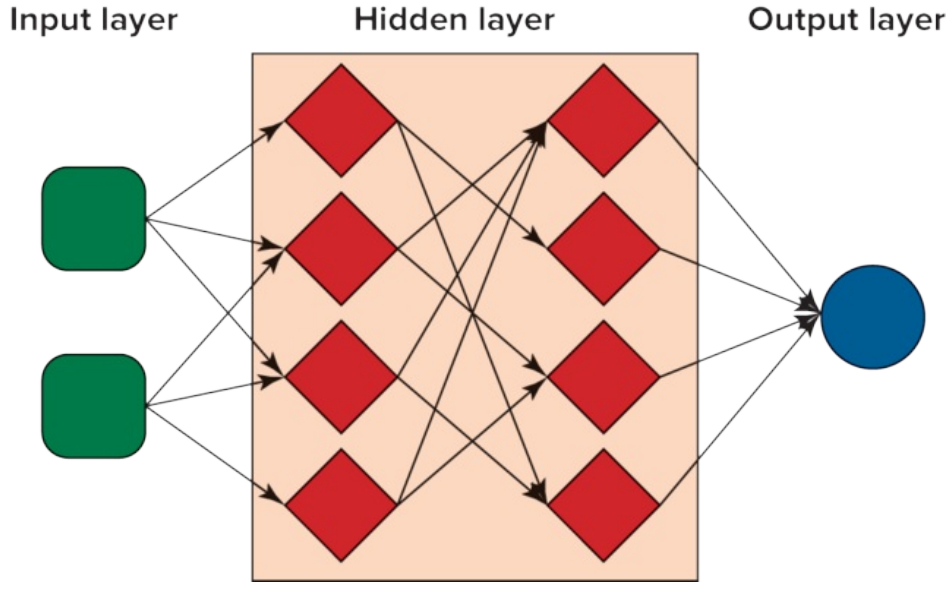
CLARIFYING NEURAL NETWORKS AND DEEP LEARNING (LO 5-1)

While the terms *neural networks* and *deep learning* are sometimes used interchangeably, they are not the same. A **neural network** is a machine learning model inspired by how biological neurons communicate and process information. Neural networks have existed for decades, typically using simple architectures with one or two hidden layers.

As shown in **Figure 5.1** (which we introduced in **Chapter 1**), a neural network is a system of connected layers that transforms input data into predictions or decisions. It starts with an **input layer** that receives raw data—like a customer's transaction history or a scanned image. That data flows through one or more **hidden layers**, where each neuron (a simple processing unit) adjusts the signal based on its internal settings. These adjustments help the network detect patterns—such as identifying whether a customer might churn or recognizing an object in a photo. The final result comes from an **output layer**, which delivers a prediction, classification, or recommendation. Neural networks improve through training. After making a prediction, the network compares its output to the correct answer and adjusts its internal signal settings, called **weights and biases**, to reduce future errors. Repeating this process across many examples helps the model get better over time.

FIGURE 5.1 Neural Network Layers

A simple artificial neural network consists of an input layer, one or more hidden layers, and an output layer. Each node (or "neuron") processes information and passes it forward, allowing the network to learn patterns and make predictions based on data.



Deep learning refers to a more recent evolution of neural networks, using many hidden layers to

learn complex patterns from large datasets. Like all other forms of AI, deep learning works best when trained on high-quality, well-curated datasets. However, deep learning models typically require much larger volumes of data to perform effectively, particularly when learning from unstructured inputs such as images, text, or audio. These deeper architectures have enabled major breakthroughs in areas such as image recognition, language processing, and autonomous systems.

However, while deep learning has expanded what AI systems can do, it is not always the best choice. For smaller datasets, limited computing power—or when model explainability is critical—traditional machine learning methods like decision trees, logistic regression, or gradient boosting can outperform deep learning in accuracy, speed, and cost. The key is to match the model to the task, the data, and the business constraints.

In this chapter, we use the term *neural networks* broadly, but when discussing large, multilayered models driving today's AI advances, we specifically mean *deep learning models*. Readers interested in a more technical description of how these systems learn—including forward propagation, backpropagation, and model architectures—can find a deeper explanation in the Technical Supplement at the end of this chapter.

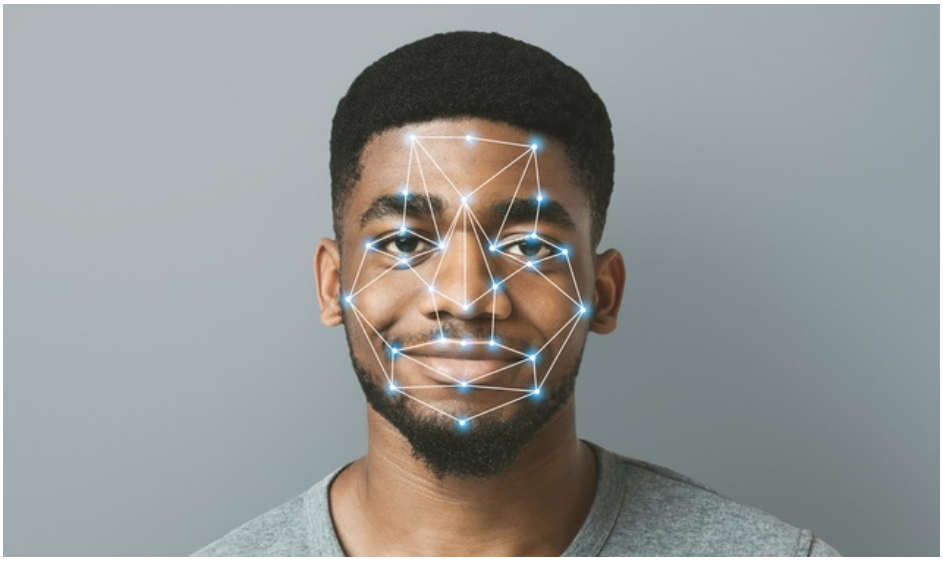
THE GROWING IMPORTANCE OF NEURAL NETWORKS (LO 5-2)

Neural networks have become a cornerstone of modern business innovation, enabling organizations to harness the full potential of their data. Their capacity to process vast amounts of information, uncover intricate patterns, and automate decision making has revolutionized industries.

Business Applications of Neural Networks

Neural networks excel in various domains, providing tangible benefits that enhance efficiency, decision making, and customer experience. Some prominent applications include:

- ***Customer service.*** Neural networks power chatbots and virtual assistants, automating responses to customer inquiries across multiple platforms. These systems provide consistent, accurate support 24/7, improving customer satisfaction while reducing operational costs.
- ***Financial forecasting.*** Businesses in the financial sector rely on neural networks to predict stock market trends, assess loan risks, and detect fraudulent transactions. For instance, neural networks can analyze historical data to identify anomalies in real time, helping prevent fraud before it occurs.
- ***Image and speech recognition.*** Neural networks are crucial in processing unstructured data like images and audio. In health care, neural networks analyze medical images to diagnose diseases accurately. In consumer electronics, voice recognition capabilities allow users to interact seamlessly with devices like smartphones and smart speakers.



Facial recognition systems identify unique features—such as the distance between the eyes, the shape of the nose, and the contour of the jawline—to create a mathematical representation of a person's face. This digital map enables AI models to accurately compare and match faces.
Prostock-studio/Shutterstock

Why Businesses Are Adopting Neural Networks

Several factors contribute to the widespread adoption of neural networks in business:

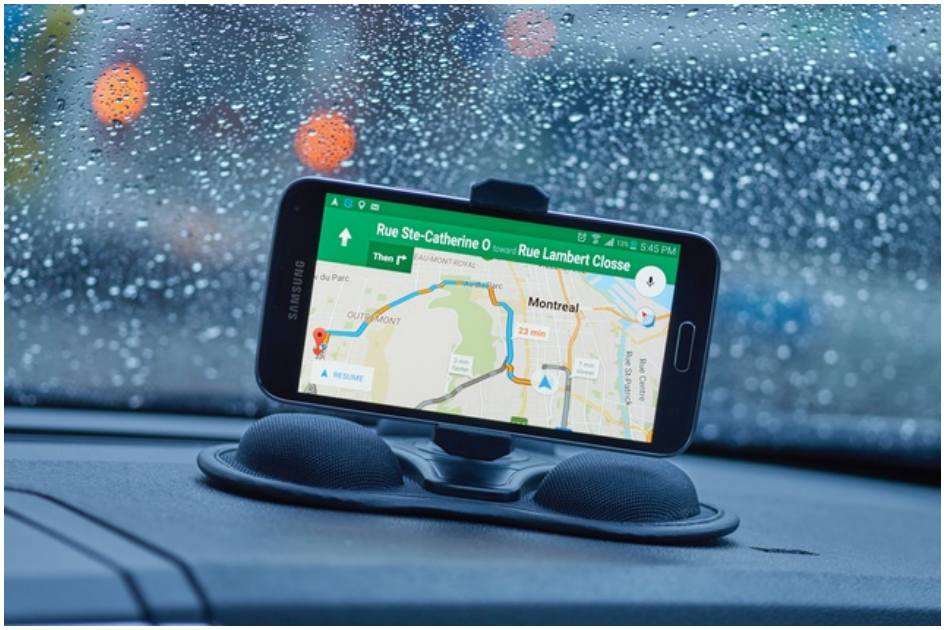
- ***Handling complexity.*** Neural networks excel at modeling complex, nonlinear relationships in data, making them suitable for solving problems that traditional algorithms cannot solve.
- ***Automation and efficiency.*** Neural networks free human resources for strategic activities by automating repetitive tasks and optimizing workflows. For example, predictive maintenance systems powered by neural networks minimize equipment downtime and operational disruptions.
- ***Personalization.*** Retail and e-commerce platforms use neural networks to provide personalized product recommendations, tailoring experiences to individual customers' preferences. This level of customization drives engagement and loyalty.
- ***Scalability.*** Neural networks can process large-scale data in real time, enabling businesses to handle the growing data deluge of the digital age.

Real-World Impacts

An effective heart disease prediction system (EHDPS) developed in India exemplifies the transformative potential of neural networks.[1] The system utilizes 15 medical parameters, including age, blood pressure, and cholesterol, to predict patients' risk levels for heart disease, enabling early intervention and treatment. The heart disease prediction model using the random forest algorithm (see **Chapter 4**) on a neural network achieved an accuracy of 90 percent.[2] Similar applications are also enhancing diagnostic accuracy and improving patient outcomes.

Retail giants like Amazon and Netflix use neural networks to understand customer behavior better and deliver personalized recommendations. Netflix analyzes viewing patterns to suggest shows and movies tailored to individual preferences, while Amazon examines purchase history and browsing habits to recommend products. These personalized experiences improve customer satisfaction, boost sales, and strengthen customer loyalty.[3]

Neural networks also play a critical role in improving supply chain efficiency by predicting demand fluctuations and optimizing inventory levels, helping businesses reduce costs and minimize waste. Companies like FedEx use neural networks to anticipate potential delivery delays and adjust routes in real time, improving both operational reliability and customer satisfaction. AI-driven logistics solutions can enhance operational efficiency by up to 30 percent through improved resource allocation and dynamic routing.[4]



Just as a GPS system recalculates a route after a missed turn or roadblock, neural networks used in logistics can adapt to unexpected events—such as traffic accidents or weather delays—by dynamically rerouting deliveries to maintain efficiency.

dennizn/Shutterstock

DEEP LEARNING (LO 5-3)

Deep learning is a specialized area within machine learning where computer systems, inspired by the human brain, learn to recognize patterns and make decisions from data. Unlike traditional machine learning, which often requires human experts to identify and handpick important signals for making predictions, deep learning automates this process. It can learn to identify significant patterns or features directly from the data, permitting it to tackle more complex problems, such as understanding natural human speech or identifying objects in images, which would be challenging with traditional machine learning.

Deep learning's strength in managing and making sense of large volumes of data makes it especially valuable in today's data-rich business environment. While traditional machine learning methods rely on structured data and manual feature engineering, deep learning models can automatically learn patterns from vast amounts of unstructured data like images, audio, and text. Deep learning typically requires more data and computing power but offers greater flexibility and accuracy in solving complex problems. A more detailed comparison of the two approaches is provided in the Technical Supplement at the end of this chapter.

Deep Learning's Role in Business

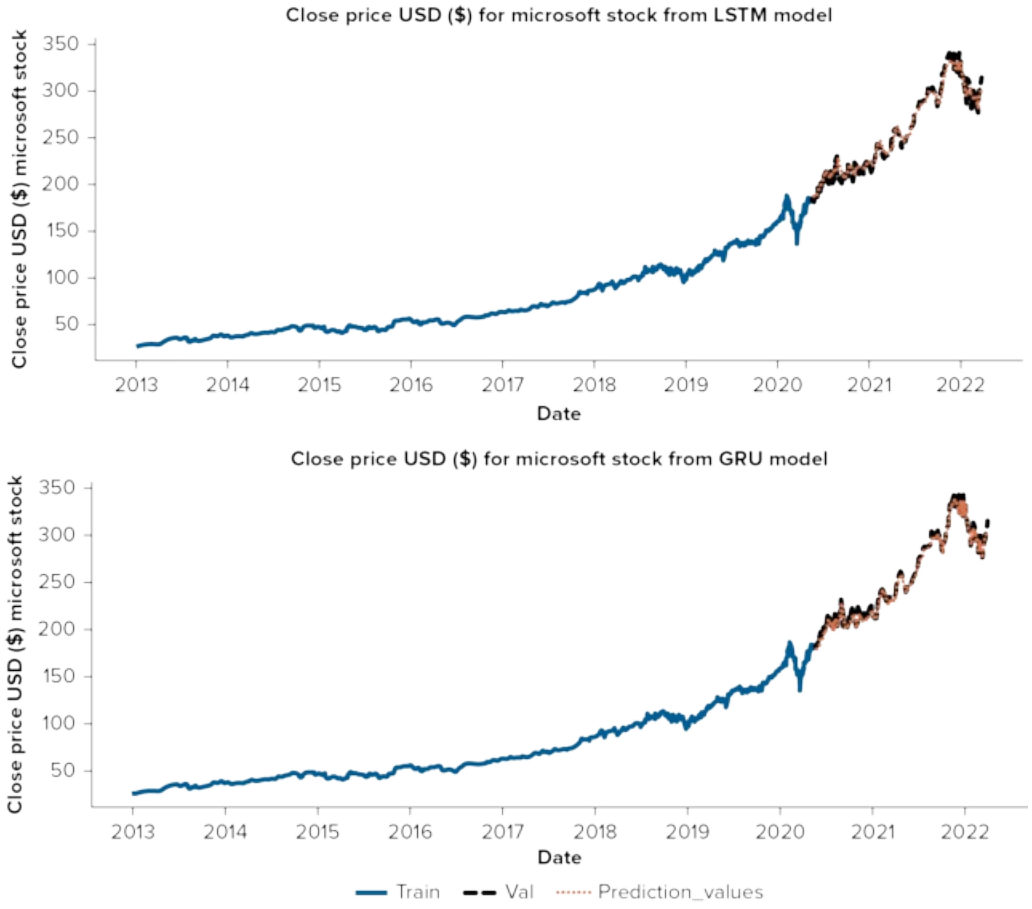
Deep learning has revolutionized how businesses approach vast and complex datasets, which are increasingly common in the digital era. Its significance in contemporary business applications is multifaceted:

- ***Capability to handle complexity.*** Deep learning models, particularly those with many layers (hence the term "deep"), excel at identifying intricate patterns in data. Traditional machine learning techniques may falter with the sheer volume and complexity of modern datasets, but deep learning thrives on them, allowing companies to uncover patterns and insights that traditional analytics tools often miss, improving the speed and quality of business decision making.
- ***Enhanced decision making.;*** In business, decision making is often contingent on analyzing large sets of variables. Deep learning automates and refines this analysis, leading to more informed and rapid decision-making processes. For example, in financial services, deep learning can predict market trends by analyzing vast arrays of past transaction data (see **Figure 5.2**).

FIGURE 5.2 Predicting Microsoft's Closing Stock Price Using Deep Learning

Deep learning models can be trained to recognize patterns in historical stock data to forecast future prices. These graphs use two different deep learning models to compare predicted and actual closing prices for Microsoft, illustrating the models' ability to capture trends. The blue lines represent the historical data that the models were built on. The red lines are the actual values during the forecast period and

the yellow lines are the forecasts.



McGraw Hill

- ***Customer insights and personalization.*** Deep learning algorithms can sift through massive quantities of customer data to provide personalized experiences. They can recommend products, tailor services, and anticipate needs by learning individual customer behaviors, a capability that is invaluable in marketing and retail.
- ***Operational efficiency.*** Businesses are leveraging deep learning to optimize operations. For instance, in supply chain management, deep learning can predict demand fluctuations and optimize inventory, reducing waste and costs.
- ***Innovation in products and services.*** Deep learning is at the heart of developing innovative products and services. For instance, it enables the creation of sophisticated AI assistants and advances in autonomous vehicles, both of which rely on deep learning's ability to process complex sensory data.
- ***Risk management.*** Analyzing vast datasets also allows for better risk management. In cybersecurity, deep learning models can detect patterns indicative of threats by analyzing network traffic, thereby enhancing security measures.
- ***Automation in social media.*** Amarra, a small New Jersey–based manufacturer and distributor of

formal gowns, adopted a generative AI platform to automate its product descriptions and social media content. In parallel, the company used AI to analyze customer browsing behavior, purchase history, and seasonal trends to better forecast demand. As a result, Amarra reduced overstock by 40 percent and increased the speed of content generation tenfold.[5]

Every industry, from health care with improved diagnostic tools to agriculture with crop yield optimization, is transforming thanks to deep learning's ability to process extensive and intricate datasets.

Deep learning is a powerful engine for extracting meaning from the data deluge characteristic of the modern business landscape. Its advanced pattern recognition capabilities, ability to learn autonomously, and scalability make it an indispensable tool for businesses looking to maintain a competitive edge through data-driven strategies.

Deep Learning in Natural Language Processing

Natural language processing (NLP) enables computers to understand, interpret, and generate human language. Businesses use NLP to power chatbots, automate document analysis, monitor customer sentiment, and translate content across languages. While early NLP systems made important progress, they were limited by rigid rule sets and statistical models. Deep learning has dramatically expanded the capabilities of NLP by allowing AI systems to learn complex patterns in language with minimal human programming.

Early Approaches to Language Translation

Early machine translation systems, such as SYSTRAN in the 1970s and Babel Fish in the late 1990s, relied heavily on predefined rules and dictionaries. Developers manually programmed how words and phrases should be mapped between languages. While these systems were groundbreaking at the time, they struggled with nuances like slang, idioms, cultural references, and context-specific meanings. Literal translations often produced awkward or even misleading results. These early systems highlighted a critical limitation: Language is highly complex and context dependent, making it difficult for rigid, hand-crafted models to perform consistently well.

How Deep Learning Changed NLP

Deep learning transformed NLP by enabling models to learn language patterns directly from massive datasets, without the need for explicit programming of every rule. Models trained on large volumes of text could recognize how words relate to one another, pick up on grammatical structures, and even capture emotional tone. Key breakthroughs included:

- ***Word embeddings.*** Techniques like Word2Vec and GloVe allowed words to be represented as vectors in a multidimensional space, capturing their meaning relative to other words.

- ***Sequence modeling.*** Recurrent neural networks (RNNs) enabled AI systems to generate and understand sequences of words, improving translation, summarization, and conversation modeling.

Deep learning made it possible for NLP systems to:

- ***Handle ambiguous language.*** Understanding context helps AI distinguish among multiple meanings of the same word.
- ***Capture long-range dependencies.*** AI could relate concepts mentioned earlier or later in a document or conversation.
- ***Improve over time.*** Deep learning models could continuously refine their language understanding as more data became available.
- ***Power business applications.*** Today, deep learning powers a wide range of NLP tasks, including automated language translation, text classification, summarization, and sentiment detection.

Businesses now rely on NLP to power customer service chatbots, translate content, classify documents, and monitor public sentiment. These applications became viable when deep learning enabled AI systems to learn language patterns directly from massive datasets.

NLP APPLICATIONS IN BUSINESS (LO 5-4)

Business NLP applications are reshaping how companies interact with customers and manage their operations. These innovations enhance efficiency and transform customer engagement and global communication dynamics. Let's explore some of these applications in the business world:

- ***Hootsuite.*** Hootsuite has emerged as a vital tool for managing and analyzing social media presence. This platform allows companies to streamline their social media activities across various channels, from scheduling posts to engaging with customers. A key feature is its capability to perform sentiment analysis using NLP. By analyzing the tone and context of social media conversations, Hootsuite assigns a positive, negative, or neutral sentiment to each mention of the company or product, providing businesses with critical insights into public perception of their brand. This functionality enables companies to swiftly identify and respond to customer sentiments, manage their online reputation, and tailor their marketing strategies to better align with audience attitudes and trends.
- ***iTranslate.*** iTranslate, a prominent application in NLP, showcases the power of AI in breaking down language barriers. With support for more than 100 languages, iTranslate offers text translation and voice-to-voice communication, enhancing its appeal for travelers and international business professionals. Its NLP capabilities extend to attempting to encompass dialects and colloquialisms, offering users accurate and contextually appropriate translations. This app highlights how advancements in NLP have made real-time, on-the-go translation a reality, transforming how we interact across different cultures and languages in our increasingly globalized world.
- ***Chatbots.*** Chatbots, too numerous to list, have revolutionized customer service and engagement. These AI-driven chatbots can understand and respond to customer inquiries in a humanlike manner, offering real-time assistance. This technology allows businesses to scale customer support efficiently while improving user satisfaction. By leveraging NLP, they interpret the nuances of human language, comprehend customer requests, and provide relevant, accurate responses. This enhances customer experience and increases efficiency by handling a high volume of queries without human intervention. Businesses across various sectors use NLP chatbots for functions ranging from answering FAQs to providing personalized recommendations, significantly enhancing customer interaction and operational efficiency. These will be discussed in detail in **Chapter 6** and **Chapter 7**.

Google Translate owes much of its efficiency to the advent of *transformer models*. Introduced by Google in 2017, these models revolutionized machine translation by enabling more nuanced and context-aware translations. Unlike earlier methods that processed text linearly, transformers handle words in relation to all other words in a sentence, capturing subtleties and complexities of language with remarkable accuracy. This breakthrough underpins Google Translate's ability to provide fluent and coherent translations across more than 130 languages, facilitating seamless communication worldwide.

Limitations of Neural Networks and Deep Learning

While neural networks and deep learning offer powerful capabilities, they also present important limitations that businesses must consider. These highly data-intensive models require large volumes of high-quality data for effective training. They are often criticized for their black-box nature, where the decision-making processes are challenging to interpret, raising concerns about transparency and trust. Deep learning systems also impose significant computational demands, often requiring specialized hardware and substantial investment to train and maintain effectively.

EMERGING CAPABILITIES IN DEEP LEARNING (LO 5-5)

Deep learning continues to evolve rapidly, sparking business innovation. Several emerging developments are expanding what neural networks can do, reshaping how companies innovate, compete, and serve customers. Understanding these trends is key to preparing for the next wave of AI-driven business transformation. Here are a few of the trends to be aware of:

- ***Unsupervised and self-supervised learning.*** An important emerging trend is the rise of **unsupervised learning** and **self-supervised learning**, where AI systems detect patterns or generate their own labels from data without requiring extensive manual annotation (see **Chapter 4**). These techniques allow businesses to train models faster and more affordably, especially when large, labeled datasets are unavailable. As AI evolves, unsupervised and self-supervised approaches open new opportunities for analyzing customer behavior, automating operations, and discovering hidden insights.[6]
- ***NLP.*** Further advances in NLP are another key trend. We are witnessing the emergence of more sophisticated chatbots that converse with unprecedented naturalness, providing seamless customer service experiences. Automated sentiment analysis is also revolutionizing, offering more profound and accurate insights into public opinion and consumer behavior. These developments in NLP are not just incremental improvements but are transformative, setting the stage for AI systems that interact, understand, and respond in ways that are increasingly indistinguishable from human communication.
- ***AI and IoT integration.*** Integrating AI with the Internet of Things (IoT) is reshaping industries by enabling more intelligent, dynamic systems. Embedding AI into IoT devices enhances real-time data collection, analysis, and decision making. From smart homes that learn occupants' preferences to industrial systems that predict maintenance needs and optimize operations, AI-driven IoT applications improve efficiency, personalization, and automation across sectors.

Foundation Models

While the concept of **foundation models**, large neural networks pretrained on broad datasets and adaptable to many tasks, emerged in research in the late 2010s, their deployment in business applications has surged only recently. These models, such as OpenAI's GPT-4, Google's PaLM, and Meta's LLaMA, are reshaping how organizations approach AI by reducing the time, cost, and effort needed to create high-performing solutions. Companies are utilizing foundation models to scale AI across departments, from customer service to risk management, without needing to build separate models for each task. This scalability and adaptability fuel faster adoption and make advanced AI more accessible to businesses of all sizes.[7]

Morgan Stanley uses GPT-4 as an internal AI assistant to support its wealth management advisors. This AI assistant helps organize and retrieve information from more than 100,000 pages of research and policy documents, demonstrating how foundation models can improve efficiency, enhance client service, and make large-scale knowledge resources more accessible.[8]

Multimodal Learning

Multimodal learning is another significant advance, enabling AI systems to process and relate information from multiple types of data simultaneously, such as text, images, audio, and video. This capability opens new frontiers for businesses like creative content generation, intelligent search engines, and advanced customer service bots. By integrating multimodal understanding, companies can build AI applications that operate more like human assistants—capable of synthesizing diverse information to deliver more natural, helpful experiences.[9]

Adobe uses multimodal AI in its Creative Cloud suite to enhance content creation tools. By integrating text, images, audio, and video, Adobe's AI-powered features help users create more engaging and dynamic content.[10]

Few-Shot and Zero-Shot Learning

Few-shot learning and zero-shot learning are reducing the amount of labeled data needed to train effective AI models. In **few-shot learning**, models generalize from just a handful of examples; in **zero-shot learning**, they generalize from none. These methods make AI deployment faster and more affordable, especially in industries where labeled data is scarce or costly to obtain.[11]

Where labeled data can be scarce, few-shot learning and zero-shot learning can enable early disease detection and personalized treatment plans. For instance, a 2023 study showed that few-shot learning models achieved an accuracy of 87 percent in diagnosing rare diseases with minimal data.[12]

Energy-Efficient Deep Learning (Green AI)

The environmental impact of large deep learning models has sparked a growing focus on **energy-efficient AI** (often called **green AI**).[13] Businesses are adopting novel approaches to build AI systems that are smaller, faster, and less power-hungry, helping lower operational costs while reducing carbon emissions.[14] By improving model efficiency without sacrificing performance, companies can scale their AI initiatives sustainably and meet growing expectations for corporate social responsibility.

Deep Learning for Personalized Adaptive Systems

Businesses are also leveraging deep learning to create **adaptive systems** that personalize experiences and respond dynamically to user behavior. From personalized education platforms to dynamic pricing engines, these systems are reshaping industries by offering highly responsive, tailored experiences. For example, Duolingo uses deep learning models to power an adaptive learning engine that adjusts lesson difficulty and content based on user performance in real time, improving engagement and learning outcomes.[15]

Deep Learning and Virtual Reality

Deep learning and virtual reality (VR) are cutting-edge technologies that can be interconnected in many ways, enhancing each other's capabilities and applications. Here are some ways they can be related.

- ***Enhanced VR experiences.*** Deep learning algorithms can generate dynamic content and adapt the VR experience based on user behavior. Real-time story generation is a good example of this.
- ***Improved object and environment recognition.*** AI can help VR systems better understand and interpret the virtual world, making interactions more natural. Deep learning can enhance object detection and tracking in VR environments. This can be used with spatial understanding as well as dynamic object properties (e.g., turning on a light switch).
- ***Emotion recognition and responsive avatars.*** Deep learning algorithms can analyze facial expressions and body language to create more realistic and emotionally responsive avatars in VR.
- ***Personalized VR experiences for training.*** Deep learning can analyze user preferences and behaviors to tailor VR content and interactions to individual users. Performance tracking and real-time adjustments in training are examples of this.
- ***Real-time language translation.*** AI can provide real-time translation services in VR environments, enabling communication between users speaking different languages. The Meta Ray-Ban glasses have functionality to provide this with a combination of augmented reality (AR) imaging and AI translation.
- ***Data visualization.*** Deep learning can help process and visualize complex data sets in VR, making it easier for users to understand and interact with information. This functionality can be used in predictive modeling, pattern recognition, and interactive what-if scenarios. Being able to "walk into" your data in VR is a game-changer.
- ***Adaptive difficulty in VR games.*** AI can dynamically adjust the difficulty of VR games based on player performance and skill level.



Artificial intelligence and virtual reality are increasingly intertwined, with deep learning enabling real-time adaptation, emotion recognition, and personalized content in immersive environments. From intelligent training simulations to dynamic storytelling and data visualization, this integration is expanding the boundaries of how we interact with digital experiences.
Gorodenkoff/Shutterstock

These are just a few examples of how deep learning and virtual reality can work together to create more immersive, intelligent, and responsive experiences. As both technologies continue to advance, we can expect to see even more innovative applications and integrations in the future.

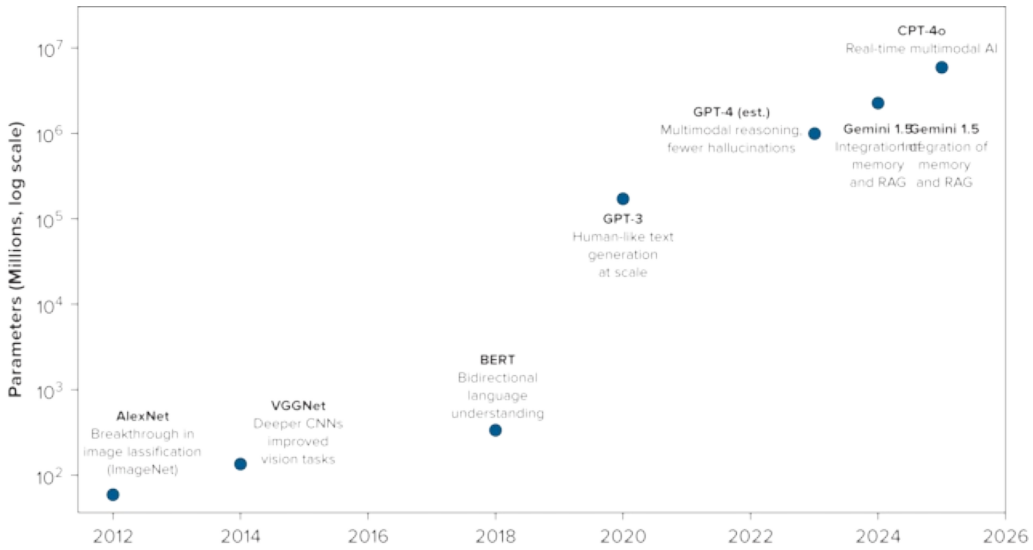
Growth of Neural Network Models

These innovations are part of a broader trend in which deep learning models are becoming more capable, more generalized, and more expansive in scale.

The pace of progress in neural network models has been nothing short of exponential. Over the past decade, the number of parameters, the internal values that help a model learn, has skyrocketed from tens of millions in early image recognition models like AlexNet to hundreds of billions in generative language models like GPT-3, and now into the trillions with emerging multimodal systems like GPT-4. As illustrated in **Figure 5.3**, this dramatic increase in scale has enabled AI systems to move from recognizing images and translating text to performing sophisticated reasoning across language, vision, and even audio. This growth reflects technical improvements and transformation in what AI can do. It has opened up business opportunities and governance challenges that come with it.

FIGURE 5.3 The Explosive Growth of Neural Network Models

The capabilities of neural networks have accelerated dramatically in recent history. Model size, performance, and versatility have increased exponentially, fueling advances in image recognition, language processing, and multimodal AI. This rapid pace requires businesses and educators alike to stay continuously informed.



Epoch AI. AI Model Timeline. https://epoch.ai/data

McGraw Hill

SUMMARY OF KEY POINTS

1. **Neural networks vs. deep learning.** A neural network is a machine learning model inspired by how biological neurons communicate and process information. Deep learning refers to a more recent evolution of neural networks, using many hidden layers to automatically learn complex patterns from large datasets.
2. **Neural networks in modern business.** Neural networks, inspired by biological processes, allow businesses to uncover patterns, automate tasks, and improve decision making.
3. **Deep learning enhances business performance.** Deep learning builds on neural networks by enabling systems to process complex, unstructured data, allowing businesses to personalize services, optimize operations, and innovate at scale.
4. **Natural language processing.** Deep learning powers applications like NLP enable businesses to automate communication, enhance customer service, and unlock insights from text data.
5. **Emerging trends.** Advances such as foundation models, few-shot learning, green AI, and deep learning–driven AR/VR create new opportunities and competitive pressures for businesses.

QUICK WINS

- **Identify opportunities to apply neural networks.** Look for business processes where pattern recognition, classification, or prediction could add immediate value.
- **Leverage deep learning to optimize operations.** Explore how deep learning applications can automate tasks, personalize customer experiences, or enhance organizational decision making.
- **Experiment with natural language processing tools.** Implement NLP solutions, such as chatbots or document summarization, to improve communication efficiency and customer engagement.
- **Monitor emerging AI trends.** Stay informed about advances like foundation models, multimodal AI, and green AI to position your business competitively in a rapidly evolving marketplace.

KEY TERMS

Adaptive systems: AI-driven systems that adjust their behavior or responses based on real-time user interactions or environmental changes.

Deep learning: An advanced form of machine learning using multilayered neural networks to learn complex patterns from large datasets automatically.

Energy-efficient AI (green AI): Efforts to design and deploy AI models that require less computational power and energy, reducing environmental impact while maintaining performance.

Few-shot learning: A machine learning method where models are trained to perform tasks based on only a small number of labeled examples.

Foundation models: Large-scale, pretrained AI models that can be adapted for a wide range of downstream tasks with minimal additional training.

Hidden layers: In a neural network, a layer of neurons located between the input layer and the output layer. Hidden layers transform the incoming data by applying weights, biases, and activation functions, enabling the network to detect patterns and relationships that are not directly observable in the raw input.

Input layer: The first layer of a neural network that receives raw data—such as numbers, images, or text—and passes it to the next layer for processing.

Multimodal learning: A method where AI systems can simultaneously process and relate information from multiple data types, such as text, images, and audio.

Neural network: Machine learning model inspired by the structure of the human brain, designed to recognize patterns and relationships in data.

Output layer: The final layer of a neural network that produces the model's result—such as a classification, prediction, or decision—based on the data processed by earlier layers.

Self-supervised learning: A machine learning approach where models generate their own labeled data from raw inputs, reducing the need for manual labeling.

Unsupervised learning: A machine learning technique where models identify patterns or groupings in data without using labeled examples.

Weights and biases: Internal values in a neural network that determine how input signals are combined and adjusted as they move through the network. Weights control the strength of each connection, while biases allow the network to shift or fine-tune its outputs.

Zero-shot learning: A deep learning technique where a model makes accurate predictions about categories it was never explicitly trained on by using patterns and relationships learned from related data.

DISCUSSION QUESTIONS

1. **Neural network advantages.** Neural networks are modeled after the human brain but operate in fundamentally different ways. What are the advantages and disadvantages of neural networks compared to traditional rule-based systems in business applications?
2. **Business applications.** How can businesses use deep learning to gain a competitive advantage in customer experience, forecasting, or product recommendations? Provide examples.
3. **NLP.** In what ways has deep learning advanced the capabilities of natural language processing? How might this reshape communication in customer service or legal/compliance settings?
4. **Deep learning capabilities.** What trends in deep learning do you believe will have the most significant impact on your industry in the next 5 years? Explain your reasoning.
5. **Skills for the future.** If deep learning tools continue to become more accessible, what skills will business professionals need to work effectively alongside these systems?

EXERCISES

1. **Role-playing a business scenario.** Assume different roles within a company (e.g., CEO, data scientist, marketing manager) and discuss implementing a deep learning project, considering aspects like budget, expected outcomes, and potential challenges.
2. **Diagram exercise.** Sketch and label a basic neural network with input, hidden, and output layers. Then, briefly describe what happens at each stage.
3. **Trend tracker.** Research a current trend in deep learning (e.g., self-supervised learning, green AI, foundation models). Summarize what it is and how it could affect one business function (e.g., HR, marketing, logistics).
4. **Evaluating AI readiness.** Perform an AI readiness assessment for a business. Identify strengths, weaknesses, opportunities, and threats (SWOT analysis) related to deep learning adoption.
5. **Future trends brainstorming.** Conduct a brainstorming session on future trends in deep learning and their implications for business. Discuss how companies can prepare for and capitalize on these trends.

CASE STUDY 5.1: IBM WATSON OPENSCALE—EXPLAINABLE AI

Artificial intelligence systems, especially those using advanced machine learning techniques like deep learning, often function as "black boxes." Their decision-making processes are not readily understandable to humans, making it difficult to explain or justify AI-generated outcomes. This lack of transparency can erode user trust, complicate regulatory compliance, and obscure hidden biases within models. The inability to explain AI decisions can have profound consequences in critical sectors like health care, finance, and criminal justice. Therefore, explainability is not simply a technical challenge; businesses must use AI ethically, transparently, and responsibly for people to trust AI decisions.

IBM Watson and OpenScale

IBM Watson emerged as one of the first widely recognized AI platforms capable of processing and analyzing large volumes of structured and unstructured data. Famous for its 2011 victory on the television quiz show *Jeopardy!*, Watson has evolved into a versatile business platform supporting finance, retail, and education applications. In finance, it enhances customer service through intelligent chatbots and aids in risk management and fraud detection by processing complex data patterns. The retail sector sees Watson powering customer engagement tools, offering personalized shopping experiences through insights derived from customer data. Additionally, in education, Watson's AI capabilities are harnessed to create adaptive learning platforms that provide customized learning experiences to students.

However, as Watson expanded into sensitive domains, IBM recognized the need to address growing concerns over AI transparency. To build trust and meet regulatory expectations, IBM developed Watson OpenScale, a tool designed to provide interpretability for AI decision making. OpenScale identifies which factors drive model predictions, allowing businesses to understand and audit outcomes such as loan approvals or medical recommendations. IBM positioned OpenScale as a key element in AI governance, helping companies monitor compliance, identify biases, and ensure fairness.

Challenges and Tradeoffs

Implementing explainability tools like OpenScale presents technical and organizational challenges. Deep learning models are inherently complex, making their decision making difficult to decipher and translate into human-understandable terms. Creating accurate, meaningful explanations often requires additional algorithms, increasing the system's complexity and cost. A

persistent tradeoff exists between model performance and transparency: Simplifying models for explainability often reduces their predictive accuracy. Furthermore, businesses must navigate technical hurdles, organizational resistance, regulatory uncertainty, and the difficulty of embedding governance tools into fast-moving AI pipelines.

Although OpenScale is designed to be model-agnostic and platform-flexible—supporting models hosted on AWS, Azure, or IBM Cloud—widespread adoption remains a work in progress. Integration across diverse AI environments is promising, but implementation challenges and evolving expectations around AI transparency suggest that these tools are still maturing.

Conclusion

Watson OpenScale reflects an effort to bridge the gap between AI capabilities and human oversight, especially in high-stakes or regulated environments. As AI becomes more deeply embedded in strategic and operational decisions, tools that offer visibility into how models make predictions will be increasingly important. However, explainability is still an evolving domain, and tools like OpenScale represent a step—not a solution—in the broader push toward responsible and trustworthy AI. The development and deployment of such tools will need to keep pace with rapid advances in AI while also addressing the real-world challenges of complexity, integration, and user trust.

Discussion Questions

1. Why is transparency important when businesses deploy AI systems based on deep learning?
2. How does IBM Watson OpenScale help businesses build trust with customers, regulators, and other stakeholders?
3. What tradeoffs do companies face between maintaining high model performance and ensuring model explainability?
4. In what ways could a lack of AI explainability create risks for businesses, even if the model appears accurate?
5. How might increasing demand for explainable AI tools influence future business strategies involving AI adoption?

Case References

IBM. (2023). *IBM Watson to watsonx*. **https://www.ibm.com/watson/industries**

IBM. (n.d.). *Watson OpenScale–IBM*. **https://www.ibm.com/docs/en/software-hub/5.1.x?topic=services-watson-openscale**

IBM. (2023, March 29). *What Is Explainable AI?* **https://www.ibm.com/think/topics/explainable-ai**

CASE STUDY 5.2: TESLA AUTOPILOT—THE PROMISE AND PERIL OF DEEP LEARNING–DRIVEN AUTONOMY

The pursuit of fully autonomous vehicles has been one of the most ambitious and transformative goals of modern transportation. Companies worldwide envision a future where human drivers are replaced by AI-powered systems capable of reducing accidents, improving traffic flow, and transforming urban mobility. Among the leading voices in this effort has been Tesla, whose autopilot and full self-driving (FSD) systems represent a bet on the power of deep learning to achieve full autonomy.

The Market for Autonomous Vehicles

The global market for autonomous vehicles has been growing rapidly. Cars in this space are classified by their level of driving autonomy, categorized by the Society of Automotive Engineers (SAE) from Level 0 (no automation) to Level 5 (full automation)[16] (see Table 5.1). A Goldman Sachs 2024 report estimates that 10 percent of new vehicle sales will be cars at Level 3 by 2030 and 30 percent will be at Level 2.[17] Tesla's autopilot system is considered Level 2, offering driver assistance but still requiring full attention.

Table 5.1 **Levels of Driving Autonomy**

Level	Description	Examples
0: No driving automation	Manually controlled, with no automation. Systems like emergency braking may assist but do not qualify as automation.	
1: Driver assistance	Single automated system for driver assistance, such as adaptive cruise control. The driver monitors other aspects of driving.	
2: Partial driving automation	Advanced driver assistance systems (ADAS) control steering and accelerating/decelerating. The driver can take control at any time.	Tesla Autopilot and Cadillac Super Cruise
3: Conditional driving automation	Vehicles have environmental detection capabilities and can make informed decisions but require a human override.	Audi A8L with Traffic Jam Pilot
4: High driving automation	Vehicles can intervene if things go wrong and operate in self-driving mode within limited areas (geofencing).	NAVYA shuttles and Waymo self-driving taxis
5: Full driving automation	Vehicles do not require human attention and can operate anywhere without geofencing. They will not have steering wheels or pedals. Fully autonomous cars are still undergoing testing.	

Deep Learning and Tesla's Autonomy Strategy

At the heart of Tesla's approach is the use of deep learning to process real-time data from cameras. Earlier versions of the system also incorporated radar and ultrasonic sensors, but these have been phased out in favor of a vision-only strategy. Tesla's vehicles act as rolling data collection hubs, sending billions of miles of driving data back to the company's servers to continually train and improve its AI models. By focusing on vision-based deep learning, Tesla is betting that camera systems combined with AI will be sufficient to achieve full autonomy.

In contrast, many of Tesla's competitors rely heavily on lidar (light detection and ranging) systems, which actively map the environment using laser pulses. Lidar is considered to perform better than camera-based systems in low-light or adverse weather conditions, offering more reliable obstacle detection when visibility is poor.

Tesla has also invested heavily in building custom AI hardware, such as its Full Self-Driving (FSD) computer, designed to accelerate deep learning inference in real time. By tightly integrating

hardware and AI software, Tesla aims to optimize performance, reduce latency, and create a scalable platform for future autonomous capabilities.

Challenges in Technology and Regulation

Despite Tesla's technological innovations, achieving safe, fully autonomous driving has proven elusive. Autopilot and FSD Beta have faced delays and repeated timeline revisions. Deep learning models can struggle with edge cases—rare, unusual driving scenarios, which are difficult to predict and train for. In addition, regulators are scrutinizing Tesla's marketing and labeling of its systems. Critics argue that terms like "full self-driving" are misleading, suggesting a level of autonomy the technology does not yet deliver.

Legal Difficulties

Tesla has faced lawsuits alleging that its autopilot system was responsible for fatal accidents. Recalls have been issued to update software safety features. In 2023, Tesla recalled more than 360,000 vehicles equipped with FSD Beta due to concerns that the system could cause crashes. Legal challenges continue to test Tesla's claims and could shape future regulatory frameworks for all autonomous vehicle companies.

Conclusion

Tesla's journey with autopilot and full self-driving highlights both the promise and peril of using deep learning for real-world autonomy. While Tesla has pioneered data-driven AI development at scale, significant technical, legal, and ethical challenges remain. As businesses look to leverage deep learning for transformational innovations, Tesla's experience offers lessons about balancing ambition, safety, transparency, and regulatory engagement.

Discussion Questions

1. How has Tesla's vision-based deep learning strategy differed from the approaches of its competitors?
2. What are some advantages and risks associated with Tesla's reliance on deep learning over lidar technology?
3. How might legal challenges and regulatory scrutiny affect the future development and adoption of Tesla's autonomous driving technologies?
4. What lessons can other businesses learn from Tesla's experience about marketing the capabilities of AI-powered systems?

5. How do technical limitations, such as handling rare driving events (“edge cases”), illustrate the current boundaries of deep learning models?

Case References

Agatie, C. (2023, November 12). “Tesla Autonomous Driving: A History of Promises and Delays.” *AutoEvolution*.

AI Magazine. (n.d.). “Top 10 Companies Developing Autonomous Vehicle Technology.” **https://aimagazine.com/technology/top-10-companies-developing-autonomous-vehicle-technology**

Musk, E. (2016, July 20). “Master Plan, Part Deux.” Tesla. **https://www.tesla.com/blog/master-plan-part-deux**

Synopsys. (n.d.). “Autonomous Driving Levels.” **https://www.synopsys.com/automotive/autonomous-driving-levels.html**

Tesla, Inc. (n.d.). *Model 3 Owner’s Manual*. **https://www.tesla.com/ownersmanual/model3/en_us/GUID-2CB60804-9CEA-4F4B-8B04-09B991368DC5.html**

CASE STUDY 5.3: FACIAL RECOGNITION TECHNOLOGY AND THE DETROIT LIONS

The Detroit Lions, like many sports franchises, are exploring how artificial intelligence can improve the fan experience and create new revenue streams. Ford Field, their home stadium, now uses facial recognition technology to speed up entry and payment processes. The cameras record facial characteristics such as eye distance, nose width, and jawline contour to generate biometric profiles while also estimating attendees' age ranges, gender presentations, emotional expressions, and movement paths throughout Ford Field. This data feeds into advanced analytics platforms, including CrowdIQ, to better understand crowd dynamics, optimize staffing, adjust merchandise and food offerings, and anticipate potential security concerns.

The Detroit Lions are not just an NFL franchise but also a for-profit entity tasked with managing a football team, marketing and revenue generation, event operations, and security. Analytics, particularly AI-enabled technologies, have quickly become an essential part of the Lions' playbook for managing these activities, improving the gameday experience, and making the Lions more profitable.

Technology Overview

One challenge faced by each team in the NFL, including the Detroit Lions, is ensuring that the appropriate people are in the correct places. Ticketholders need to get to their seats, and employees, contractors, and vendors need access to spaces that are relevant to their duties. Facial recognition technology uses AI to scan still or video images for specific facial features or faces. It can then compare the identified faces to a set of faces, such as fans who have been banned from the stadium for unruly behavior. Facial recognition can also be used to identify employees, credentialed media, vendors, and other qualified personnel and allow them to be entered into restricted areas, such as the press box or a locker room. Facial recognition is used in these ways as a form of access control. The technology must have a comparison set of faces to allow specific people to enter specific areas or to keep predetermined people out of the stadium.

Applications and Business Value

CrowdIQ uses similar AI technology but does not use a comparison set of known faces. Its software and cameras can determine the number of people in a particular area, the colors of clothing worn (great for determining how many fans from opposing teams are in attendance), gender, and facial expressions. The Detroit Lions use CrowdIQ to monitor lines at concession stands, assist with crowd control, and even nudge fan behavior.

Long lines at concession stands lead to reduced sales. The information coming from CrowdIQ alerts management when and where they need to increase staffing or open additional lines/concession stands so more fans can be served quickly, leading to increased sales, which are generally highly profitable. However, CrowdIQ can also determine if attendees are fans of the visiting team and are purchasing increasing amounts of alcoholic beverages; staff can direct more security to specific areas and even slow down or halt sales of alcohol in favor of water, soft drinks, and food. CrowdIQ can even tell when fans are getting bored with the game and start playing upbeat music and offer additional promotions and activities for fan engagement.

CrowdIQ has also enabled additional business intelligence. For example, the Lions learned that they had many more female fans at games than previously, allowing them to adjust the product mix at concession stands, bars, and merchandise kiosks to reflect a more diverse set of tastes among game attendees. By offering an attractive mix of food, beverage, and merchandise to this larger-than-initially thought and growing fan segment, the Detroit Lions are winning both on and off the field.

Discussion Questions

1. Are there additional opportunities for the Detroit Lions and other NFL teams to increase their use of facial recognition technologies?
2. The use of facial recognition technologies is considered problematic and even a violation of civil liberties by critics. How might these concerns be addressed? What other problems can you identify?
3. The CrowdIQ system has led the Lions to change their assumptions about gameday attendees and gender, finding that there were far more female fans at games than previously thought. What other assumptions about gameday can an AI system such as CrowdIQ challenge?
4. How could deep learning models improve the accuracy and fairness of facial recognition systems used in stadium environments?
5. What strategies could the Detroit Lions implement to increase fan trust and transparency when using AI-driven technologies like facial recognition?

Case Reference

Lemire, J. (2022). “Detroit Lions’ Analytics War Room Is Helping the Fan Experience with Real-Time Fan Tracking Solution from CrowdIQ,” *Sports Business Journal*.

ENDNOTES

[1] Singh, P., Singh, S., & Pandi-Jain, G. S. (March 15, 2018). "Effective Heart Disease Prediction System Using Data Mining Techniques." *International Journal of Nanomedicine* 13(T-NANO 2014 Abstracts), 121–124. **https://doi.org/10.2147/IJN.S124998**

[2] Sharma, A., Singh, A., Kumar, A., & Elaiyaraja, P. (May 2023). "Heart Disease Predictor." *Journal of Emerging Technologies and Innovative Research* 10(5), 64-70. **https://www.jetir.org/papers/JETIR2305838.pdf**

[3] Pandita, Rahul. "How Netflix & Amazon Use AI to Do Better Customer Segmentation." *Teqfocus Consulting LLC*, July 14, 2024. **https://www.teqfocus.com/blog/how-netflix-amazon-use-ai-to-do-better-customer-segmentation**

[4] Rafalski, Kacper. "AI in Logistics—Top Benefits and Practical Applications." *Netguru*, April 24, 2025. **https://www.netguru.com/blog/ai-in-logistics**

[5] Robin, Meredith. "A Global Dress Distributor Gave AI a Chance. Now It Makes Creative Content Faster, and Overstock Is Down 40%." *Business Insider*, March 7, 2025. https://www.businessinsider.com/wholesale-formal-gown-distributor-using-ai-for-ecommerce-operations

[6] LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep Learning." *Nature* 521(7553), 436–444. **https://doi.org/10.1038/nature14539**

[7] Bommasani, R., Hudson, D. A., Adeli, E., et al. (2021). "On the Opportunities and Risks of Foundation Models." *arXiv preprint*. **https://arxiv.org/abs/2108.07258**

[8] WealthStack Roundup. "AI @ Morgan Stanley Assistant Now Live." https://www.wealthmanagement.com/artificial-intelligence/wealthstack-roundup-ai-morgan-stanley-assistant-now-live

[9] InterVision Systems. "What is Multimodal Gen AI and How Can it Evolve Your Business Strategy?" March 13, 2025. https://intervision.com/blog-what-is-multimodal-gen-ai-and-how-can-it-evolve-your-business-strategy/.

[10] Rubio-Licht, Nat. "Adobe's Multimodal Model Could Make AI Summaries Clearer." *The Daily Upside*, January 25, 2024. https://www.thedailyupside.com/adobe-multimodal-model-ai-summaries-clearer

[11] Brown, T. B., Mann, B., Ryder, N., et al. (2020). "Language Models Are Few-Shot Learners." *Advances in Neural Information Processing Systems,* 33, 1877–1901. **https://arxiv.org/abs/2005.14165**

[12] Yoo, T. K., Choi, J. Y., & Kim, H. K. "Feasibility Study to Improve Deep Learning in OCT Diagnosis of Rare Retinal Diseases with Few-Shot Classification." *Medical & Biological*

Engineering Computing, 59(2), 401–415. doi: 10.1007/s11517-021-02321-1. Epub 2021 Jan 25. PMID: 33492598; PMCID: PMC7829497.

[13] Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). "Green AI." *Communications of the ACM* 63(12), 54–63. **https://doi.org/10.1145/3381831**

[14] Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). "Green AI." *Communications of the ACM* 63(12), 54–63. **https://doi.org/10.1145/3381831**

[15] Henry, Parker. "How Duolingo Uses AI to Create Lessons Faster." Last modified June 22, 2023. https://www.duolingo.com/blog/how-duolingo-uses-ai

[16] Synopsys Editorial Staff. "The 6 Levels of Vehicle Autonomy Explained." February 15, 2025. https://www.synopsys.com/automotive/autonomous-vehicles/6-levels-of-vehicle-autonomy.html

[17] Synopsys Editorial Staff. "The 6 Levels of Vehicle Autonomy Explained." February 15, 2025. https://www.synopsys.com/automotive/autonomous-vehicles/6-levels-of-vehicle-autonomy.html.

CHAPTER 5S
Neural Networks and Deep Learning Concepts Technical Supplement

Understanding how neural networks and deep learning models operate can provide valuable context for business professionals and technical teams seeking to deploy AI more effectively. This supplement offers a deeper look into the structures, learning processes, and key innovations that underlie modern AI systems.

TRADITIONAL MACHINE LEARNING vs. DEEP LEARNING

It's important to note that deep learning is not universally better than traditional machine learning. Many business problems—especially those involving structured data—are best solved with simpler, more interpretable models. The choice depends on the nature of the data, the complexity of the problem, and the need for transparency.

Machine learning and deep learning differ significantly in processing information, building models, and tackling increasingly complex and high-dimensional problems. Traditional machine learning refers to earlier approaches—such as decision trees, logistic regression, and support vector machines—that typically require manual feature engineering, where human experts define the characteristics the model should learn from. These models often perform well on structured data but can struggle with unstructured inputs like images or text.

In contrast, deep learning models use neural networks to discover useful features during training, allowing them to handle much larger and more complex datasets.[1]

Table 5S.1 summarizes key differences between traditional machine learning and deep learning approaches.

Table 5S.1 **Traditional Machine Learning vs. Deep Learning**

Facet	Traditional Machine Learning	Deep Learning
Data	Structured or semistructured	Raw/unstructured
Training	Learns patterns from data through algorithms	Learns patterns directly from raw data using a neural net
Date requirements	Works well with small to medium-sized data sets	Requires very large data sets needed for best performance
Feature extraction/engineering	Manual with human experts needed	Automatic using neural nets
Training time	Faster	Typically slower
Complexity of models	Simpler: decision trees, support vector machines (SVMs), linear regression	Complex: convolutional neural networks (CNNs), recurrent neural networks (RNNs), transformers
Applications	Structured data tasks like fraud detection, customer segmentation, or simple predictive analytics	Unstructured data tasks like image recognition, natural language processing, and autonomous driving
Computational requirements	Can run on standard computers	Needs significant computing power with specialized hardware
Explainability	High	Can be challenging

HOW NEURAL NETWORKS WORK

Neural networks are powerful machine learning models inspired by the human brain's structure. They consist of layers of interconnected processing units called neurons, each designed to detect patterns, adjust to input data, and contribute to a final prediction or decision.[2]

Neural Network Architecture

As noted in the chapter, a basic neural network includes three types of layers (see **Figure 5.1** in the main chapter):

- ***Input layer.*** Receives the raw data—such as customer purchase histories, images, or sensor readings—and prepares it for processing.
- ***Hidden layers (one or more).*** Perform intermediate transformations, learning patterns and relationships within the data. Deep learning models have multiple hidden layers.
- ***Output layer.*** Produces the final output, such as a product recommendation, credit approval decision, or image classification.

Each connection between neurons is assigned a weight, representing its importance. Each neuron also has a bias, allowing the model to shift decision boundaries as needed.

Activation Functions

Neurons act as small decision makers within the network. Each neuron receives inputs, multiplies them by assigned weights, adds a bias, and passes the result through an *activation function*. These weighted signals combine across layers, enabling the network to refine its understanding of the data progressively. Training a neural network involves adjusting these weights and biases to minimize the difference between the model's predictions and the actual outcomes.

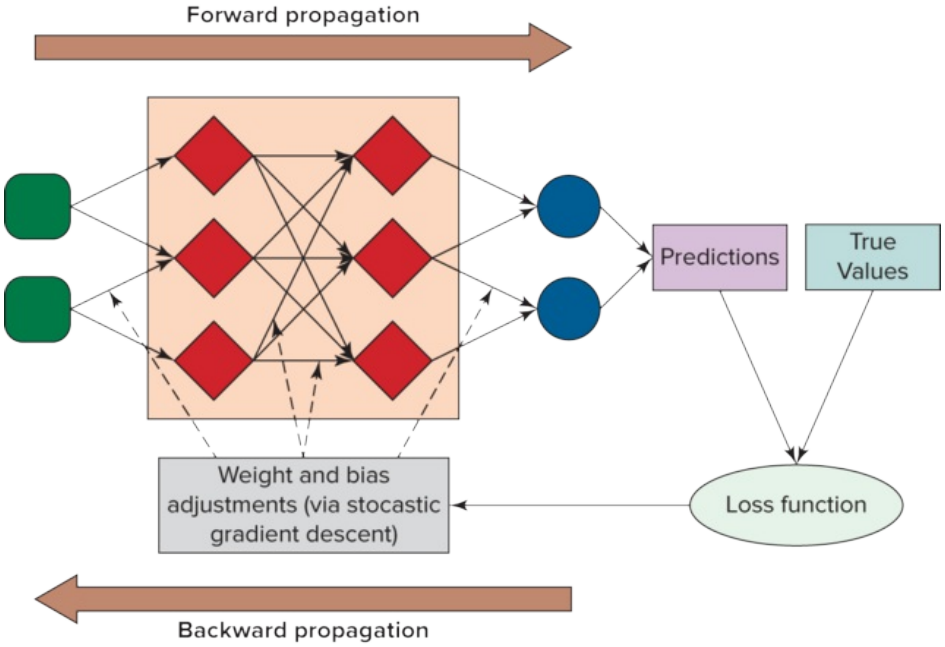
Activation functions give neural networks the ability to learn complex patterns in data. Without them, the network would behave like a simple calculator, only able to draw straight lines and solve fundamental problems. You can think of an activation function like a dimmer switch: Rather than simply turning a neuron on or off, it controls how much signal gets passed to the next layer. This flexibility allows the network to form more nuanced relationships and make better decisions based on subtle patterns. Activation functions make deep learning models robust enough to handle tasks like recognizing faces in photos or interpreting the tone of a sentence.

Forward Propagation and Backpropagation

In *forward propagation*, data flows from the input layer through the hidden layers to the output layer. Each neuron processes its inputs and passes signals forward, with each layer learning to detect more valuable and meaningful features, like edges, shapes, or categories, based on what came before (see **Figure 5S.1**).

FIGURE 5S.1 Forward and Backward Propagation in a Neural Network

In forward propagation, input data flows through the neural network layers to generate a prediction. During training, backward propagation, more commonly known as backpropagation, adjusts the network's internal weights based on the prediction error, helping the model learn from its mistakes and improve over time. During training, stochastic gradient descent (SGD) updates weights using small batches of data to improve learning speed and generalization.



After the network generates a prediction, it evaluates how far its output deviates from the actual result by calculating a *loss function*, a measure of prediction error. This loss function quantifies the difference between the predicted output and the true value, guiding the network on how to adjust its internal parameters.

Using *backpropagation*, the network then moves backward, adjusting its weights and biases to reduce the error. This iterative process—forward prediction followed by backward correction—allows the network to "learn" over time and improve its accuracy across many training examples.[3]

Training with Stochastic Gradient Descent

To efficiently train a neural network, the model uses a method called *stochastic gradient descent (SGD)*. Rather than updating the weights after seeing the entire dataset, SGD updates weights incrementally, using small subsets of the data called *minibatches*. This approach speeds up

learning and introduces randomness that helps the model escape local minima during optimization. Training typically occurs over multiple epochs, where an epoch is one complete pass through the entire training dataset. Each epoch is divided into many minibatches, with the model adjusting its internal parameters after each. Each batch update is referred to as an *iteration*. Repeating this process helps the model gradually improve its performance. This technique is now standard in training deep learning models.[4]

Office Building Analogy

To visualize how a neural network functions, imagine an office building:

- The reception desk (input layer) collects incoming information, such as client forms or inquiries.
- Departments (hidden layers) process the information. Each department specializes in a different aspect—finance, marketing, legal—analyzing and refining the input.
- The executive suite (output layer) makes the final decision based on all the departmental analyses.
- Performance reviews and promotions (backpropagation) happen after each major project. Based on the results, departments adjust their processes, collaborate differently, or receive new resources to improve future outcomes.

This analogy captures how information moves through a neural network and how learning happens by continuously refining internal processes based on feedback.

Overfitting

Like other machine learning models, *overfitting* can occur when a neural network learns the training data too well, including its noise and outliers, leading to poor performance on new, unseen data. It's like memorizing answers for a test rather than understanding the underlying concepts. Techniques such as dropout, early stopping, and regularization are employed to mitigate overfitting, helping the model generalize better to new inputs:

- *Dropout* randomly disables some neurons during training so the network doesn't become too reliant on any one path.
- *Early stopping* monitors performance on validation data and halts training before the model begins to memorize noise.
- *Regularization* adds a small penalty to overly complex models, discouraging extreme weight values and helping the network stay flexible rather than overtuned to the training set.

DEEP LEARNING MODEL TYPES

Not all deep learning models are the same. Different architectures are designed to handle different types of problems. Two of the most important and widely used model types are Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs).

Convolutional Neural Networks: Specialists in Visual Data

Convolutional neural networks (CNNs) are designed to process visual information, such as photographs, videos, medical images, or product scans. Instead of analyzing the entire image at once, CNNs use specialized layers called *convolutional layers* to scan small patches of the image, looking for basic patterns like edges, textures, or colors. The network identifies increasingly complex features, such as shapes, objects, or faces, as the data flows through deeper layers. CNNs are especially effective because they can recognize patterns regardless of their location in the image—an ability known as *translation invariance*.[5]

Key business applications of CNNs include:

- Image classification.
- Medical image analysis.
- Automated defect detection in manufacturing.

While CNNs excel at analyzing spatial relationships in images, they are less effective for tasks involving sequences, such as predicting future events based on past behavior.

Recurrent Neural Networks: Specialists in Sequences and Time Series

Recurrent neural networks (RNNs) are built for tasks involving sequences—where the order and timing of data points matter. In an RNN, each layer's output feeds into the next layer and back into itself. This looping architecture gives RNNs a kind of short-term memory, allowing them to remember earlier steps in a sequence while processing later steps.

RNNs are especially valuable for tasks such as:

- ***Finance.*** Forecasting stock prices, revenue trends, or risk factors based on historical data.
- ***Customer engagement.*** Predicting customer churn based on sequences of interactions.
- ***Language processing.*** Generating coherent sentences word-by-word.

Despite their strengths, traditional RNNs can struggle with long-term dependencies—remembering information far back in a sequence. Newer architectures like *long short-term memory (LSTM)*[6] networks and *gated recurrent units (GRUs)* were developed to address this limitation, enabling deeper memory across longer sequences.

TRANSFORMERS AND ATTENTION MECHANISMS

While deep learning dramatically improved natural language processing, earlier models like RNNs still struggled with long-range dependencies—remembering relationships between words far apart in a sentence or paragraph. In 2017, researchers at Google introduced a breakthrough architecture called the *transformer*, which fundamentally reshaped how AI systems process language and other sequential data.[7]

What Are Transformers?

Transformers are deep learning models built around the idea of attention—the ability to weigh the importance of different parts of the input data, no matter how far apart they are. Unlike RNNs, which process information sequentially, transformers analyze all input parts simultaneously, allowing them to capture complex relationships more efficiently and at a greater scale. This parallel processing ability makes transformers much faster to train and better suited for handling large datasets, which is critical for building sophisticated language models.

Deep learning transformed natural language processing by allowing models to learn the structure and meaning of language directly from large collections of text. Instead of relying on manually programmed rules, these models learned how words relate to one another based on how they appear in real sentences. This enabled AI systems to understand grammar better, detect tone, and generate more natural responses—improvements that laid the foundation for today's advanced language models.[8]

How Attention Mechanisms Work

The core innovation behind transformers is the *attention mechanism*. This architecture improves training efficiency and enables the model to understand relationships between words or elements, even when they are far apart in the input—a key capability for translating long sentences or answering questions based on complex text. For example, when translating a sentence from English to French, the model can pay more attention to specific words that carry the core meaning, even if they appear earlier or later in the sentence.[9]

In technical terms, attention assigns different weights to different input tokens (words, phrases) based on their importance for the current task. This enables the model to understand context flexibly and generate more coherent and accurate outputs.

Transformers in Modern AI Applications

Transformers have become the foundation for many of today's most advanced AI systems, including:[10]

- ***BERT (bidirectional encoder representations from transformers).*** Powers Google search improvements, questions answering systems, and documents summarization tools.
- ***GPT (generative pretrained transformer) models.*** Underpin AI tools capable of generating humanlike text, assisting in marketing, customer service, creative writing, and business communication.

Transformer models are widely used for tasks such as text generation, document summarization, machine translation, and conversational AI.

Conclusion

Neural networks and deep learning models form the technical backbone of many modern AI systems. A basic understanding of these architectures and learning processes provides a foundation for further exploration of AI capabilities and future innovations.[11]

ENDNOTES

[1] LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep Learning." *Nature* 521(7553), 436–444. **https://doi.org/10.1038/nature14539**

[2] McCulloch, Warren S., & Pitts, Walter. (1943). "A Logical Calculus of the Ideas Immanent in Nervous Activity." *The Bulletin of Mathematical Biophysics* 5(4), 115–133. **https://doi.org/10.1007/BF02478259**

[3] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). "Learning Representations by Back-Propagating Errors." *Nature* 323(6088), 533–536. **https://doi.org/10.1038/323533a0**

[4] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. Available online: **https://www.deeplearningbook.org/**

[5] LeCun, Y., Lon Bottou, L., Bengio, Y., & Haffner, P. (1998). "Gradient-Based Learning Applied to Document Recognition." *Proceedings of the IEEE* 86(11), 2278–2324. **https://doi.org/10.1109/5.726791**

[6] Hochreiter, Sepp, & Schmidhuber, Jürgen. (1997). "Long Short-Term Memory." *Neural Computation* 9(8), 1735–1780. **https://doi.org/10.1162/neco.1997.9.8.1735**

[7] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Aidan N. Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). "Attention Is All You Need." *Advances in Neural Information Processing Systems* 30(NeurIPS 2017). **https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf**

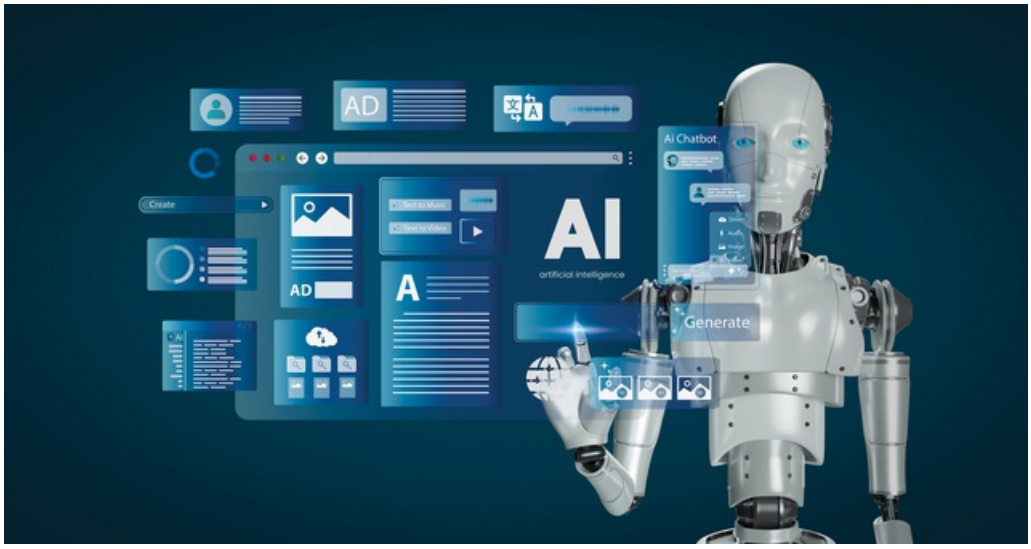
[8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Aidan N. Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). "Attention Is All You Need." *Advances in Neural Information Processing Systems* 30(NeurIPS 2017). **https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf**

[9] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Aidan N. Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). "Attention Is All You Need." *Advances in Neural Information Processing Systems* 30(NeurIPS 2017). **https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf**

[10] Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). "Recent Trends in Deep Learning Based Natural Language Processing." *IEEE Computational Intelligence Magazine* 13(3), 55–75. **https://doi.org/10.1109/MCI.2018.2840738**

[11] Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). "Recent Trends in Deep Learning Based Natural Language Processing." *IEEE Computational Intelligence Magazine* 13(3), 55–75. **https://doi.org/10.1109/MCI.2018.2840738**

CHAPTER 6
Generative AI in Business: Transforming Industries with Innovative Technology



BOY ANTHONY/Shutterstock

Learning Objectives

- **LO6-1** Define generative AI, and distinguish it from predictive or task-specific AI.
- **LO6-2** Describe how generative AI models are trained and how they produce content.
- **LO6-3** Identify and evaluate key business applications of generative AI.
- **LO6-4** Explain the risks associated with generative AI, including hallucinations, bias, and intellectual property concerns.
- **LO6-5** Describe the business models used to monetize generative AI, and explain how organizations capture value from these technologies.

INTRODUCTION

Until recently, most business applications of artificial intelligence focused on analyzing data to classify customers, detect fraud, or predict future trends. These predictive AI tools have become essential for improving efficiency and decision making. But a new generation of AI tools is shifting the conversation from *What will happen?* to *What could I create?*

Generative AI refers to systems that can produce original content, such as text, images, audio, and even product designs, based on patterns learned from data. These models don't just analyze, they generate. Businesses are now using generative AI to write marketing copy, design prototypes, create virtual assistants, and much more. In a study of nearly 1,000 occupations, researchers found that for 80 percent of U.S. workers, at least 10 percent of the tasks they do could be done by a generative AI.[1]

As seen in **Chapter 5**, the exponential growth in neural network parameters has fueled the remarkable advances in generative AI. These ever-larger models are now capable of producing humanlike text, imagery, and even video with increasing accuracy and realism.

This chapter introduces the core ideas behind generative AI and explores how it is being applied in real-world business contexts. You will learn how these systems generate content, how they differ from predictive models, and where they are already delivering value—or causing concern. This chapter takes a relatively nontechnical approach to covering generative AI. The **Technical Supplement** at the end of the chapter presents a deeper dive into the algorithms that power generative AI.

While this chapter highlights the business potential of generative AI, it is essential to also recognize its limitations. Generative models can confidently produce biased, misleading, or factually incorrect content. These issues are discussed later in the chapter and are a central focus of **Chapter 8**, which explores how prompt engineering techniques can help reduce risk and improve the quality of AI-generated content. **A Technical Supplement at the end of the chapter p rovides more detailed explanations of algorithms used for generative models for interested readers.**

WHAT IS GENERATIVE AI? (LO6-1)

Generative AI is a branch of artificial intelligence focused on producing original content such as text, images, audio, video, or designs based on patterns learned from vast amounts of data. Unlike predictive or task-specific AI models, which are typically used to classify, forecast, or detect patterns in existing data, generative AI creates something new in response to a user's prompt.

At the core of generative AI are advanced deep learning models capable of generating realistic outputs that resemble the data they've been trained on. These models don't simply copy; they generate original content informed by patterns they have learned. For example, some models can compose humanlike text from a short prompt, while others can produce realistic images or videos. Tools like *ChatGPT* (text), *DALL-E* (images), and *Synthesia* (video) demonstrate how these capabilities are being applied in business. Companies are using them to design new products, personalize marketing messages, improve customer service, and explore novel ways of interacting with customers. These distinctions are beginning to blur. For instance, while DALL-E originally operated as a standalone diffusion model, newer systems like GPT-4o integrate text, image, and audio generation in a single model—enabling more seamless and context-aware outputs across media types.

While tools like ChatGPT and DALL-E may feel like overnight sensations, the technologies behind them have been in development for decades. **Table 6.1** provides a brief timeline of major milestones in generative AI.

Table 6.1 **Milestones in the Development of Generative AI**

Year	Milestone
1956	Term "artificial intelligence" coined at the Dartmouth Conference.
2014	Generative adversarial networks (GANs) introduced—early breakthrough in image generation.
2017	Transformers architecture introduced—foundation for models like GPT and BERT.
2020	OpenAI releases GPT-3, a large-scale language model that gains widespread attention.
2022	DALL-E 2 and Stable Diffusion popularize generative image creation.
2022–2023	ChatGPT launches and quickly becomes a public-facing generative AI milestone.
2024	GPT-4o integrates text, image, and audio generation in a unified multimodal model.

Predictive vs. Generative AI

These milestones mark the evolution of generative AI. Still, they build on decades of progress in what is now often called **predictive AI**, systems designed to classify, forecast, or detect patterns in data. As generative models have captured recent headlines, it is helpful to distinguish them from the predictive systems that form the backbone of most traditional business AI applications. Understanding the difference between these approaches helps clarify how generative AI expands the landscape of possibilities.

Predictive AI typically starts with structured input data, such as spreadsheets, user profiles, transactions, time series trends, images, or medical records, and uses a pretrained model to deliver specific outputs like classifications, risk forecasts, or anomaly alerts. These systems are frequently used in business to automate decision making, flag risks, or optimize operations. Other predictive outputs, such as recommendations, risk scores, or diagnostic predictions, are also common but are not shown here to maintain focus. The term **traditional AI** is commonly used in popular discussions to refer to predictive systems that have been used for decades to solve narrow, well-defined tasks.

In contrast, generative AI begins with a user-provided prompt (e.g., “Write a product description” or “Create an image of an intern”) and uses a deep learning model trained on massive datasets to generate entirely new content. Outputs may include paragraphs of text, synthetic images, or AI-generated video avatars. Generative AI tools expand creative and communication possibilities for businesses.

Figure 6.1 illustrates the conceptual difference between predictive and generative AI systems. **Table 6.2** compares predictive (or task-specific) AI with generative AI, highlighting differences in purpose, underlying methods, outputs, and user interaction.

FIGURE 6.1 Predictive AI vs. Generative AI Workflows

Predictive AI uses structured input data and rule-based or statistical models to produce specific outputs such as classifications, forecasts, or alerts. In contrast, generative AI responds to user prompts using deep learning models to generate new content, such as text, images, or videos.

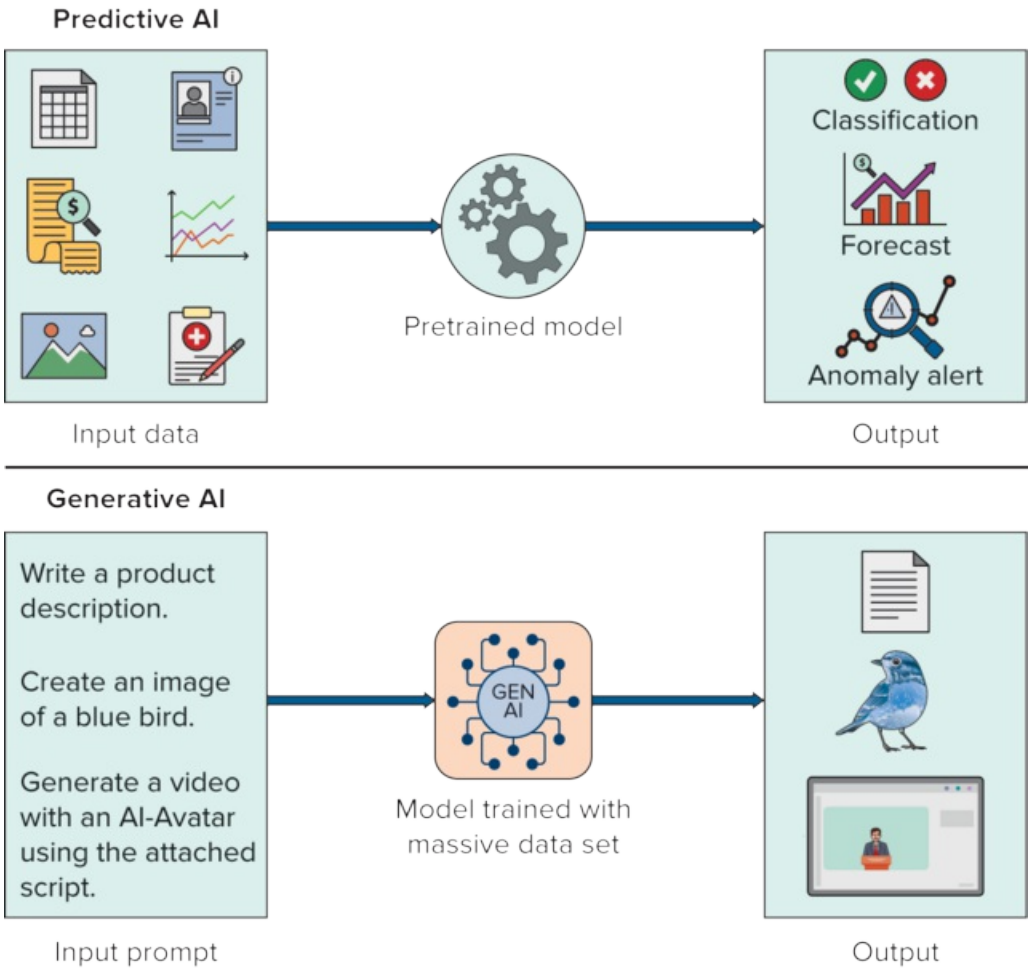

Predictive AI
Classification
Forecast
Anomaly alert
Pretrained model
Input data
Output
Generative AI
Write a product description.
Create an image of a blue bird.
Generate a video with an AI-Avatar using the attached script.
GEN AI
Model trained with massive data set
Input prompt
Output

Table 6.2 **Comparing Predictive and Generative AI**

Feature	Predictive/Task-Specific AI	Generative AI
Primary function	Classify, predict, or optimize outcomes based on input data	Generate new content in response to input prompts
Common techniques	Rules-based systems, decision trees, logistic regression, discriminative neural networks	Transformers, GANs, variational autoencoders (VAEs)
Underlying methods	May use rules or traditional machine learning	Relies heavily on deep learning architectures
Input/output flow	Input → Label or score	Prompt → Original content (text, image, video, etc.)
Example uses	Fraud detection, medical diagnosis, recommendation engines	Text generation, image synthesis, avatar video creation
Output type	A label, score, or category	Human-readable or perceptual content (text, image, audio)
Data requirements	Often trained on labeled data	Typically trained on large, unlabeled datasets
Transparency	Often more interpretable and auditable	Can be difficult to explain ("black box")
User interaction	Often fixed-task and automated	Interactive; responds to varying user prompts
Reliability and risk	Predictable, easier to validate	Can hallucinate or reflect training bias

Types of Generative AI

Generative AI encompasses a range of applications, with different types of models tailored for specific tasks. There are three primary types of generative AI models: text-based, image-based, and multimodal:

- **Text-based generative AI** models, such as OpenAI's *GPT*, have gained widespread recognition for generating humanlike text. These models can be used for automated content creation, chatbots, language translation, and sentiment analysis. In business analytics, they are crucial in generating reports, summarizing data, and even providing textual insights into data patterns. 618Media, a boutique digital marketing agency, illustrates how even small companies can use text-based generative AI effectively. Integrating Claude AI into its daily operations allows the firm to create content for blogs, newsletters, and ad copy faster. Claude also assists in market research, competitor analysis, and social media analysis.[2]

- **Image-based generative AI** leverages neural networks, such as convolutional neural networks (CNNs), to generate images and manipulate visual data. This type of generative AI is used in creative applications, such as generating art, modifying photos, and creating realistic computer-generated imagery (CGI). Business analytics can be employed for data visualization and image-based insights.
- **Multimodal generative AI** combines text and image generation, enabling the creation of content that seamlessly integrates both forms. It can be used for tasks like generating multimedia presentations, creating interactive data dashboards, and developing content-rich reports that combine visual elements with textual insights.

HOW GENERATIVE AI WORKS (LO6-2)

Generative AI systems are built on advanced neural network architectures that can learn patterns from vast datasets and use that knowledge to create new content. These models aren't just memorizing—they learn the structure and context of language, images, or code well enough to generate original outputs.

Training Generative Models

Training a generative AI model typically involves two major stages: *pretraining* and *fine-tuning*. During **pretraining**, the model is exposed to a massive volume of text or image data—often drawn from books, websites, and other open sources—to learn general patterns and relationships. Well-known generative AI tools such as *ChatGPT*, *Claude*, and *Gemini* are examples of pretrained **foundation models** that can be used across a wide range of topics. The letters GPT stand for *generative pretrained transformer*.[3]

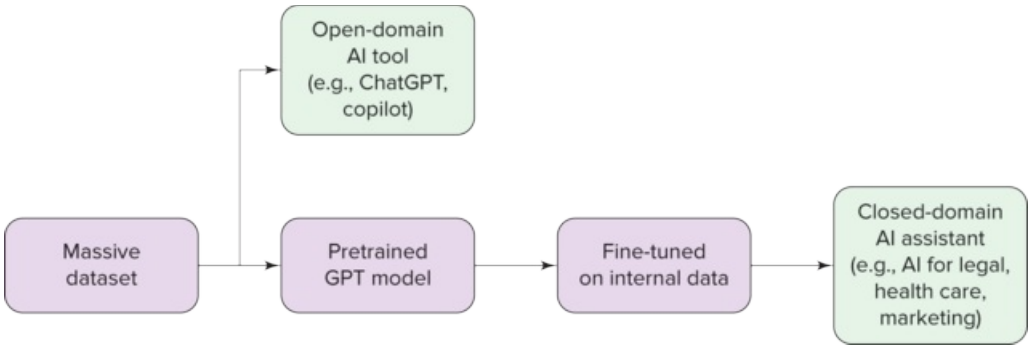
In business, these pretrained models are often fine-tuned on a smaller, more specific dataset to tailor it to particular tasks or industries. For example, a company might adapt a pretrained model to assist with legal writing, respond to health care queries, or reflect a brand's specific voice in marketing content. This **fine-tuning** process produces what many businesses call a **custom model**, one that is adapted to their domain, voice, or proprietary knowledge.

Most custom models are produced in this fashion, but, in rare cases, a company may train a model entirely from scratch using its own data. This requires significant technical expertise and computing power.

Figure 6.2 illustrates how a generative AI model can follow one of two paths: remaining open domain for general use or being fine-tuned to serve a specialized business function.

FIGURE 6.2 From Pretraining to Business Application in Generative AI

This figure illustrates how a generative pretrained transformer (GPT) model is first trained on a massive, general-purpose dataset. From there, it can either be used directly as an open-domain tool (such as ChatGPT or Copilot) or fine-tuned on internal data to serve specialized business needs in fields like law, health care, or marketing.



Text Generation: Step by Step

When a user enters a prompt into a generative AI system, such as "Write a product description for a new smartwatch," the model begins a process of breaking that prompt down, interpreting its meaning, and generating a response, one word at a time.

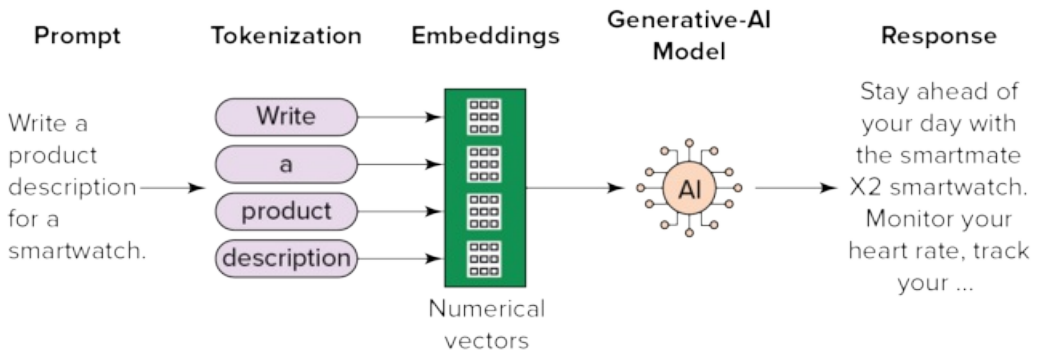
The process starts with **tokenization**, where the input text is split into smaller units called *tokens*. These could be words, parts of words, or characters, depending on the model. The tokens are then converted into *embeddings*, mathematical vectors that represent the meaning and relationships between words based on what the model learned during training. A **mathematical vector** is a list of numbers that captures information in a form the AI model can understand and process.

These embeddings are passed into the model, which uses them to predict the most likely next token. This step is repeated over and over, with the model using its own previous output to predict the next part of the response. This is called **autoregressive generation**. The result is a sequence of words or sentences that appears to flow naturally, even though it is being constructed one piece at a time.

Figure 6.3 illustrates this process, showing how the input prompt flows through the model to produce a final response.

FIGURE 6.3 How a Generative AI Model Produces a Text Response

This figure illustrates the process a generative AI model follows to produce a response to a user prompt. The input is first broken into tokens–smaller units of text–each of which is converted into a numerical vector (embedding). These vectors are passed into the model, which generates the output one piece at a time based on learned patterns and context.



Why Responses Can Vary

One of the most interesting—and sometimes confusing—features of generative AI is that it can produce different responses to the same prompt. This happens because the model doesn't retrieve a prewritten answer. Instead, it generates each response word by word, choosing the next word based on probabilities learned during training. For any given prompt, there are many plausible continuations, and the model selects among them using a sampling strategy. The same model, given the prompt, "How are you," might respond with "I'm fine, thank you. How about you?" in one instance and "Doing well, thanks! And you?" in another.

Image Generation

As noted, generative AI can also generate original images from text prompts. These tools interpret written descriptions and construct images that match them with varying levels of detail and realism. For example, a prompt like "Create the image of a modern smartwatch on a marble countertop under soft lighting" can yield a photorealistic image.

Two of the most common model types used for image generation are **generative adversarial networks (GANs)** and **diffusion models**. GANs work through a creative feedback loop between two models: one that generates an image and another that critiques it until the result is convincing. Diffusion models, used in tools like *DALL-E 2*, *Midjourney*, and *Stable Diffusion*, take a different approach. They start with a noisy, random image and gradually refine it by reversing the process that added the noise in the first place. (The **Technical Supplement** at the end of this chapter presents a deeper explanation of these model types.)

Businesses are adopting image-generation tools in fields such as marketing, product development, and content creation. These models allow teams to visualize ideas quickly, generate product mockups or social media visuals, and even prototype design concepts without the time or cost

typically required for photography or manual design work.

Multimodal and Video Generation

As noted earlier, some of the newest generative AI tools go beyond a single type of content. These are known as **multimodal models**, systems that can work with more than one kind of input or output. For example, a multimodal model might accept a combination of text and images as input and generate a response in either format. Tools like *GPT-4 with vision*, *Gemini*, and *Claude 3* are capable of reading an image, interpreting its contents, and responding to questions about it in natural language. Other tools allow users to describe an image and then generate a detailed visual or even request a voiceover script to be turned into an animated video.

Generative AI is also being used to create videos using tools like *Synthesia*, which allows companies to generate professional-looking videos with AI avatars reading user-provided scripts. These tools combine text, voice synthesis, and facial animation to produce realistic video content without the need for filming. Other models, like *Runway* and *Pika Labs*, focus on generating short, animated video clips from text prompts or image inputs.

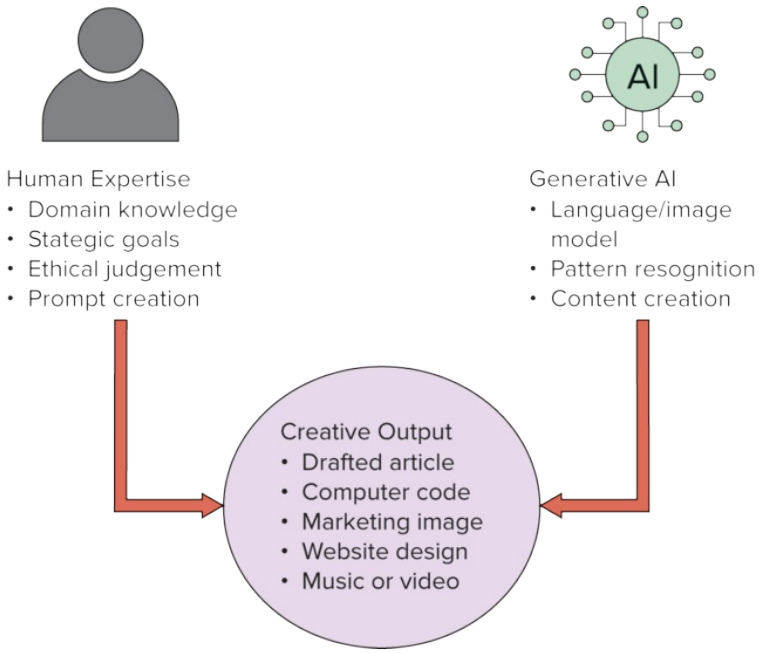
BUSINESS APPLICATIONS OF GENERATIVE AI (LO6-3)

Generative AI is a technological advancement that opens a wide range of strategic opportunities for companies. Businesses leverage its potential to generate novel solutions and products, automate processes, and personalize customer experiences.

Generative AI models represent a paradigm shift in the digital landscape, transitioning from static repositories of information to dynamic creators. With the capability to write articles, generate images, compose music, and even design websites, these models have opened a new realm of creative possibilities. However, it is crucial to recognize that these models do not operate in isolation as autonomous creators (see **Figure 6.4**). Their true potential is realized when combined with the expertise and direction of knowledgeable individuals.

FIGURE 6.4 Generative AI as a Creative Partner, Not a Standalone Creator

This figure illustrates how generative AI works best when guided by human insight. The outputs, such as text, images, or product designs, are the result of a collaborative process where humans provide context, direction, and oversight, while the AI generates content based on learned patterns.



Text-Based Applications

One of the most common uses of generative AI in business is for generating written content. These tools can draft emails, write reports, summarize documents, and even tailor messages for different audiences. In marketing, generative AI can create social media posts, product descriptions, and personalized advertising copy. In internal communications, companies are using generative models to draft announcements, summarize meeting notes, and respond to customer inquiries at scale.

Generative AI tools are also being used to assist with strategic business communication. Language models can help teams craft responses that are more concise, customer friendly, or aligned with brand voice. These tools can be particularly helpful for employees who may not have strong writing skills or for teams working under tight deadlines. The ability to generate multiple versions of a message with slight variations makes generative AI useful for A/B testing and other data-driven communication strategies.

Example: Using Generative AI to Improve Marketing Effectiveness

A/B testing, also known as split testing, compares two versions of content—like an email subject line—to see which performs better. For example, a team might test "Unlock Our Secret Tips for Success!" against "Exclusive Success Strategies Inside!" by sending each to a small, random sample of subscribers. The subject line with the higher open rate is then used for the rest of the audience.

Generative AI tools make this process faster and more effective by automatically generating multiple message variants tailored to tone, audience, and intent. Instead of brainstorming manually, marketers can prompt the AI to produce a range of compelling options in seconds. Platforms like *SendX* now integrate generative AI to help teams set up A/B tests, track real-time performance, and even automate optimization, replacing guesswork with rapid, data-driven decision making.[4]

Custom GPTs

As noted earlier, some businesses and individuals are now creating custom GPTs—specialized versions of large language models tailored to specific tasks. For example, OpenAI's *GPT Store* features tools like a Creative Writing Coach for helping writers brainstorm ideas, a Laundry Buddy for explaining fabric care instructions, and Game Time, which generates quick party games based on group size and themes.[5] These micro-tools demonstrate how generative AI can be embedded into daily business functions or consumer experiences with minimal coding.

CMY Cubes, an Australian STEAM-focused company, specializes in visually striking, sensory-based educational tools. Their flagship product, a colorful, precision-engineered cube, demonstrates additive color mixing and is used by educators, parents, and designers worldwide. With a small team and a growing global presence, CMY Cubes needed a way to scale marketing and customer engagement without hiring a large content team. In response, they built a custom

GPT trained on their product catalog, brand tone, and internal communications. The model now helps generate search engine optimized blog posts, draft customer service replies, and analyze marketing campaign results. They maintain consistency across all communications by embedding brand language and tone directly into the GPT.[6]

Importantly, CMY Cubes follows a *human-in-the-loop approach*. Staff members review all AI-generated content before publishing or sending, ensuring accuracy and brand alignment. In this fashion, they increased content output, had more responsive marketing, ensured the veracity of AI content, and used the saved time for creative and strategic tasks.

Chatbots and Customer Engagement

The rise of chatbots in the business world has been remarkable, reflecting a significant shift in how companies engage with customers and manage operations. Powered by advancements in AI, particularly in natural language processing (NLP) and generative AI, today's chatbots have evolved far beyond simple scripted responders. They can now handle complex customer interactions, provide personalized support, and even conduct transactions. This proliferation of chatbots is driven by their ability to offer 24/7 service, reduce operational costs, and enhance customer experience.

Chatbots are employed across various sectors, from retail for personalized shopping experiences to banking for automated customer service and financial advice. Several companies have successfully integrated AI chatbots to enhance their services:

- Lufthansa Group deployed an AI chatbot within its HR department to assist employees with questions about policies, benefits, and onboarding.[7]
- Elomia offers a mental health chatbot that mimics supportive therapy conversations, providing accessible mental health support.[8]
- Humanly is a recruiting platform that leverages generative chat to screen job candidates and share company culture insights, streamlining the hiring process.[9]
- Duolingo uses generative AI to create realistic language practice scenarios, offering personalized feedback to learners.[10]

Beyond these examples, industry wide adoption of AI chatbots is growing rapidly. In early 2025, 16 percent of companies had already adopted AI chatbots, with an additional 55 percent planning to implement them.[11] With 90 percent of customer queries resolved in fewer than 11 prompts, issues are getting resolved in minutes, setting a new standard in customer service.[12] Over time, they learn from each interaction, refining their ability to handle complex queries and provide relevant solutions.



Chatbots are employed across various sectors.
ImageFlow/Shutterstock

Generative Visual Content and Virtual Try-On

Generative AI is also transforming how businesses create visual content. Instead of relying on stock photos or traditional design processes, companies can now use image-generation tools to produce original visuals in seconds. By entering a simple text prompt, users can generate high-quality marketing images, product mockups or images for digital campaigns.

Tools like *Dream Studio*, *DALL-E*, and *Midjourney* are driving this shift:

- *Dream Studio* transforms photos into surreal or abstract artworks by enhancing patterns and features. Users can upload images and choose from artistic styles guided by sample prompts.
- *OpenAI's DALL-E* creates unique and imaginative images directly from text descriptions, useful for concept development and creative ideation.
- *Midjourney* excels at producing high-quality visual outputs, from photorealistic scenes to stylized illustrations, making it popular with marketing teams and designers.

These tools are especially valuable for organizations with limited design resources or fast-moving content cycles. Marketing teams use them to generate ads, websites, and social media visuals. Product teams employ them to explore early design concepts or test visual variations in A/B campaigns. The ability to produce custom visuals at scale has become a key competitive advantage.

Sephora, a leading beauty retailer, uses generative AI to power its virtual try-on experience. This feature allows customers to see how makeup products like lipstick, eyeshadow, or foundation would look on their own face using their phone or computer camera. The model generates realistic renderings of how different products appear under various lighting conditions and skin tones. This enhances the online shopping experience by reducing uncertainty and encouraging product exploration.[13] The ability to try on products virtually is becoming common with many

online retailers.

Product Design

Airbus uses generative AI in its design process for both parts and production.[14] This approach uses AI algorithms to explore many design possibilities, focusing on optimizing for specific criteria like weight reduction, strength, and aerodynamic efficiency. The AI system can generate and evaluate thousands of potential designs, identifying the best performance according to the defined parameters. For example, Airbus used generative AI to design a cabin partition for its A320 fleet.[15] It allowed the company to get a part with the desired parameters and explore new manufacturing options that increased production efficiency. This method allows Airbus to innovate more rapidly and efficiently, leading to lighter, more fuel-efficient aircraft designs that would be challenging to conceive using traditional methods.

BMW also uses generative AI in its design process, particularly for components like custom wheel rims.[16] Designers input performance and aesthetic requirements into the system, and the AI generates multiple design alternatives that balance weight, material constraints, and brand style. These AI-generated design candidates can then be tested digitally before moving to production. Like Airbus's use of generative design, this application highlights how AI can augment creativity and engineering precision in product development.



BMW uses generative AI in its design process, particularly for components like custom wheel rims.
photootohp/123RF

Generative AI in Contract Review

JPMorgan Chase's implementation of generative AI in its COIN (contract intelligence) platform

highlights the power of AI in financial services.[1] COIN uses advanced models to analyze complex legal documents, such as commercial loan agreements, which traditionally require approximately 360,000 hours of human work annually. By automating this process, the platform significantly reduces the time involved—completing tasks in seconds that used to take hours—and minimizes errors inherent in manual review. This efficiency gain allows employees to redirect their focus to more strategic tasks, enhancing overall productivity and operational effectiveness within the bank.

Training and Onboarding

Generative AI is also being used to streamline corporate training and onboarding. Video creation tools like *Synthesia* allow organizations to generate realistic training videos using AI avatars that speak user-written scripts in multiple languages. These tools eliminate the need for professional actors, filming, and editing, enabling rapid content creation at a fraction of the traditional cost. For global organizations, it allows them to create and communicate a consistent message to employees in their primary language.

Platforms like *InVideo* offer AI-assisted editing and video generation tools that let users create onboarding videos, explainers, or short social media clips from simple text prompts. These tools use generative AI to suggest visual layouts, add transitions, and synchronize narration with visuals. Companies can scale training programs and personalize onboarding experiences across departments, geographies, and languages.

Multimodal and Interactive Experiences

As noted earlier, some of the most advanced generative AI tools combine multiple types of input and output, creating multimodal experiences. Businesses use these tools to analyze charts, describe product photos, or generate creative content blending language and imagery. In customer experience, multimodal AI powers interactive shopping assistants and multimodal design brainstorming sessions. A variety of generative AIs produce videos. Some of the more prominent are *InVideo*, *Synthesia*, and *DeepBrain*. Common features include text-to-speech, avatar choice, and animation.

Specialized Use Cases across Industries

Beyond content creation and customer engagement, generative AI is being adapted for specialized applications in fields like law, health care, and software development. In legal services, tools like *Harvey* use fine-tuned language models to help draft contracts and summarize case law. In health care, AI assistants help generate patient education materials and explain complex clinical information. In software, tools like *GitHub Copilot* assist programmers by suggesting code and

documentation. These examples show how generative AI can be tailored to serve industry-specific needs by combining general language capabilities with business rules, regulatory standards, and proprietary data.

RISKS AND LIMITATIONS OF GENERATIVE AI (LO6-4)



While generative AI offers substantial business value, it also introduces new challenges and risks. Companies must be aware of the limitations of these tools, particularly when it comes to trust, accuracy, ethics, and compliance. Because generative models are trained on large and sometimes unverified datasets, they may produce outputs that are incorrect, biased, or inappropriate.

Hallucinations and Factual Inaccuracy

One of the most prominent limitations of generative AI is the phenomenon known as **hallucination**—when a model generates content that sounds plausible but is factually incorrect or entirely fabricated. For example, a language model might produce a confident answer to a legal or medical question that includes false citations or nonexistent data. This poses significant risks in business settings where decisions may be based on AI-generated outputs.

Hallucinations are particularly common in open-domain AI models—systems designed to respond to a wide range of topics using probabilistic reasoning. Because these models aim to produce the most likely next word or phrase, they can create text that seems reasonable even when it is ungrounded or inaccurate. In contrast, closed-domain AI systems, which are fine-tuned for specific company knowledge and typically include safeguards against answering out-of-scope questions, tend to reduce the risk of hallucination. These safeguards might include refusing to answer unsupported queries or defaulting to verified knowledge bases.

The risk of hallucinations has also been mitigated by the use of **retrieval-augmented generation (RAG)** techniques. RAG models access up-to-date information—such as pulling from trusted websites or internal databases—to ground responses in current and verifiable sources. This is particularly important for time-sensitive or domain-specific queries where static training data may be outdated. While hallucinations remain a concern, these strategies are helping businesses manage the risk more effectively.

It is essential to remember that all generative AI outputs should be verified, especially in high-stakes or regulated environments. As OpenAI warns on its ChatGPT platform: “ChatGPT can make mistakes. Check important info.” Treating generative AI as a drafting partner, not a definitive source, helps avoid the risks of acting on incorrect or misleading information. (See **Chapter 8** for strategies on prompt engineering and verifying AI-generated content.)

Bias and Fairness

As we have discussed elsewhere in this book, generative AI models can reflect or even amplify biases present in the data they were trained on. This can result in outputs that reinforce stereotypes or exclude certain groups. For example, image generators might default to certain demographics unless explicitly prompted otherwise, or language models might reflect cultural or gender bias. These outcomes stem from the statistical nature of how models learn from patterns in training data. Addressing bias requires not only more inclusive training data, but also evaluation, auditing, and ongoing adjustment.[18]

Some companies have explored using generative AI to make chatbot interactions more inclusive. Instead of assuming a binary gender identity in conversation, AI systems can be designed to ask for or infer user preferences respectfully, adjusting names, pronouns, and tone accordingly. While generative models can unintentionally default to biased language, companies are beginning to train and fine-tune bots with datasets that reflect a broader range of identities and social norms.[19] (For a broader discussion of ethical AI design and bias mitigation strategies, see **Chapter 7**.)

Intellectual Property and Legal Concerns

Because generative models are trained on publicly available content, questions have emerged about ownership, copyright, and legal liability. For instance, if a model generates text or images that closely resemble copyrighted works, it may raise intellectual property concerns. Companies need to consider the legal implications of using generative content for public-facing materials, especially when generated by third-party tools. This is an evolving area, and regulations may vary across jurisdictions. (See **Chapter 3** for related issues in data governance and compliance.)

Transparency and Explainability

Unlike rule-based systems, generative models often operate as “black boxes”—they can produce useful results, but it may be unclear how they arrived at those results. This lack of transparency makes it difficult to audit decisions, identify errors, or explain outcomes to stakeholders. In regulated industries like finance or health care, this may limit how and where these models can be used. (**Chapter 7** explores how businesses evaluate and adopt AI tools, including considerations for explainability.)

Ethical Use and Governance

The power of generative AI also raises broader ethical questions. How do organizations ensure that these tools are used responsibly? Who is accountable for harm caused by AI-generated

misinformation or offensive content? Responsible governance of generative AI includes clear use policies, disclosure when content is AI-generated, and thoughtful consideration of its societal impact. Ethics review boards, audit trails, and transparent deployment strategies can help businesses reduce risk and build public trust. (For more on organizational policies and risk management, see Chapter 7.)

Work Role Changes

Generative AI is reshaping not just what work gets done but who does it. While broader workforce implications, such as job displacement or reskilling, are discussed elsewhere in this book, it is also important to consider how specific roles evolve within organizations. In many cases, generative AI doesn't eliminate jobs outright but reshapes tasks, leading to hybrid models where human workers and AI systems collaborate. The shift is especially visible in customer-facing roles, where AI handles routine interactions and humans take on more specialized or oversight functions.

Klarna, a Swedish financial technology company, is renowned for its "buy now, pay later" program, which allows online shoppers to split large purchases into multiple payments. Popular among Generation Z and TikTok users, Klarna has recently faced scrutiny for its shift from human employees to AI, particularly in customer service. The company's AI assistant now handles the equivalent of 800 full-time roles. The company is piloting a new employment model for remote customer service workers, offering flexible schedules and competitive pay to complement its AI setup rather than replace it.[20]

BUSINESS MODELS IN GENERATIVE AI
(LO6-5)

Developing and training generative AI systems, especially large models like GPT-4, is expensive. It requires significant computing power, data, and engineering talent. To make these efforts sustainable, companies that build or distribute generative AI must find ways to recover their costs and generate revenue. In other words, they need a business model.

Not every company uses generative AI in the same way. Some create powerful models and give other developers access to them, typically through an **API**, or **Application Programming Interface**. An API lets another program "talk to" the model and use it for tasks like writing, summarizing, or translating. Other companies use these same models to build finished tools, like writing assistants or design generators, that they sell directly to businesses or consumers.

Still others focus on specific industries, such as legal services or health care, where the models are fine-tuned to comply with rules or vocabulary unique to that field. Some companies don't even build the models; they sell the computing infrastructure and software tools needed to host and run them. And finally, a growing number of companies offer their models as open source, meaning anyone can download and use them for free, but they make money by offering paid support, hosting, or customization services.

Table 6.3 summarizes five common business models that have emerged in the generative AI space. Understanding these business models is useful when evaluating how a generative AI company creates value.

Table 6.3 **Business Models in the Generative AI Landscape.**

Model Type	Description	Examples
Customizable foundation models (GPTs)	These are powerful models trained on vast data sets, made available to other developers through APIs. An API lets you connect to the model to perform tasks like summarizing, translating, or generating content.	OpenAI (GPT-4),[21] Claude (Anthropic),[22] Gemini (Google),[23] Cohere[24]
Productized AI tools	These are ready-to-use tools built on top of generative models. Users typically pay a monthly fee to use them for specific tasks like writing, editing, or summarizing.	Jasper,[25] GrammarlyGO,[26] Copy.ai[27]
Vertical-specific solutions	These models are tailored for a particular industry, like legal, health care, or coding. They often include language or compliance rules specific to that domain.	Harvey (legal),[28] Nabla (healthcare),[29] Replit (coding)[30]
AI infrastructure-as-a-service	These companies provide the cloud tools and computing resources needed to host, train, or deploy AI models—mainly for data science or engineering teams.	AWS Bedrock,[31] Azure AI Studio,[32] NVIDIA DGX[33]
Open-source + service model	These companies release their models for free public use, then make money by offering paid services like hosting, support, or enterprise customization.	Hugging Face,[34] Mistral,[35] Stability AI[36]

SUMMARY OF KEY POINTS

1. **Generative AI.** This type of AI creates new content like text, images, audio, or designs by learning patterns from data. Unlike predictive AI, which forecasts or classifies, generative AI generates original outputs based on prompts.
2. **How generative AI works.** Generative AI uses deep learning models trained on large datasets to generate content by predicting pieces one at a time, using tokenization and autoregressive techniques.
3. **Generative AI business applications.** Generative AI is currently employed to automate marketing copy, aid in product design, simulate human voices, analyze contracts, and improve customer service through chatbots and avatars. Its capability to swiftly tailor and generate content renders it valuable across various industries.
4. **Generative AI risks and limitations.** Generative AI has potential issues, including factual inaccuracy, misuse for impersonation or fraud, bias amplification, and lack of transparency. Businesses should implement safeguards, verify outputs, and address ethical considerations before using these tools.
5. **Common generative AI business models.** Companies generate revenue from generative AI through various models: selling access to customizable foundation models, offering tools for specific tasks, tailoring solutions for industries such as health care or law, providing infrastructure-as-a-service, or supporting open-source tools with paid services.

QUICK WINS

1. **Craft better prompts.** Generative AI provides responses to prompts rather than merely structured data. How a question is asked or a request is articulated significantly influences the quality of the response.
2. **Understand text generation.** The process of text generation involves segmenting language into tokens. The artificial intelligence system predicts each token sequentially, utilizing probabilities acquired during its training to produce responses.
3. **Explore beyond text.** Generative AI isn't limited to words—it can clone voices, generate images, create avatars, and produce deepfake videos using similar foundational architectures.
4. **Verify outputs carefully.** Generative AI tools can hallucinate, replicate bias, or be exploited for misinformation. Treat outputs as drafts, not unquestioned truth.
5. **Evaluate monetization models.** Companies use varied business models, from open APIs to vertical-specific tools, to generate revenue. Understanding these helps assess their offerings and incentives.

KEY TERMS

Application Programming Interface (API). A software interface that enables applications to send requests to a foundation model (such as for summarization or translation) and receive generated outputs in response.

Autoregressive generation. A sequential process whereby a model forecasts the most probable subsequent token and appends it to the output sequence, continuing this procedure until the response is fully generated.

Custom generative AI model. A version of a generative AI model that has been adapted to meet the specific needs of a business or domain. Most custom models are created by fine-tuning a large, pretrained model on proprietary data to reflect a company's voice, terminology, or tasks.

Diffusion models. A type of generative AI model that creates images by starting with random noise and gradually refining it to match a text prompt. Used in tools like DALL-E 2 and Stable Diffusion, these models "reverse" the process of adding noise to images during training.

Fine-tuning. The process of adapting a pretrained model to a specific domain or task by continuing training on a smaller, specialized dataset. Fine-tuning helps the model reflect particular vocabulary, tone, or rules.

Foundation model. Also known as a pretrained model; a large, general-purpose model such as ChatGPT 5, Claude, or Gemini, trained on extensive public data. These models can be utilized directly or adjusted for specific tasks.

Generative adversarial network (GAN). An image-generation method where two networks, called "generator" and "discriminator," interact. The generator creates images, while the discriminator evaluates them, until realistic outputs are achieved.

Generative AI. Creates original content—text, images, audio, video, or designs—based on patterns from large datasets; unlike predictive AI, which forecasts or classifies, generative AI produces new creations.

Hallucination. A phenomenon where a generative model produces information that appears credible but is inaccurate or unsubstantiated, thereby presenting significant reliability concerns.

Image-based generative AI. A type of generative AI that creates original images from text prompts or other inputs. These models are used in marketing, design, and art creation.

Mathematical vector. A list of numbers that represents information in a way that AI models can understand and process. In generative AI, vectors are used to encode the meaning of words, images, or other inputs.

Multimodal generative AI. Generative AI systems that can handle and combine different types of inputs or outputs—such as text, images, and audio—in a single model. These tools enable experiences like describing an image in words or generating a video from a script.

Multimodal model. A system capable of processing multiple data types (e.g., text and image),

which allows tasks such as describing a photo or generating an image from a text prompt.

Pretraining. The initial phase of training a generative AI model using large datasets from diverse sources to learn general language or image patterns. This step creates a foundation model that can be adapted for many tasks.

Predictive AI. Involves traditional models that classify, score, or forecast outcomes using structured data (e.g., fraud detection, risk scoring), rather than creating new outputs.

Retrieval-augmented generation (RAG). An AI technique that combines a language model with an external search component. Instead of relying only on what the model has memorized, RAG retrieves relevant documents or facts in real time to help generate more accurate and up-to-date responses.

Text-based generative AI. A type of generative AI that produces humanlike text based on input prompts. It can be used for writing emails, reports, product descriptions, or chatbot responses.

Tokenization. Involves dividing input text into smaller units, such as words or characters, which can then be processed by the model effectively.

Traditional AI. AI systems designed to perform narrowly defined tasks such as classification, prediction, or optimization using structured data and rule-based or statistical methods. Often referred to as predictive or task-specific AI.

DISCUSSION QUESTIONS

1. **Spot the difference: Predictive vs. generative AI.** Table 6.2 compares predictive AI (which spots patterns) with generative AI (which creates new content). Name one big difference between the two, and give one quick example of each.
2. **Why choose a pretrained model?** Companies can fine-tune their own model or just call a pretrained one like ChatGPT. For a small firm on a tight budget, give one reason it might choose the pretrained option.
3. **Explain a real-world use case.** The chapter describes real-world uses of generative AI—such as Airbus designing lighter plane parts and JPMorgan's COIN system that reviews contracts. Pick one of these examples and explain, in your own words, why generative AI was useful.
4. **Why are hallucinations a business risk?** The text says generative AI can sometimes hallucinate—produce answers that sound right but are wrong. Why is that a problem for businesses that rely on the tool?
5. **Pick the best business model for a Sstartup.** Table 6.3 lists five business models. Which model seems easiest for a brand-new startup, and why?

EXERCISES

1. **Define generative AI.** In one or two sentences, write your own definition of *generative AI*, being sure to state how it differs from predictive or task-specific AI.
2. **Compare predictive vs. generative AI.** Create a two-column T-chart: on the left, list three traits of predictive AI, and on the right, list the matching traits of generative AI, drawing directly from **Table 6.2**.
3. **Match companies to AI benefits.** The chapter names Airbus, BMW, JPMorgan, and Sephora. Match each company to the benefit they gained from generative AI in an explanatory paragraph.
4. **Manage a generative AI risk.** Choose one risk from the "Risks and Limitations" section and write a three-point manager's checklist for reducing that risk.
5. **Test and compare prompt effectiveness.** Write two different prompts for ChatGPT that ask for the same information. Compare the outputs and note which prompt worked better, tying your observations to the "**Why Responses Can Vary**" box.

CASE STUDY 6.1: WHEN AI MAKES FALSE CLAIMS—THE OPENAI LIBEL CONTROVERSY

In June 2023, Mark Walters, a radio host from Georgia, became the central figure in one of the first defamation lawsuits involving a generative AI tool. Walters discovered that ChatGPT, the widely known chatbot created by OpenAI, had fabricated a story about him in response to a user query. The AI asserted, falsely and without evidence, that Walters was the defendant in a federal embezzlement case—something entirely untrue.

The user who received the false information was a journalist researching an unrelated court case involving the Second Amendment Foundation (SAF), a firearms advocacy group. Curious to learn more, the journalist asked ChatGPT for a summary of the complaint in the case. The AI generated a detailed, confident-sounding summary that included Walters as a named defendant. The fabricated summary even referenced a legitimate case citation and alleged that Walters had been accused of fraudulently misdirecting funds intended for SAF programs. None of this was true. Walters had no connection to the case.

Walters filed a defamation lawsuit against OpenAI after learning of the fabricated information. The case quickly gained national attention—not just for its novelty, but also for what it revealed about the limitations and legal risks associated with generative AI tools. It posed difficult questions for the courts: Could an AI system be held liable for defamation? If not, then who—if anyone—was legally responsible for its false statements?

OpenAI responded by emphasizing that ChatGPT is a tool, not a legal authority, and that users are informed of the system's limitations when they log in. Still, the incident raised serious concerns for businesses and regulators. At the heart of the issue was AI hallucination, where generative models create convincing but factually incorrect content. These hallucinations are not bugs in the traditional sense; they feature how large language models (LLMs) generate responses based on probability rather than fact-checking.

The Walters case highlighted a broader challenge for companies using generative AI: What happens when AI makes an error and someone gets hurt? From chatbots providing incorrect customer service advice to AI-generated reports riddled with inaccuracies, the potential for reputational and legal risk is substantial.

For businesses, the lawsuit was a wake-up call. Many had rushed to integrate generative AI tools into their operations, from customer service to HR to marketing, without fully understanding how these tools generate content or how to audit their accuracy. Legal experts have warned that disclaimers may not be sufficient, especially if the AI tool is presented as authoritative or if its content directly impacts individuals' reputations, finances, or health.

As of April 2025, the Walters case remains unresolved in court, yet it has already significantly influenced how companies approach AI governance. Some organizations have begun implementing stricter hallucination mitigation strategies, including human-in-the-loop review processes, retrieval-augmented generation (RAG) systems drawn from vetted sources, and automated reasoning tools validating content before it reaches users.

The lawsuit also sparked a renewed focus on transparency and explainability in AI systems. Users need to know what the AI says and why it says it. In highly regulated industries such as health care, law, and finance, companies are increasingly treating generative AI outputs with the same level of scrutiny as financial audits or clinical reports.

Ultimately, the case served as a poignant reminder of the evolving legal and regulatory landscape surrounding AI. Lawsuits like Walters may prompt courts and legislatures to define new standards for liability, content verification, and risk management. In the meantime, companies are advised to proceed cautiously, striking a balance between innovation and oversight.

The Walters incident, while extreme, is unlikely to be the last of its kind. As generative AI becomes increasingly pervasive in business and society, organizations will need to build trust in the technology and the governance systems that guide its use.

Discussion Questions

1. What responsibilities do companies have when deploying AI tools that could potentially produce inaccurate or defamatory content?
2. Do you think disclaimers are enough to protect businesses from liability when AI-generated outputs are incorrect? Why or why not?
3. What role should human oversight play in the deployment of generative AI tools, particularly in customer-facing roles?
4. If the AI output is technically generated by an algorithm, who should be held accountable when it causes harm—developers, users, or the company deploying the tool?
5. How might cases like this influence future regulations on AI content, particularly in areas like media, law, or health care?

Case References

Reason Foundation. “Response to State’s Motion to Dismiss.” PDF, filed January 2024. **https://reason.com/wp-content/uploads/2024/01/WaltersvOpenAIStateMTDResponse.pdf**

Walters v. OpenAI, L.L.C., 1:23-cv-03122, (N.D. Ga.). Last updated February 10, 2025. **https://www.courtlistener.com/docket/67654321/walters-v-openai-llc/**

CASE STUDY 6.2: AI VOICE CLONING

AI voice cloning has emerged as one of the most emotionally powerful and ethically complex forms of generative AI. This technology uses advanced machine learning to accurately replicate the intonation, rhythm, and accent of a person's voice. From giving back voices to those with speech loss to generating dynamic entertainment content, its potential is undeniable. But so is its capacity for harm. Voice cloning lies at the intersection of accessibility, innovation, and exploitation.

How It Works

Voice cloning models are trained using neural networks that analyze audio samples to learn the unique fingerprint of a speaker's voice. These models can then generate synthetic speech using text input, sometimes needing as little as five seconds of recording. Tools like *Descript's Overdub* and *Podcastle's Revoice* have made the process commercially accessible, sparking both creativity and concern.

Uses—and Abuses

Voice cloning technology has already had real-world consequences. In early 2024, New Hampshire voters received robocalls that sounded like President Joe Biden discouraging them from voting in the state's Democratic primary. The voice was a deepfake. Thousands of voters were misled by an AI-generated voice clone, a tactic now explicitly banned by the FCC under new robocall guidelines.

In Slovakia's 2023 election, deepfake recordings emerged days before the vote, falsely portraying opposition leader Michal Šimečka discussing election fraud and adding a tax on beer. His party lost narrowly, fueling accusations of digital interference. And in 2019, a UK-based company was defrauded of €220,000 when a scammer used an AI-generated voice clone of the CEO to request an urgent transfer.

Phishing attempts are also evolving. Attackers can now scrape audio from social media to build voice clones of a victim's friends or family. Combined with location data, these clones can deliver disturbingly convincing messages, such as a fake call from a spouse in distress, to trick people into surrendering money or confidential information.

When Voice Cloning Helps

Despite these risks, the same technology is being used to change lives for the better. *Project*

Revoice, for example, helps people with ALS restore their own voices using AI. By recording speech early in the disease's progression, users can continue speaking in their natural voice even after losing the ability to vocalize. Similarly, *Voiceitt* helps individuals with speech impairments interact with Alexa and other digital assistants. By learning and translating atypical speech patterns, it offers users a new level of independence.

Entertainment and education are also embracing the technology. Video game developers are using voice clones to dynamically generate dialogue, increasing immersion. Language learning platforms can offer customized pronunciation in different dialects. In customer service, Voice AI agents provide a consistent tone, no matter when or where a customer calls. Actor James Earl Jones gave permission to use AI to clone his voice for the character Darth Vader in future projects, ensuring continuity while allowing him to retire. This move sets a precedent for proactive, ethical voice rights management.

Grief and the Gray Areas

Perhaps the most emotionally charged use of voice cloning lies in grief counseling. AI can be used to recreate the voices of deceased loved ones to craft messages for birthdays or milestones. For some, this is comforting; for others, unsettling or even manipulative. Consent is rarely clear, especially posthumously, and raises profound questions about emotional impact and autonomy.

Legal and Ethical Challenges

Globally, the legal system is racing to catch up. In the United States, a patchwork of state laws defines whether voices fall under rights of publicity. The FCC's 2024 ruling banned AI-generated robocalls under the Telephone Consumer Protection Act, giving state attorneys general new tools to prosecute fraud. California, Michigan, and Minnesota have passed laws regulating political deepfakes during elections. At the same time, voice is a form of biometric data—meaning cloned voices could potentially defeat voice authentication systems, raising new concerns about identity theft and data security.

Discussion Questions

1. What safeguards should be in place before voice cloning technology is made available to the public?
2. Who owns the rights to a person's voice? Should individuals have legal protection over how their voices are used, and how does this apply after death?
3. What are the ethical differences between using voice cloning to preserve a legacy versus impersonating someone for emotional impact (e.g., grief or fraud)?
4. How should companies balance innovation and accessibility with the potential for abuse,

especially in politically sensitive or emotionally charged contexts?

5. In what ways might voice cloning disrupt existing security practices such as biometric authentication—and what alternatives should be considered?

Case References

"AI Character Voice Generator: How to Clone Any Voice for Video Games." HeyGen. February 21, 2025. **https://www.heygen.com/blog/ai-character-voice-generator**

Seitz-Wald, A., & Memoli, M. "Fake Joe Biden Robocall Tells New Hampshire Democrats Not to Vote Tuesday." *NBC News*, January 22, 2024. **https://www.nbcnews.com/politics/2024-election/fake-joe-biden-robocall-tells-new-hampshire-democrats-not-vote-tuesday-rcna134984**

Ducu, Gavril. "The Election That Wasn't Stolen—But Something Worse Happened." NAFO Forum, January 31, 2024. **https://nafoforum.org/magazine/the-election-that-wasnt-stolen-but-something-worse-happened**

Damiani, Jesse. "A Voice Deepfake Was Used to Scam a CEO Out of $243,000." *Forbes*, September 3, 2019. **https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/**

Lammers, Tim. "James Earl Jones Signed over Rights for AI to Recreate Darth Vader's Voice." *Forbes*, September 9, 2024. **https://www.forbes.com/sites/timlammers/2024/09/09/james-earl-jones-signed-over-rights-for-ai-to-recreate-darth-vaders-voice/**

National Conference of State Legislatures. "Deceptive Audio or Visual Media ('Deepfakes') 2024 Legislation." December 2024. **https://www.ncsl.org/technology-and-communication/deceptive-audio-or-visual-media-deepfakes-2024-legislation.NCSL**

"Project Revoice," Dentsu Health. **https://www.dentsuhealth.com/work/project-revoice**

"The Dangers of Voice Cloning and How to Combat It," The Conversation. **https://theconversation.com/the-dangers-of-voice-cloning-and-how-to-combat-it-239926**

"The Rise of AI Voice Cloning: How Scammers Are Using Technology to Mimic Your Loved Ones," RealCall, April 24, 2025. **https://www.realcall.ai/blog/the-rise-of-ai-voice-cloning-how-scammers-are-using-technology-to-mimic-your-loved-ones/**

"Voiceitt—Inclusive Voice AI," Voiceitt. **https://www.voiceitt.com/**

"Voter Suppression AI Robocall Investigation Update." New Hampshire Department of Justice, February 6, 2024. **https://www.doj.nh.gov/news-and-media/voter-suppression-ai-robocall-investigation-update**

"9 Must-Have Interactive Language Pronunciation Tools," Strikingly. **https://www.strikingly.com/blog/posts/9-must-have-interactive-language-pronunciation-tools**

CASE STUDY 6.3: REVOLUTIONIZING SOCCER TACTICS WITH TACTICAI

In elite sports, tiny advantages can decide championships. For Liverpool FC—one of the most decorated football clubs in the world—set pieces are recognized as moments of strategic opportunity. A *set piece* refers to any situation where play is restarted following a stoppage, such as free kicks, corner kicks, penalty kicks, or throw-ins. These structured opportunities allow teams to position their players and execute practiced strategies. Decisions around these plays are largely based on intuition, experience, and video analysis. However, AI is entering into the decision process with programs like *TacticAI*, a generative and predictive AI system developed in collaboration with Google DeepMind.

The Role of TacticAI

TacticAI is designed to assist Liverpool's coaching staff in evaluating and optimizing set-piece strategies. The system combines predictive AI, which forecasts likely outcomes of a given play setup, with generative AI, which proposes alternative player positions and movements to improve outcomes. It is trained on detailed positional data from thousands of corner kicks across several seasons of Premier League play. The model can simulate how different tactical changes might affect the outcome of a corner, such as player placement, marking assignments, or delivery trajectory.

This goal against Liverpool, scored by Fikayo Tomori after a miscleared corner and deflection, highlights the kind of set-piece vulnerability that TacticAI was designed to address. By analyzing thousands of corner kick scenarios, the AI system now helps Liverpool refine player positioning and responses to improve both offensive and defensive outcomes in moments like these.

TacticAI doesn't aim to replace coaches; instead, it enhances decision making by generating viable alternatives that may not be immediately obvious. According to a paper published in *Nature Communications*, Liverpool's coaching staff preferred the AI-generated tactics to their original set-ups 90 percent of the time when given the option blind.

How It Works

At the core of TacticAI is a graph neural network that captures the spatial and relational structure of players on the field. This model predicts not just binary outcomes (e.g., goal or no goal), but a distribution of likely consequences based on each setup. Once a baseline is established, the generative module proposes variations, slightly shifting a defender's starting position, altering a run path, or adjusting spacing, that are likely to improve the predicted success rate.

These proposals are not simply visualizations; they are data-grounded recommendations, allowing coaching staff to explore “what-if” scenarios in ways that were previously impossible without manually coding and reviewing endless hours of footage.

Business and Competitive Value

The potential value of such a system extends beyond the football pitch. Sports organizations are increasingly seeing AI as a core part of operations—affecting recruitment, injury prevention, fan engagement, and tactical preparation. In the context of set pieces, TacticAI helps clubs identify areas of inefficiency or predict vulnerability in opposing teams and multiplies the speed and scope at which new strategies can be tested and refined.

Potential Risks and Limitations

While TacticAI represents a breakthrough, it also introduces new challenges.

- ***Over-reliance on AI.*** There is a risk that coaching staff could come to rely too heavily on AI-generated suggestions, sidelining intuitive, experience-based insights that remain critical in high-pressure or fluid game scenarios.
- ***Data bias and model overfitting.*** If the training data includes tactical biases or reflects historical trends that don’t apply to future opponents, AI recommendations may reinforce outdated strategies.
- ***Competitive equity.*** Clubs with access to high-quality AI systems may gain disproportionate advantages, especially as tools like TacticAI become proprietary. This raises questions about fairness in leagues where resources are already unevenly distributed.
- ***Interpretability.*** Understanding why TacticAI recommends a particular formation is not always straightforward. Coaches may hesitate to adopt strategies they can’t fully justify to players or management.

The Future of Sports AI

As generative AI systems like TacticAI expand, they will likely play a growing role in other sports. Basketball, American football, and baseball, where structured plays and spatial dynamics are central, are ripe for AI-driven tactical analysis. Sports organizations will need to weigh potential performance gains against issues of transparency, ethics, and competitive integrity.

Discussion Questions

1. What are the benefits and limitations of using generative AI to optimize in-game strategy?
2. How should coaching staff balance AI recommendations with human intuition and real-time judgment?
3. What ethical concerns might arise if some teams gain access to significantly better AI tools than others?
4. In what ways could biased or outdated training data affect the performance of a system like TacticAI?
5. Should AI-generated strategies be explainable to players or is performance alone sufficient justification?

Case References

Katwala, Amit. "Google DeepMind's New AI Model Can Help Soccer Teams Take the Perfect Corner." *Wired*, March 19, 2024. **https://www.wired.com/story/deepmind-is-helping-soccer-teams-take-the-perfect-corner/**

Wang, Z., Veličković, P., Hennes, D., Tomašev, N., Prince, L., Kaisers, M., Bachrach, Y., et al. (2024, March 19). "TacticAI: An AI Assistant for Football Tactics." *Nature Communications* 15(1), 1906. **https://www.nature.com/articles/s41467-024-45965-x**

ENDNOTES

[1] Eloundou, T., Manning, S., Mishkin, P., & Rock, D. “GPTs Are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models.” *OpenAI*, March 2023. **https://arxiv.org/abs/2303.10130**

[2] 618Media. “Services 618Media: #1 Digital Marketing Agency.” https://www.618media.com/services.

[3] Radford, Alec, et al. “Improving Language Understanding by Generative Pre-Training.” *OpenAI*, 2018. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

[4] SendX. “A/B Testing: How to Optimize Your Email Campaigns.” SendX Blog. **https://www.sendx.io/blog/email-subject-line-ab-testing#:~:text=You%20send%20each%20version%20to%20a%20small%2C,for%20the%20rest%20of%20your%20email%20list**

[5] Tom’s Guide. “OpenAI Launches Custom Chatbots—This Is How They Work and What They Can Do.” Last modified November 7, 2023. https://www.tomsguide.com/news/openai-launches-custom-chatbots-this-is-how-they-work-and-what-they-can-d

[6] Boyer, David. “Case Studies: How SMEs Are Using AI to Compete with Big Players.” *Inside Small Business*, November 4, 2024. **https://insidesmallbusiness.com.au/technology/case-studies-how-smes-are-using-ai-to-compete-with-big-players**

[7] Lufthansa Industry Solutions. “SmartAssistantAI: AI Chatbot Implementation in Line with Your Needs.” *Lufthansa Industry Solutions*. **https://www.lufthansa-industry-solutions.com/de-en/solutions-products/artificial-intelligence/smartassistantai-ai-chatbot-implementation-in-line-with-your-needs**

[8] Elomia. “Elomia: Your Mental Health AI Companion.” *Elomia*. **https://elomia.com/**

[9] Humanly. “Conversational AI for Hiring.” *Humanly.io*. https://www.humanly.io/

[10] Duolingo. “Introducing Duolingo Max: AI-Powered Learning with GPT-4.” *Duolingo Blog*, March 14, 2023. **https://blog.duolingo.com/duolingo-max/**

[11] Tidio. “80+ Chatbot Statistics & Trends in 2025 [Usage, Adoption Rates].” Last modified October 18, 2024. https://www.tidio.com/blog/chatbot-statistics/

[12] Tidio. “80+ Chatbot Statistics & Trends in 2025 [Usage, Adoption Rates].” Last modified October 18, 2024. https://www.tidio.com/blog/chatbot-statistics/

[13] Parsani, Puran. “Beauty and the Bot: How Sephora Reimagined Customer Experience with AI.” *Cut the SaaS*, February 23, 2024. https://www.cut-the-saas.com/ai/beauty-and-the-bot-how-sephora-reimagined-customer-experience-with-ai

[14]

[14] Autodesk. (2019, November 19). “Autodesk and Airbus Demonstrate the Impact of Generative Design on Making and Building.” *Autodesk News.* **https://adsknews.autodesk.com/en/news/autodesk-airbus-generative-design-aerospace-factory/**

[15] Autodesk. (2019, November 19). “Autodesk and Airbus Demonstrate the Impact of Generative Design on Making and Building.” *Autodesk News.* **https://adsknews.autodesk.com/en/news/autodesk-airbus-generative-design-aerospace-factory/**

[16] **BMW.com**. “AI Design: Artificial Intelligence Gets Creative.” https://www.bmw.com/en/design/ai-design-and-digital-art.html

[17] Superior Data Science. “J.P Morgan—COiN—A Case Study of AI in Finance.” Last modified 2024. https://superiordatascience.tech/jpmorgan-coin-case-study

[18] UNESCO. “Generative AI: UNESCO Study Reveals Alarming Evidence of Regressive Gender Stereotypes.” *UNESCO*, March 7, 2024. **https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regressive-gender-stereotypes**; Esposito, P., Atighehchian, P., Anastasis Germanidis, A., & Ghadiyaram, D. “Mitigating Stereotypical Biases in Text-to-Image Generative Systems.” *arXiv*, October 10, 2023. **https://arxiv.org/abs/2310.06904**

[19] Superior Data Science. “J.P Morgan—COiN—A Case Study of AI in Finance.” Last modified 2024. https://superiordatascience.tech/jpmorgan-coin-case-study

[20] Allyn, Bobby. “Lucrative Tech Jobs Vanish as Klarna Replaces Workers with AI Chatbot.” *SFGATE*, May 17, 2024. https://www.sfgate.com/tech/article/lucrative-tech-jobs-vanish-klarna-gig-work-20339137.php

[21] OpenAI. “GPT-4.” https://www.openai.com/gpt-4

[22] Anthropic. “Claude.” https://www.anthropic.com/claude

[23] Google. “Gemini.” **https://www.google.com/gemini**

[24] Cohere. “Enterprise Language Models.” https://www.cohere.com

[25] Jasper. “Generative AI for Marketing.” https://www.jasper.ai

[26] Grammarly. “GrammarlyGO.” https://www.grammarly.com/grammarlygo

[27] **Copy.ai**. “AI Writing Tools.” https://www.copy.ai

[28] Harvey. “AI for Legal Tasks.” https://www.harvey.ai

[29] Nabla. “AI in Healthcare.” **https://www.nabla.com**

[30] Replit. “AI-Powered Coding Environment.” https://www.replit.com

[31] Amazon Web Services. “AWS Bedrock.” **https://aws.amazon.com/bedrock**

[32] Microsoft Azure. “Azure AI Studio.” **https://azure.microsoft.com/en-us/services/machine-learning/**

[33]

[33] NVIDIA. “DGX Systems.” **https://www.nvidia.com/en-us/data-center/dgx-systems/**

[34] Hugging Face. “Open Models and Enterprise Support.” **https://www.huggingface.co**

[35] Mistral. “Open-Source Models.” https://www.mistral.ai

[36] Stability AI. “Open Models and Services.” https://www.stability.ai

CHAPTER 6S
Generative AI Technical Supplement

Generative AI differs from traditional predictive AI because it doesn't just analyze or classify data but creates new content. These models synthesize outputs based on statistical patterns learned from large datasets. This ability powers text, images, audio, video, and code generation, enabling new applications in marketing, design, entertainment, customer service, and more. This supplement is intended to give you insight into how generative AIs create their output.

Note: For a foundational explanation of neural networks, attention mechanisms, and transformer architecture, see the ***Chapter 5 Technical Supplement****. This section builds on that material by introducing additional model types–such as GANs and variational autoencoders (VAEs)–that are particularly important in generative AI.*

KEY TECHNICAL COMPONENTS

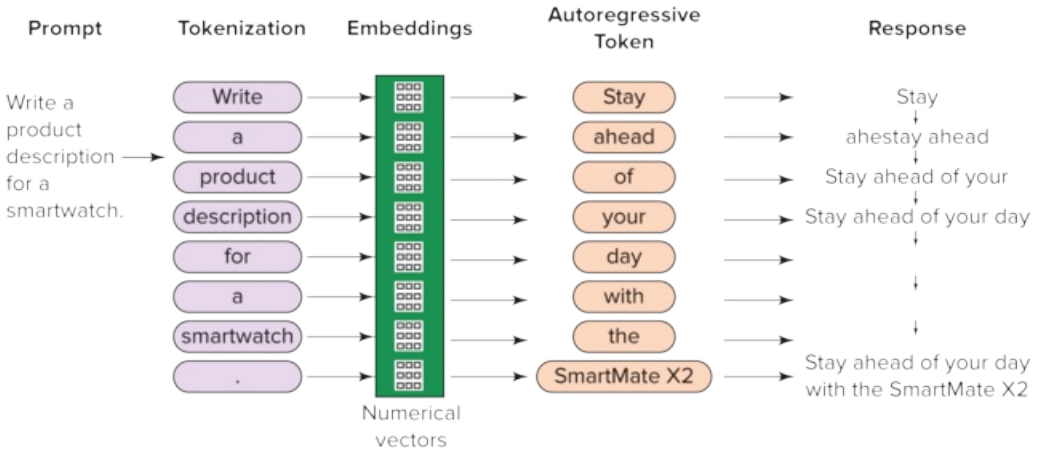
This section explores the technical building blocks that power generative AI systems—how they process inputs, generate outputs, and differ from predictive models.

Tokenization and Embeddings

Before generative models can work with inputs like text or images, they convert them into numerical form. Tokenization breaks input (e.g., text) into smaller units like words or subwords. A *subword* is a unit smaller than a word, which can be a character, a sequence of characters, or a morpheme (the smallest meaningful unit in a language). Each token is mapped to an *embedding*—a vector capturing its contextual meaning.[1] Embeddings are fundamental to understanding how models process and predict. As shown in Figure 6S.1, each token in a prompt is transformed into a vector representation that preserves contextual meaning. These embeddings guide the model's ability to predict the most likely next token, one at a time, allowing it to build a coherent response step by step. This is called *autoregressive*, and it means that with each successive token, the model generates its response based on the context of both the original prompt and everything it has generated so far.

FIGURE 6S.1 Autoregressive Text Generation in a Chatbot

This diagram illustrates how a generative AI model transforms a prompt into a response by predicting one token at a time. The model tokenizes the input, converts it into embeddings, and then iteratively generates each new token in an autoregressive loop—building the response step by step based on prior context.



Transformer Architectures and Autoregressive

Generation

Transformers are the dominant architecture in generative AI. A *transformer* is a type of neural network specifically designed to handle sequences of data, such as sentences or code, by evaluating all elements in parallel rather than one at a time. This enables faster processing and better understanding of context. Transformers use *attention mechanisms* to evaluate how much focus to place on different parts of the input when generating a response.[2] In practice, attention mechanisms allow the model to focus more on the most relevant words or phrases—determining, for example, that in the sentence "A smartwatch that tracks your steps and...," the word "tracks" is closely related to "steps" and, therefore, influences the prediction of what comes next. This improves the chances of generating a coherent and contextually relevant phrase like "monitors your heart rate" or "sends notifications."

Some transformer-based models are designed to generate responses only based on previously seen text—these are sometimes called *decoder-only models*, like GPT. Others process an input (like a question or image prompt) and then generate a response—these are known as *encoder-decoder models*, such as T5 or DALL-E.[3]

While the terms can be technical, the key difference is that decoder-only models are like writing a story one word at a time based on what's been written so far. In contrast, encoder-decoder models first process the full input and then generate an output based on that analysis. These models are well suited for tasks requiring long-range context, such as summarization, captioning, or dialogue.

CORE GENERATIVE MODEL TYPES

Generative AI encompasses several model types, each with strengths, weaknesses, and best-use scenarios. While they all aim to create new content, the way they do so varies widely, from generating text word-by-word to producing highly realistic images. This section introduces the primary families of generative models that power today's AI tools. They are organized in the order that reflects their real-world relevance and growing conceptual complexity from widely used text models to more specialized or foundational systems. Each model is explained in practical, business-relevant terms to help you understand how these tools function and why they matter.

Transformers

As noted earlier, transformers use attention mechanisms to model complex relationships across long input data sequences. Unlike earlier models that processed one token at a time, transformers can evaluate the context of the entire input at once, making them well-suited for tasks like language translation, summarization, and conversational AI. They are typically pretrained on massive general-purpose datasets and then fine-tuned on smaller, domain-specific datasets to perform specialized tasks such as customer service chat, legal document analysis, or code generation.

- ***Use cases.*** Text generation, chatbots, summarization, coding.
- ***Strengths.*** Scalable, high quality, adaptable.
- ***Limitations.*** Data and computing-intensive.

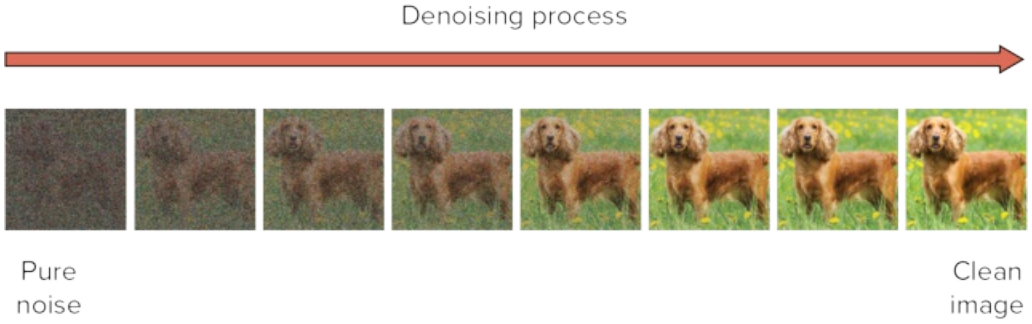
Diffusion Models

Diffusion models begin with a canvas of random noise—random values that contain no meaningful structure—and generate images by gradually removing the noise over many steps. This process is known as *denoising*. At each stage, the model refines its prediction of what a clean image should look like based on patterns it has learned from training data.[4] This process is similar to sculpting: Just as a sculptor chips away at a block of marble to reveal a detailed figure, a diffusion model removes layers of noise to reveal a coherent image. This denoising process allows diffusion models to produce highly detailed and realistic images, making them a popular choice for applications in digital art, advertising, and media generation. **Figure 6S.2** presents an example of the denoising process.

- ***Use cases.*** Art generation (e.g., Midjourney, Stable Diffusion).
- ***Strengths.*** High visual fidelity.
- ***Limitations.*** Slower generation, tuning complexity.

FIGURE 6S.2 Diffusion Model Process

A step-by-step visual of the forward noise process and reverse generation process in a diffusion model.



Labrador Photo Video/Shutterstock

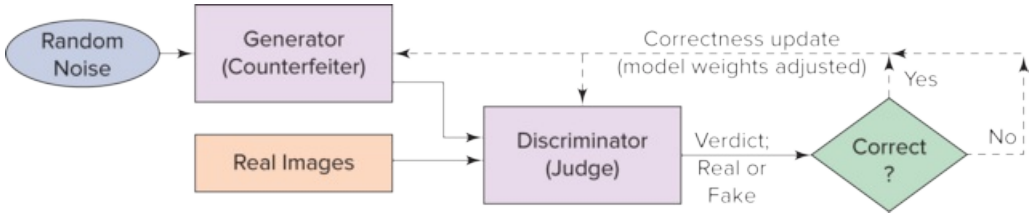
Generative Adversarial Networks

A *generative adversarial network (GAN)* is composed of two neural networks: a generator and a discriminator. The *generator* creates synthetic data, while the *discriminator* evaluates whether the input is real (from the dataset) or fake (from the generator). These networks are trained together in a competitive process, sometimes described as a "minimax game" in the literature.[5] The generator tries to fool the discriminator, and the discriminator tries to avoid being fooled. Over time, this adversarial setup pushes the generator to create increasingly indistinguishable outputs from real data. **Figure 6S.3** presents an overview of how GANs work.

- ***Use cases.*** Synthetic images, fashion design, avatars.
- ***Strengths.*** High-quality, lifelike visuals.
- ***Limitations.*** Training instability, limited to structured outputs.

FIGURE 6S.3 How GANs Work

A visual overview of how GANs operate shows how the generator creates synthetic (fake) data from noise, and the discriminator evaluates whether samples are real or fake. Both networks improve through feedback as they compete in an adversarial loop.



Variational Autoencoders

Variational autoencoders (VAEs) work by compressing input data, such as an image or product design, into a simplified, lower-dimensional representation called a *latent space*.[6] This compressed version captures the most essential features of the input. The model then reconstructs the original input from this latent space with slight variations, allowing for the creation of new but similar outputs. Unlike models focused on perfect realism, VAEs prioritize interpretability and control. This makes them especially useful in applications where the ability to tweak or interpolate between outputs is valuable, such as in product design, scientific modeling, or creating prototype variations.

- ***Use cases.*** Scientific modeling, product design, drug discovery.
- ***Strengths.*** Controlled generation, interpretable.
- ***Limitations.*** Less realistic than GANs.

Recurrent Neural Networks

Recurrent neural networks (RNNs) are one of the earliest types of neural networks designed to handle sequential data.[7] These models process input step-by-step, maintaining a memory of prior inputs to inform each new prediction. This ability to carry forward context makes RNNs useful for tasks like speech recognition, early text generation, and forecasting in time series data. However, they struggle with retaining information over long sequences, which limits their performance on more complex language or visual generation tasks.

- ***Use cases.*** Time series, early NLP.
- ***Strengths.*** Simpler, intuitive memory.
- ***Limitations.*** Poor with long-range dependencies.

Before exploring the mechanics of how generative models produce output, it helps to compare the major types. **Table 6S.1** summarizes each model's core function, common use cases, and practical strengths and limitations in a business context.

Table 6S.1 **Comparison of Generative AI Model Types**

Model Type	How It Works	Primary Use Cases	Strengths	Limitations
Transformers	Uses attention mechanisms to analyze all input tokens in parallel and predict output sequences one step at a time	Text generation, image captioning, code generation, summarization, chatbots	Scalable, accurate, handles long-range dependencies	Requires large datasets and compute power
Diffusion models	Starts with noise and iteratively denoises to produce content	Image generation (e.g., Midjourney, Stable Diffusion), animation	Produces highly detailed, diverse visuals	Slower inference, more complex to tune
GANs (generative adversarial networks)	Two networks (generator and discriminator) train together to produce increasingly realistic outputs	Image and video synthesis, deepfakes, synthetic photography, design prototyping	Produces high-quality images, realistic outputs	Harder to train, may be unstable
VAEs (variational autoencoders)	Compresses input into latent space then reconstructs output with variation	Product design, data augmentation, scientific modeling	Controlled, interpretable generation	Less realism than GANs
RNNs (recurrent neural networks)	Processes sequences one step at a time, using prior steps as memory	Early text generation, speech recognition, time series forecasting	Handles short sequences, simpler architecture	Struggles with long-term dependencies

HOW GENERATIVE AI BALANCES CREATIVITY AND COHERENCE

Generative models produce new content in different ways depending on their architecture. Some, like transformer-based language models, generate output one token at a time—building a sentence or paragraph step-by-step. Others, like GANs and diffusion models, generate entire images or outputs all at once or in iterative stages.

Sampling Strategies in Generative AI

A key factor influencing output in generative AI is the *sampling strategy*, a method for selecting among many statistically probable next tokens. These models do not "decide" based on understanding or intent. Instead, they rely on probability distributions learned during training. Sampling strategies guide how the model navigates this distribution to produce either consistent or varied outputs. This helps explain why the same prompt can yield different responses: The model is sampling from a range of likely continuations, not retrieving a fixed answer.

To manage this balance between coherence and originality, developers use sampling strategies like the following:[8]

- ***Temperature.*** Adjusts randomness. Higher values make the model more exploratory; lower values make it more predictable (more likely to take the most probable output).
- ***Top-k sampling.*** Limits predictions to the top k most likely tokens, focusing the output.
- ***Nucleus (top-p) sampling.*** Chooses from a dynamic set of top choices that together make up a specified probability threshold p.

These techniques help control the tone, diversity, and creativity of generated outputs.

Many open-domain AI tools, like ChatGPT, allow users or system designers to adjust temperature settings to match different use cases—ranging from highly factual assistant personas to more imaginative or creative ones. **Table 6S.2** shows the setting you would use for each method for a desired level of creativity in a generative AI.

Table 6S.2 **Sampling Strategies: Adjusting Probabilistic Outputs in Generative AI**

Sampling Method	Highly Predictable	Balanced Creativity	Highly Creative / Random
Temperature	0.0–0.3—Factual, consistent	0.4–0.9—Varied, natural	1.0–2.0—Surprising, imaginative
Top-k Sampling	Low *k* (e.g., 5–20)—Safe, limited options	Mid *k* (e.g., 40–80)—Focused but expressive	High *k* (e.g., 100+)—Expansive, diverse
Nucleus (Top-p)	Low *p* (e.g., 0.7)—Narrow, confident	Moderate *p* (0.85–0.9)—Balanced tradeoff	High *p* (0.95–1.0)—Risk taking, unexpected
Typical Use Cases	Legal writing, FAQs, tech support	Email drafts, product copy, education	Fiction, ad slogans, ideation tools

Generative AI Isn't Done Evolving

The field of generative AI is evolving at extraordinary speed. During the writing of this textbook, we have updated materials multiple times to reflect new tools, capabilities, and terminology—including the shift from standalone diffusion models to integrated multimodal systems like GPT-4o. What was once cutting-edge quickly becomes foundational. This pace demands a new model for sharing knowledge. Rather than relying on new editions every few years, we are embracing an *evergreen* approach to publication—one that supports ongoing updates and digital supplements so that students and instructors can engage with the most current understanding of this fast-moving field. As generative AI expands in scope and influence, we aim to provide a resource aligned with the latest real-world developments.

ENDNOTES

[1] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). “Efficient Estimation of Word Representations in Vector Space.” arXiv preprint *arXiv*:1301.3781. **https://arxiv.org/abs/1301.3781**

[2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). “Attention Is All You Need.” *Advances in Neural Information Processing Systems* 30. **https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf**

[3] Roberts, Jesse. “How Powerful Are Decoder-Only Transformer Neural Models?” *International Joint Conference on Neural Networks (IJCNN)*, 2024. **https://arxiv.org/abs/2305.17026**

[4] Ho, J., Jain, A., & Abbeel, P. (2020). “Denoising Diffusion Probabilistic Models.” *arXiv* preprint arXiv:2006.11239. **https://arxiv.org/abs/2006.11239**

[5] Roberts, Jesse. “How Powerful are Decoder-Only Transformer Neural Models?” *International Joint Conference on Neural Networks (IJCNN)*, 2024. **https://arxiv.org/abs/2305.17026**

[6] Roberts, Jesse. “How Powerful are Decoder-Only Transformer Neural Models?” *International Joint Conference on Neural Networks (IJCNN)*, 2024. **https://arxiv.org/abs/2305.17026**

[7] Ho, J., Jain, A., & Abbeel, P. (2020). “Denoising Diffusion Probabilistic Models.” *arXiv* preprint arXiv:2006.11239. **https://arxiv.org/abs/2006.11239**

[8] Maxime Lebonne. “Decoding Strategies in Large Language Models.” *Hugging Face Blog*, 2025. **https://huggingface.co/blog/mlabonne/decoding-strategies**

CHAPTER 7
Strategic AI Implementation: People, Process, and Purpose



Owlie Productions/Shutterstock

Learning Objectives

- **LO7-1** Describe factors that affect organizational readiness for AI.
- **LO7-2** Identify business functions that are suitable for AI integration.
- **LO7-3** Describe how organizations implement AI solutions, including selecting tools, working with vendors, and using frameworks like Agile to support successful integration.
- **LO7-4** Explain strategies to monitor and measure AI effectiveness.
- **LO7-5** Describe how to navigate basic ethical, legal, and regulatory considerations involved in AI implementation.
- **LO7-6** Recognize the unique challenges and risks of implementing generative AI, and identify strategies such as human oversight, platform safeguards, and legal protections to ensure its responsible use in business settings.

INTRODUCTION

Many organizations are eager to adopt artificial intelligence but are unsure where to start. Boards and business leaders increasingly see AI as essential for staying competitive. However, success requires more than simply choosing a new tool. It begins with understanding where AI fits and how to prepare the organization to utilize it effectively.

This chapter offers a guide for organizations at various stages of their AI journey. It covers how to:

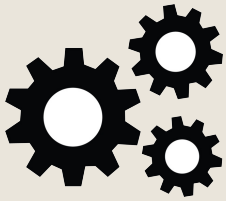
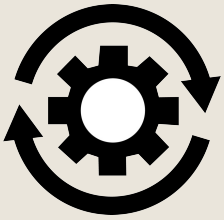
- Assess organizational readiness, including people, processes, data, and culture.
- Identify high-impact business functions where AI can create value.
- Integrate AI effectively using tools, vendors, and frameworks like Agile to support iterative improvement.
- Monitor and measure AI performance to ensure it delivers meaningful outcomes.
- Address ethical, legal, and regulatory challenges to build trust and ensure responsible use.

Whether the organization is just beginning or looking to expand its AI capabilities, this chapter provides the strategic foundation and operational insight needed to confidently move forward.

ASSESSING ORGANIZATIONAL READINESS FOR AI (LO7-1)

AI can potentially change how businesses operate, but success doesn't happen automatically. Before jumping into a project, organizations need to assess whether they're truly ready. That means asking practical questions: Do we have the right data? The right technology? Clear processes? People with the necessary skills? Without these foundations, even the best AI tools will struggle to deliver results. **Table 7.1** summarizes the major dimensions of readiness and the key questions or indicators of each. We will be looking at each of these in depth.

Table 7.1 **Assessing Organizational Readiness for AI**

Dimension	Key Questions or Indicators
Data 	• Do we have sufficient high-quality, accessible data? • Are data silos and standards addressed? • Are security and compliance policies in place?
Technology 	• Can our IT infrastructure support AI workloads? • Are legacy systems limiting implementation? • Do we need cloud, GPU, or vendor tools?
Processes 	• Are target workflows stable and well documented? • Are we solving a real, measurable problem? • Can we track input and output data?
Workforce 	• Do employees have necessary AI/data skills? • Are reskilling plans and training in place? • Is HR tracking change readiness?
Culture and leadership 	• Is leadership committed to AI initiatives? • Do we foster innovation and tolerate failure? • Is there cross-departmental collaboration?

Data Readiness

AI needs plentiful, high-quality data to operate effectively. Developing effective data management systems to generate, maintain, and utilize high-quality, accessible, and well-structured data is a

prerequisite for implementing AI. A 2024 report from the Association for Intelligent Information Management found that while 80 percent of organizations believed they were ready for AI, 52 percent encountered challenges, including poor internal data quality and organizational issues when introducing AI solutions.[1] In some cases, companies discovered that data had not been managed effectively for years, resulting in inefficiencies and errors.

Key Factors

When managing data for AI, it is important to consider a few critical areas we discussed in **Chapter 3**:

- ***Data availability.*** Sufficient data must be available to train, test, and sustain AI tools. Diverse and representative datasets that capture the range of scenarios that AI systems are expected to encounter are required for effective modeling.
- ***Data quality.*** Datasets must be accurate, complete, consistent, and free from biases. Poor data quality affects the performance of AI systems, generating unreliable results and perpetuating biases.
- ***Data accessibility.*** Data must be stored in formats compatible with AI tools and accessible only to authorized users, without unnecessary barriers. Efficient access to data accelerates AI development and implementation.
- ***Data security.*** Strong cybersecurity practices must be in place to protect data from breaches and misuse. Organizations should adopt best practices such as encryption, multifactor authentication, and real-time monitoring to safeguard sensitive information.
- ***Regulatory compliance.*** Companies must adhere to relevant laws and industry regulations when handling data.
- ***Data governance.*** Clear rules must be set for who owns the data, how it can be used, and how it will be monitored.
- ***Scalability and sustainability.*** Data systems must be able to grow as the organization's AI use expands without causing slowdowns or breakdowns. Sustainable data storage and processing practices can reduce environmental impact and support social responsibility goals.

Data Silos and Lack of Standardization

Many organizations have data stored in isolated systems, departments, or platforms that don't communicate well with each other. Such **data silos** often arise from legacy systems, mergers and acquisitions, or a lack of coordination across business units. One team might use a cloud-based CRM, while another tracks performance in Excel files saved on a local server. These systems may function just fine on their own—but when it's time to train an AI model, their incompatibility becomes a barrier.

Even when data is shared across teams, a lack of standardization can make it difficult to use. **Stan dardization** refers to having consistent formats, labels, units, and definitions across datasets.

Labels may differ ("Cust ID" vs. "Customer Number"), units may not match, and definitions may vary. What constitutes a "conversion" in one part of the business may have an entirely different meaning elsewhere. Standardization makes possible the linking of datasets, their effective cleaning, and the training of models that accurately reflect the actual business environment.

Integrating Data across the Organization

Once silos and standardization issues have been addressed, the next challenge is data integration, bringing together data from across the organization in a way that supports AI initiatives. Without integration, AI can't combine customer feedback with transaction history or link inventory data to supplier performance. Integration ensures that systems can share information, that data flows smoothly across departments, and that AI models can access a complete and accurate picture of business operations.

Successful data integration requires deliberate planning, collaboration across departments, and, often, the implementation of a new data management system. Some common strategies include:

- Building a *data lake* or *data warehouse* that centralizes information from different sources.
- Using *data integration tools* to automate the syncing and standardization of data fields.
- Establishing a shared *data dictionary* to ensure teams use the same terms and formats.
- Implementing *data governance policies* that promote consistency, ownership, and quality.

These efforts require more than just IT involvement. Business leaders, analysts, and system owners must work together to define priorities and coordinate changes. Data integration done well lays the foundation for effective AI training, modeling, and analytics.

Technological Readiness

Companies need sufficient computing power, storage capacity, and IT infrastructure to implement AI effectively to support data-heavy applications. Many older systems were not designed to handle modern data types like video, audio, or large volumes of unstructured text. These systems may also lack the speed and flexibility to run AI models efficiently. In some cases, companies rely on legacy platforms that use proprietary software—code that only the original developers can modify—which makes updates and integration with newer tools difficult and costly.

The technology, infrastructure, and expertise a company needs will vary depending on the scope of its AI initiative. One common starting point is the use of chatbots for customer service—a practical and visible application of AI that spans from simple tools to advanced systems. Let's look at this type of implementation to see the range of requirements needed:[2]

- *Cloud-based chatbots (software as a service, or SaaS)* require no internal hardware and are easy to set up. These are a good fit for small businesses because they are relatively inexpensive and don't require specialized technical skills.

- *Custom chatbots using machine learning* involve more complexity. These may require high-performance computers with large memory or access to cloud servers equipped with *graphics processing units (GPUs),* which are commonly used for AI tasks. Companies building these systems typically need at least one *machine learning engineer* or *data scientist* on staff to design and manage the chatbot.
- *Enterprise-level AI chatbots*, such as those powered by large language models (LLMs), require substantial computing resources—including multi-GPU servers and advanced cloud infrastructure. They also demand a full technical team, including AI/ML specialists, engineers who manage AI in production (*MLOps*), and cloud architects to scale and maintain the system. **MLOps** is short for "machine learning operations"; it refers to the practices and tools used to deploy, manage, and monitor machine learning models in production.

While advanced systems offer more customization and long-term scalability, they also come with significantly higher costs and greater staffing needs.[3] For many companies, these technical and resource requirements are important in deciding whether to build a solution in-house or work with a vendor.

Business Process Readiness

Before implementing AI, organizations must also clearly understand the business processes they want to improve. AI is most effective when applied to well-defined workflows with consistent inputs, measurable outputs, and existing performance benchmarks. If the process is broken, inconsistent, or poorly documented, AI may only reinforce inefficiencies or generate unreliable results.

Assessing process readiness involves asking:

- ***Is the process well understood?*** Can stakeholders describe how it currently works, what inputs it requires, and what outcomes it produces?
- ***What problem will the AI solve?*** Does the organization know what it hopes to improve or optimize (e.g., speed, cost, accuracy, customer experience)?
- ***Is the process stable enough to automate?*** AI struggles with constantly shifting workflows or one-off exceptions.
- ***Do we have the right data at each step?*** AI solutions require data-rich environments where actions and outcomes are already tracked or can be tracked reliably.

For example, applying AI to customer service chat interactions without clear escalation pathways or resolution criteria will limit its effectiveness. Similarly, using AI for procurement forecasting will fail if departments bypass centralized systems or lack data discipline.

Organizations should also assess whether existing digital systems—like ERP (enterprise resource planning), CRM, or ticketing platforms—accurately reflect the processes AI will interact with. AI cannot learn or operate effectively if data is stored in disconnected spreadsheets or inconsistently

logged.

By mapping and evaluating the processes AI is meant to enhance, organizations reduce risk, improve clarity, and are more likely to select an appropriate use case and implementation approach.

Workforce Readiness

Once organizations have confirmed that their data, technology, and business processes are ready for AI, they must turn to the readiness of their workforce. Many companies underestimate the significant impact of human behavior on the success or failure of AI projects. Implementing AI is more than installing new software. It is also about transforming the way people work, make decisions, and think about technology.

Adding AI to daily operations means verifying whether employees possess the necessary skills and providing support if they do not. Employees who use or maintain AI systems require specific technical and operational skills. AI implementation frequently reshapes workflows, meaning some employees may need to be reskilled or transitioned to new roles. A proactive approach to workforce readiness is crucial for ensuring smooth adoption and maximizing the value of AI initiatives.

HR's Role in Building AI-Ready Teams

Human resources (HR) is often the department responsible for ensuring a workforce is prepared for AI adoption. It assesses employee skills, conducts a **skills gap analysis**, identifies training needs, and gauges employee readiness for change. Gaps may include specialized skills such as machine learning and natural language processing and general competencies like data literacy and security. The skills needed will depend on the type of AI the organization plans to implement. HR helps plan workforce transitions through reskilling, redeployment, recruitment, or outsourcing.

Organizations can use specific workforce metrics to quantify their AI skills gap. These include tracking the percentage of AI-impacted roles currently staffed with trained personnel, measuring employee confidence in using AI tools, or calculating the average time it takes workers to gain proficiency in new AI systems. A training gap ratio can help prioritize investments in upskilling, while an internal **AI readiness index** can benchmark preparedness over time.[4]

The AI Skills Gap

Every AI project requires a team of people to collect the data, build the systems, and interpret the results. Since 2019, job postings for AI roles have increased by 21 percent annually, while compensation has risen by 11 percent per year.[5] However, a shortage of skilled professionals in AI and related technologies has emerged. Many positions remain

unfilled, and nearly half of executives cite a lack of in-house AI expertise as a major barrier to progress. The shortage of AI professionals is not confined to a single region. Across the globe, countries are experiencing serious gaps between demand and available talent. In the United States, projections suggest that up to 50 percent of AI jobs could go unfilled by 2027.[6]

The HR department also often focuses on addressing potential employee resistance through early communication and ensuring compliance with labor laws and ethical AI practices. In addition, HR must consider the diverse needs of the workforce during technology transitions. Ensuring all employees, regardless of age, background, or digital fluency, have access to training opportunities helps foster inclusion and reduce resistance to change. By tackling these issues upfront, HR minimizes disruption in an AI implementation.

Workforce Reskilling and Support

A huge issue for organizations is whether to hire new employees with needed AI skills or to reskill current employees. Employers may be more interested in upskilling/reskilling their current labor force if demand for AI-related skills far outpaces the available supply. Factors such as government tax credits, union activity and collective bargaining agreements, training program availability, regulatory schemes, and employee interest likely drive these workforce decisions.

The first step for companies is to identify the skills in high demand due to AI integration and develop a comprehensive reskilling program for their employees. This program should not only focus on technical skills related to AI but also on enhancing soft skills like critical thinking, problem solving, and adaptability. By investing in their employees' growth and development, companies can ease the transition to AI-augmented roles and improve employee loyalty and retention. It is equally important to foster a culture of continuous learning, encouraging employees to engage with evolving technologies and industry trends proactively.

Open and honest communication about AI's changes and impacts is also crucial. Employees should be informed about why AI is being integrated, how it will affect their jobs, and the company's steps to support them. Involving employees in the AI integration process itself can also be beneficial. This could be through feedback sessions, collaborative workshops, or inclusion in decision-making processes related to AI deployment. Such involvement helps address employees' concerns and provides insights that can guide the implementation of AI.

Support systems such as counseling, career guidance, and mental health resources should also be available to assist employees through the transition. Recognizing and addressing the emotional and psychological impacts of workplace changes is key to maintaining a healthy and productive work environment.

Note that federal, state, and local grant money or tax credits may be available to organizations to invest in reskilling local workforces. Community colleges, vocational schools, and universities often offer such opportunities, sometimes at lower costs. Companies can provide in-house training

and professional development activities, either through internal experts or by the use of external vendors. Strategic workforce development should ensure access to training that supports long-term employability and competitiveness across skill levels.

One company that recently launched a multiyear reskilling initiative to help its employees adapt to changing technology demands, including AI, is AT&T.[7] The company provided training in data analytics, machine learning, and other digital skills, giving employees pathways to transition into new roles. This reskilling initiative has supported AI adoption and promoted internal job mobility by preparing employees for emerging roles throughout the company.



Like missing puzzle pieces, the right skills are essential to a successful AI implementation.
Lightspring/Shutterstock

Cultural and Leadership Readiness

While data quality, technical infrastructure, and workforce skills are essential for AI readiness, they are insufficient to drive a successful AI implementation. Organizations must have a culture that permits change and a strong leadership commitment to improve processes to ensure a supportive environment for AI initiatives.

Cultural Readiness

Cultural readiness involves an organization's openness to change, innovation, and cross-functional collaboration. A culture that encourages experimentation without a fear of penalties for failing, values data-driven decision making, and supports continuous learning is more likely to integrate AI technologies successfully.

People and organizations can be resistant to change. There is comfort in routines and familiarity. Naturally, resistance arises when the status quo is compared to something that feels unfamiliar,

overhyped, and potentially threatening—especially if it could disrupt jobs or alter the way work gets done. Reducing resistance to AI requires a culture that supports learning and collaboration. When employees understand how AI can support or enhance their work, rather than replace it, they are more likely to engage with new tools. Involving the right people early, encouraging cross-functional conversations, and creating space for questions helps build trust and buy-in.

Leadership Readiness

Leadership plays a pivotal role in setting the tone for successful AI implementation. Effective leaders champion AI initiatives by articulating a clear vision, allocating necessary resources, and fostering an environment where cross-departmental collaboration is the norm. They also address potential resistance by engaging stakeholders early and promoting transparent communication. Their engagement through funding, messaging, and modeling AI use signals to the rest of the organization that AI is not a short-term experiment but a strategic priority.

A 2024 systematic literature review emphasizes that successful AI integration is closely linked to cultural alignment and proactive leadership. The study highlights that organizations with leaders promoting innovation and adaptability are better positioned to navigate the changes associated with AI adoption.[8] Ultimately, adopting AI isn't just about tools. It is about adopting a change mindset. Organizations need to cultivate cultures where experimentation is encouraged, failure is viewed as a learning opportunity, and AI is seen as a tool for empowerment. This involves providing space for training, encouraging questions, and acknowledging small wins. Companies that take the time to manage change effectively are more likely to unlock AI's full potential, not just technically, but also organizationally.

In many high-performing organizations, leadership readiness also includes a clearly articulated **innovation strategy**.[9] Rather than treating AI as a one-off technical upgrade, leaders integrate it into the organization's broader innovation goals—such as new product development, customer engagement, or business model evolution. By linking AI to strategic innovation, leaders create a shared sense of purpose, encourage forward-thinking experimentation, and help align teams around long-term value creation. This strategic framing of AI helps shift organizational focus from short-term efficiency gains to sustained transformation.

IDENTIFYING BUSINESS FUNCTIONS READY FOR AI INTEGRATION (LO7-2)

An important part of planning and executing AI projects is to identify the business problem to be solved. It might take some detective work to differentiate a symptom, such as the return rate of a particular item being higher than usual, from the actual problem, which could be a change in vendor, a problem with the manufacturing of that item, an issue somewhere in the supply chain, or several other possibilities.

Before jumping into specific departments, it's important to take a step back and ask: *What business problem are we really trying to solve?* Not every issue requires AI—and not every function is a good starting point.

The best AI projects begin with a well-scoped problem that is:

- Clearly defined and measurable.
- Rooted in an existing workflow.
- Supported by available data.
- Aligned with strategic priorities.

In fact, many successful AI implementations begin not by asking "Where can we use AI?" but by identifying a persistent bottleneck or inefficiency—and then exploring whether AI is the right solution. **Figure 7.1** presents a list of good and bad characteristics of processes to apply AI solutions.

FIGURE 7.1 What Makes a Good Problem for AI?

Use the checklist below to evaluate whether a business challenge is a strong candidate for AI implementation—especially in the early stages. Look for characteristics marked with ✅ and avoid those marked with ❌ until your organization gains more experience.

	Characteristic	Description
✅	**Repetitive tasks**	Involves rule-based, high-volume work like data entry or basic customer service
✅	**Structured or semi-structured data**	Reliable data exists and can be accessed for training and ongoing use
✅	**Clear objective or outcome**	Success can be measured (e.g., faster processing, improved accuracy, higher revenue)
✅	**Stable process**	The process doesn't change frequently and follows a consistent pattern
✅	**Feasible scope**	Can be implemented in manageable steps and scaled over time

✅ **Business alignment**	Addresses a real pain point or opportunity linked to strategy or operations
✅ **Opportunity for prediction or automation**	AI can add value by forecasting outcomes or reducing manual workload
❌ **Rare or one-off events**	Problems that are highly unique, infrequent, or lack data aren't good AI use cases
❌ **Noisy, incomplete, or sparse data**	Poor data quality undermines model performance and outcomes
❌ **Unstable or undefined processes**	AI will only replicate existing confusion or inefficiency
❌ **High ethical sensitivity without oversight**	Avoid applying AI to areas like hiring, lending, or health care without proper safeguards

Many organizations use structured process improvement and problem-solving frameworks like *agile*, *lean*, or *design thinking* to identify and prioritize the right business challenges to solve. Agile emphasizes iterative development and continuous feedback. Lean focuses on eliminating waste and maximizing value. Design thinking encourages empathetic exploration of user needs to inspire innovative solutions. These frameworks encourage teams to focus on user needs, process inefficiencies, and measurable outcomes before proposing solutions. They provide a tested way to uncover pain points across departments and evaluate whether AI is the right fit. These same approaches are often used again during implementation, helping teams prototype AI solutions, collect feedback, and iterate quickly.

While every organization's needs will differ, three business functions frequently emerge as early targets for AI integration: customer service, marketing, and operations and supply chain management. These are high-impact areas in many companies where AI can enhance efficiency, reduce costs, and improve the customer experience. Research from McKinsey[10] and Deloitte[11] shows that these functions lead in AI adoptions, while areas like legal, compliance, and HR lag due to data and privacy concerns. Let's examine the types of companies that are well-suited to start with one of these functions.

Customer Service

Companies with large call centers or extensive customer service operations, such as telecommunications firms, banks, and e-commerce platforms, can use AI to handle repetitive inquiries, enhance response times, and provide 24/7 service. Businesses that view customer satisfaction and loyalty as key success factors, such as those in the retail and hospitality sectors, can likewise use AI to deliver personalized and efficient service. Organizations experiencing rapid growth or seasonal spikes in customer inquiries, such as online retailers during the holiday season, can implement AI to scale up their customer service capacity without proportionally increasing

staff.

For example, generative AI, in the form of chatbots and virtual assistants, has been used to automate customer interactions. Chatbots and virtual assistants can handle routine customer inquiries and send more complex issues to humans. More complex implementations have incorporated machine learning to analyze customer data, enabling the prediction of product or service issues and the proactive offering of solutions.

One company that has taken advantage of AI is Zappos, which began integrating AI chatbots into its customer service in 2019 to enhance efficiency and personalization. The chatbots handle routine inquiries, provide immediate responses, and offer 24/7 support, significantly reducing wait times and operational costs. The AI-driven tools also offer personalized assistance based on customer history, resulting in user satisfaction above 90 percent. As a result, Zappos has seen a notable reduction in customer service costs, along with increased customer engagement and satisfaction.[12] As part of its personalization effort, it also created a virtual try-on environment. The application, which utilizes predictive models, has significantly reduced sizing-related return rates without compromising the customer experience.



Zappos uses AI-powered chatbots and personalization tools to enhance customer service and reduce return rates—delivering fast, tailored support around the clock.
mikewaters/123RF

Government agencies are also adopting AI to improve service delivery. For example, the U.S. Internal Revenue Service (IRS) utilizes AI to identify potential tax return fraud and prioritize audits.[13] State-level unemployment agencies have implemented AI to streamline claims processing and flag irregularities, thereby reducing wait times and improving service accuracy.

Marketing

Companies that rely heavily on data analytics to drive their marketing strategies, such as tech

firms and e-commerce platforms, are candidates to use AI to analyze customer data, predict trends, and optimize campaigns. Businesses seeking to deliver highly personalized customer experiences, such as luxury brands and online retailers, can leverage generative AI and machine learning to tailor and customize marketing messages, product recommendations, and offers based on user behavior and preferences. This level of personalization significantly enhances customer engagement and conversion rates. AI can also be used to optimize marketing campaigns in real-time by adjusting ad placements and targeting specific audiences with personalized messages.

Another widely used AI application in marketing is **sentiment analysis**, which uses natural language processing (NLP) to evaluate how customers feel based on written content—such as social media posts, product reviews, or survey responses.[14] By gauging whether sentiment is positive, negative, or neutral, companies can better understand customer perception of their brand or campaigns in real time. For instance, a company launching a new product might use sentiment analysis to monitor social media chatter and quickly identify emerging concerns or praiseworthy features. These insights help marketers fine-tune messaging, respond to customer needs, and manage brand reputation more effectively.

Organizations with extensive marketing operations, such as multinational corporations and large consumer goods companies, can use AI to automate and scale their marketing efforts, from content creation to social media management. Finally, companies in highly competitive industries, such as fashion and automotive, can employ AI to stay ahead of market trends by analyzing consumer behavior, predicting demand shifts, and tailoring marketing campaigns to promote customer engagement.

Coca-Cola embraced generative AI technology to enhance its marketing efforts in early 2023.[15] Partnering with OpenAI and Bain & Company, Coca-Cola launched the “Create Real Magic” platform, which allows digital artists to create original artwork using Coca-Cola’s iconic assets. This initiative has significantly boosted consumer engagement and creativity, highlighting the potential of AI in marketing. Despite some challenges, such as maintaining brand consistency and addressing ethical concerns, Coca-Cola’s use of AI has led to innovative marketing campaigns and improved customer interactions.

Operations and Supply Chain Management



Images from Coca-Cola's "Create Real Magic" campaign, where generative AI empowered digital artists to remix iconic brand elements and engage audiences in new, creative ways.
pitinan/123RF

AI applications in operations and supply chain management often rely on proven techniques for prediction and optimization. These tools enable companies to improve demand forecasting, inventory management, dynamic pricing, predictive maintenance, and logistics. Companies with intricate supply chains, such as manufacturing giants, logistics providers, and multinational corporations, are prime candidates to use AI in these ways.

Businesses that need to quickly adapt to market changes or disruptions, like those in the food and beverage industry or electronics manufacturing, can also leverage AI to enhance their supply chain resilience and agility. By analyzing real-time sales data from electronic point-of-sale (EPOS) systems and external market trends, AI can help businesses maintain optimal stock levels, avoiding both overstocking and stockouts while improving responsiveness to customer demand.

Firms that rely heavily on data for operational decisions, such as automotive manufacturers and large retailers, can integrate AI to analyze vast amounts of data and improve supply chain efficiency. Processes that are repetitive, rule based, and time consuming are ideal candidates for integration with AI and robotics. In logistics, AI can optimize delivery routes by analyzing variables such as traffic, weather, and delivery schedules, thereby improving efficiency, reducing transportation costs, and enhancing on-time performance.

Unilever provides a good example of how AI can be applied across multiple areas of the supply chain.[16] The company utilizes predictive analytics to optimize inventory and replenishment, thereby reducing costs and enhancing efficiency. It also applies AI for route planning, significantly

lowering transportation expenses and delivery times. In its broader supply chain strategy, Unilever has even incorporated satellite imagery to support sustainability efforts, enhancing traceability and transparency in sourcing and logistics.



Unilever, parent company of Ben & Jerry's, uses AI to optimize supply chain operations from inventory forecasting to route planning.

Holly Curry/McGraw Hill

Other Departmental Applications

While these three areas are the most common entry points for AI integration, many companies are beginning to use generative AI within them. Tools like ChatGPT, Claude, and Google Gemini can assist with answering customer questions, generating marketing content, and summarizing internal reports. Because these tools are cloud based and easy to test, they offer a low-risk way to explore AI. For organizations just getting started, generative AI can deliver quick wins and help teams gain hands-on experience.

AI is also used to support **dynamic pricing**, adjusting prices in real time based on demand patterns, inventory levels, competitor pricing, and customer behavior. This allows companies to maximize revenue while maintaining market competitiveness. Industries such as retail, hotels, airlines, and ridesharing commonly employ dynamic pricing to respond quickly to changing market conditions.

Aligning AI Projects with Business Objectives

It is easy to get caught up in the hype of AI, but like all technologies, AI is a tool, not a goal. Its real value comes from how well it helps an organization achieve its business objectives. As noted earlier, decisions to integrate AI into company processes should be driven by its ability to create value for customers, improve operations, or support strategic priorities, rather than a vague sense of needing to "do something with AI."

Beware of the Pilot Trap

The **pilot trap** refers to the common scenario where AI projects remain stuck in the pilot or testing phase and fail to scale into full production.[17] When generative AI appeared on the business horizon in 2022, many companies gave their tech teams the freedom to experiment with this technology. About 70 percent of those projects never progressed beyond the lab. Two major obstacles typically prevent scaling:

1. Many early models suffered from issues with accuracy and reliability.
2. Teams often lacked access to the corporate data needed to power AI at scale. Without high-quality, well-governed data, even the most promising pilot struggles to deliver value.

To avoid the pilot trap, companies should break AI projects into small manageable projects that can deliver results quickly. Projects that can demonstrate meaningful productivity gains or revenue impact within 12 months are far more likely to gain continued support and scale.

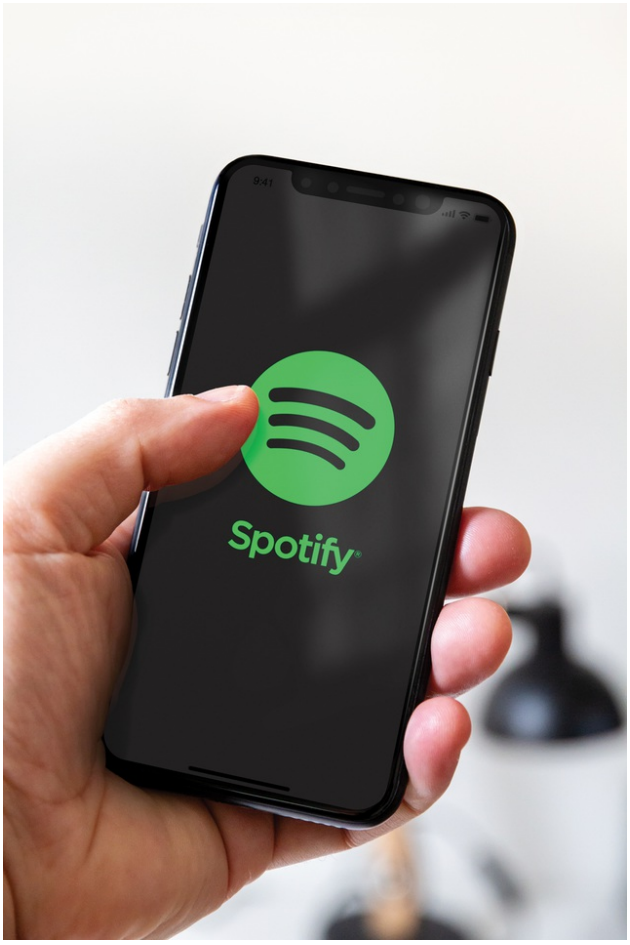
Many companies have projects that fall into the pilot trap. Surveys suggest this is a widespread problem. A 2024 report from Informatica[18] found that two-thirds of organizations struggle to move generative AI projects beyond the pilot stage, with nearly all reporting difficulty in proving business value. Yet few companies want to be publicly associated with AI project failures, which makes these real-world examples hard to find.

Success Stories

There are, however, examples of organizations that have successfully navigated this challenge by aligning their AI efforts with defined needs and existing data ecosystems. General Mills, for instance, developed *MillsChat*, an internal generative AI tool built on Google's platform.[19] The tool assists employees with writing, summarizing, and brainstorming tasks, focusing on enhancing communication with customers and suppliers as part of an effort to increase supply chain efficiency.

Procter & Gamble has taken a different but equally strategic approach with its internal generative AI tool *ChatPG*, aiming to develop employee skills and responsible-use knowledge regarding generative AI.[20] To ensure the integrity and security of company data, the team established controls and safety measures. Employees must undergo governance training before using ChatPG. In both cases, success was made possible by strategic alignment, access to data, and a clear operational purpose.

One company that has used the Agile framework is Spotify. The audio streaming service needed to enhance its music recommendation system to provide more personalized user experiences. To this end, it used its own Agile framework, known as the "Spotify Model," which emphasizes team autonomy and iterative development.[21] Cross-functional teams, including data scientists and engineers, worked in sprints to develop and refine AI algorithms for music recommendations. The iterative approach enabled Spotify to continually improve its recommendation system based on user feedback, resulting in increased user engagement and satisfaction.[22]



Spotify continuously improves its AI-powered recommendation system through cross-functional teams using the Spotify Model—its Agile framework that emphasizes autonomy, collaboration, and rapid iteration.

Daboost/Shutterstock

IMPLEMENTING AI SOLUTIONS (LO7-3)

The organization is ready for AI and has identified the business processes where it can add value. What comes next? As earlier chapters have shown, there are many types of AI algorithms to choose from—each suited to different tasks. The right algorithms must be selected, tested, and aligned with organizational needs. Just as important, companies must decide how to implement the solution: Should they build it in-house, subscribe to a software-as-a-service (SaaS) platform, or work with outside vendors to develop a custom system? Each approach has its own tradeoffs in cost, speed, flexibility, and control. Before diving into vendors or deployment models, we will look at the AI tool selection process.

Selecting the Right AI Tools

Selecting the right AI tools begins with understanding which business activity can best advance business goals with AI. A variety of tools are available, each suited to specific functions, such as content creation, customer support, product development, or data analysis. Table 7.2 outlines common business activities and highlights AI apps that organizations can use to enhance performance, reduce costs, or improve decision-making in each area.

Table 7.2 **Sample AI Tools for Business Activities**

Activity	Description	Example Tools
Content Creation and Communication		
Creating and customizing content	Marketing, advertising, and public relations businesses can benefit from generative AI tools. These tools excel at producing creative content, such as marketing copy or visuals. They can also tailor messages for different audiences and adapt based on engagement feedback.	Jasper and Canva with AI integrations
Language translating and localization	Natural language processing (NLP) models are the best choice for organizations operating across multiple linguistic regions. These AI systems translate and localize content while preserving meaning and cultural nuances, making them ideal for creating training materials or global communications.	Google Translate, Microsoft Translator
Customer Experience and Personalization		
Personalizing customer experiences	AI systems using machine learning effectively analyze customer data to provide personalized shopping experiences, such as tailored recommendations or customized communication	Amazon Personalize, Salesforce Agentforce 2.0

	recommendations or customized communication. Machine learning models analyze patterns in user behavior to create dynamic, individualized interactions.	Agentforce 2.0
Providing customer support and service	Chatbots powered by NLP are ideal for handling customer inquiries, providing instant responses, and escalating complex cases to human representatives. These systems enhance efficiency while maintaining a high standard of customer service.	Zendesk, IBM Watson Assistant
Design, Innovation, and Prototyping		
Developing new products	Generative AI is a powerful ally in innovation and prototyping. It can simulate outcomes, generate new ideas, and predict market reactions, streamlining the research and development process.	Autodesk Fusion 360 with Generative Design, Adobe Firefly
Assisting with software development	AI coding assistants can help developers generate code snippets, explain existing code, suggest bug fixes, and automate repetitive tasks. These tools enhance productivity, reduce development time, and support onboarding of less experienced programmers.	GitHub Copilot, Amazon CodeWhisperer, Tabnine
Business Intelligence and Analytics		
Analyzing high-volume data	For businesses dealing with large datasets, machine learning excels in processing and interpreting information, whether it is market research, customer feedback, or financial analysis. These models can identify trends and generate actionable insights.	Tableau with Einstein AI, Microsoft Power BI
Predicting outcomes and assessing risk	Machine language models are effective at predictive analytics in finance and supply chain management domains. These tools forecast market trends, assess risks, and can recommend mitigation strategies based on historical and real-time data.	Aible, Alteryx
Employee Development		
Training and simulation	Generative AI is highly effective in creating realistic training simulations in health care, aviation, or customer service industries. These tools enhance the learning experience by providing dynamic and interactive training environments.	Suki.AI, Microsoft's Custom Neural Voice

outputs, enabling efficient automation without the need for machine learning.

AI, however, extends automation into more complex territory, handling unstructured or semi-structured data, making decisions, and adapting to changing conditions. It is better suited for tasks that involve natural language processing, analyzing complex datasets, or predicting outcomes from

patterns. The most effective automation strategies often integrate RPA and AI to achieve **intelligent process automation (IPA)**. In this model, RPA manages structured workflows such as triggering actions or retrieving data, while AI handles cognitive tasks like interpreting the text, making decisions, or adapting workflows in real time.

One company that implemented IPA and RPA is Sappi. Headquartered in South Africa, Sappi is one of the world's leading producers of pulp and paper, with manufacturing operations on three continents and customers in more than 150 countries. Processing large volumes of customer orders, creating invoices, and tracking deliveries involved time-consuming and repetitive tasks. Sappi implemented IPA to improve efficiency, combining RPA with tools that could handle more complex decisions.[23] Software bots now manage routine tasks like extracting order details and updating systems, which has reduced processing time and cut down on errors. To handle less structured work—like interpreting customer emails or spotting unusual ordering patterns—Sappi added automation tools that could learn from data. This combination allows Sappi to manage the full process more efficiently while still involving employees when needed. The result has been faster service, more accurate billing, and better insight into customer demand.

AI Agents

AI agents such as UiPath, Blue Prism, and Automation Anywhere are emerging tools that can carry out multistep tasks with minimal human input. Unlike RPA, which follows fixed rules, AI agents can adapt to changing information, make context-based decisions, and operate autonomously across multiple systems. As they develop, these agents may extend automation into areas that previously required human oversight, blurring the line between simple task automation and autonomous problem solving. This synergy allows businesses to automate a broader range of processes, from routine operations to more dynamic, decision-intensive tasks.

One company that makes use of AI agents is U.S.-based insurance company Lemonade. Unlike traditional insurers, Lemonade uses a flat-fee model and operates primarily through a mobile app and website. At the heart of Lemonade's digital infrastructure are AI agents like "AI Jim," which handles customer service and claims processing. AI Jim goes far beyond a typical chatbot.[24] It uses machine learning and natural language processing to analyze customer-submitted video claims, assess emergencies, detect fraud, and determine whether to pay or escalate the claim. Lemonade's fully digital claims process allows customers to explain their situation in a short video, which AI Jim evaluates for consistency, urgency, and risk. This approach enables near-instant approval for straightforward cases, sometimes in under 3 minutes, and also captures valuable metadata often missed by human agents that could indicate fraud.[25] Claims flagged for further review are passed to a human adjuster, who can replay the customer's video as needed, improving the accuracy and fairness of decisions.

Selecting the Right AI Vendors

Building a robust AI system from scratch requires specialized knowledge and significant resources, which are often beyond the reach of companies at the start of their AI journey. To overcome these challenges, many organizations turn to third-party vendors and external experts who provide ready-made AI solutions or customizable platforms. These partnerships can significantly accelerate AI adoption, allowing businesses to focus on their core competencies while leveraging state-of-the-art technology.

Selecting the right third-party AI vendors is critical to successfully integrating AI into business operations. This decision affects not only the immediate functionality of the AI solution, but also its long-term outcomes, including scalability, data security, and alignment with business objectives.

In-house development teams often require skills in data engineering, machine learning, prompt engineering (for generative AI), and cloud deployment. While some AI tools are plug-and-play, custom solutions often require a dedicated team with experience in AI development, integration, and ongoing maintenance.

Many organizations choose to build or deploy AI applications through established platforms such as Amazon Web Services (AWS), Microsoft Azure, Google Cloud, or OpenAI's API services. These providers offer scalable infrastructure, pretrained models, and integration tools that can accelerate implementation and reduce development costs.

A thoughtful evaluation process helps ensure that the chosen vendor can meet the organization's unique needs, mitigate risks, and deliver measurable value. The following sections outline key considerations when assessing potential AI vendors, offering a practical framework to guide this pivotal decision.

1. Alignment with Business Needs

Before selecting a vendor, organizations should assess whether the solution aligns with their industry context and specific business objectives:

- ***Domain expertise.*** Choose vendors with a demonstrated track record in your industry. For example, those specializing in health care will be more familiar with compliance frameworks such as HIPAA, while retail-focused vendors may offer advanced recommendation engines.[26]
- ***Use-case fit.*** Ensure the vendor's AI tools are designed to meet your business objectives, such as improving customer personalization, automating internal processes, or extracting insights from large datasets.[27]

2. Scalability and Flexibility

AI solutions should be able to grow with your business and adapt to changing needs:

- ***Growth potential.*** Assess whether the vendor's offerings can scale as your organization expands, supporting increased data volumes, users, or use cases.[28]
- ***Customization options.*** Determine whether the AI tools can be tailored to your workflows and

integrated with existing platforms, such as ERP or CRM systems.[29]

3. Data Security and Privacy

Strong data protection practices are essential when working with AI vendors. Ensure their systems meet regulatory, technical, and ethical standards:

- ***Compliance with regulations.*** Vendors must adhere to data protection standards, such as the GDPR, the AI Act, and the CCPA, ensuring data privacy and ethical handling (see Chapter 3). Additional regulations will apply in some instances, such as:[30]
  - Health Insurance Portability and Accountability Act (HIPPA) for medical records.
  - Family Educational Rights and Privacy Act (FERPA) for educational records.
  - Fair Credit Reporting Act (FCRA) for consumer reports.
  - Federal Information Security Management Act (FISMA) for federal data.
  - Personal Information Protection Law (PIPL) covering residents of China.
- ***Data ownership.*** Confirm that your organization retains ownership and control over its data and any AI-generated outputs.[31]
- ***Security measures.*** Evaluate the vendor's approach to cybersecurity, including practices such as encryption, secure data storage, and regular vulnerability assessments.[32]

4. Ease of Integration

AI tools should integrate seamlessly with your existing digital environment to minimize disruption:

- ***Compatibility.*** Ensure that the AI system is compatible with your existing software, systems, and workflows. Look for open application programming interfaces (APIs) and robust documentation that support integration.[33]
- ***Onboarding support.*** Choose vendors who provide thorough implementation support, including training for IT teams and end users to accelerate adoption and reduce friction.[34]

5. Performance and Reliability

Reliable AI systems should deliver measurable results and consistently meet performance expectations:

- ***Proven track record.*** Select vendors who can provide case studies, client testimonials, or benchmarks that demonstrate consistent results in real-world applications.[35]
- ***Metrics and KPIs.*** Ask vendors how they measure success and ensure their proposed performance indicators align with your business priorities.[36] Common key performance indicators (KPIs) include model accuracy, precision and recall, uptime or response latency for

AI services, user adoption rates, and measurable impacts on business outcomes such as cost savings or revenue growth.

6. Technical Support and Maintenance

Ongoing support is critical for long-term success, especially as technology and organizational needs evolve:

- ***Post-implementation support.*** Confirm that vendors offer responsive, accessible support channels to resolve issues as they arise.[37]
- ***Updates and upgrades.*** Ensure the vendor provides regular software updates to improve performance, address vulnerabilities, and incorporate new features as the technology advances.[38]

7. Vendor Reputation

Reputation reflects both the vendor's experience and their standing in the market:

- ***Research.*** Research vendor reviews, third-party ratings, and industry recognition to assess their credibility and customer satisfaction.[39]
- ***Innovation.*** Look for vendors who are actively investing in R&D and staying ahead of trends in AI technology.[40]

8. Cost and ROI

AI solutions should demonstrate clear value relative to their cost and implementation effort:

- ***Transparent pricing.*** Favor vendors who offer clear, upfront pricing models without hidden costs or complex licensing structures.[41]
- ***Value proposition.*** Evaluate the total cost of ownership against the expected business benefits, such as time savings, cost reductions, or revenue growth.[42]

One area where costs and return vary widely is chatbot implementation. **Table 7.3** outlines typical development and operational costs for three types of chatbot solutions, along with required expertise and common use cases.

Table 7.3 **Cost and Resource Requirements for Different Chatbot Types**

This table illustrates how AI implementation costs scale with complexity and business needs—from plug-and-play SaaS options to enterprise-scale large language model (LLM) systems.[43]

	Software as a Service (SaaS) Chatbot (Plug-and-Play)	Custom ML Chatbot	Enterprise-Scale LLM Chatbot
Infrastructure needs	Cloud-hosted (no internal hardware required)	1–4 GPUs or high-RAM CPU servers (cloud/on-premises)	Multi-GPU clusters, 128–512 GB RAM, scalable cloud AI
Development costs	$10,000–$30,000	$50,000–$100,000	$100,000–200,000
Estimated monthly cost	$50–$1,000	$1,000–$20,000+	$20,000–$100,000+
Internal expertise required	Basic IT setup; no coding or AI expertise required	ML engineer or data scientist; development for integration	Full AI/ML team; MLOps engineer; cloud architects
Deployment context	Small businesses; fast deployment using tools like Zendesk, Intercom, or Drift	Companies needing brand-specific tone, domain-specific logic, or analytics	Global firms with high traffic, multilingual AI assistants, or sensitive data compliance needs

Example: Electrolux and Phenom—AI-Driven Recruitment

Electrolux, a *Fortune* Global 500 manufacturer of appliances, was having difficulties finding high-quality job candidates.[44] The company decided to partner with Phenom, an AI-powered talent experience platform, to modernize its recruitment process. Phenom offered a suite of tools, including a personalized career site, an internal talent marketplace, and automated candidate matching. The integration of AI-driven fit scoring, one-way video interviews, and automated scheduling streamlined the hiring process, resulting in an 84 percent increase in application conversion rates and a 9 percent reduction in time-to-hire. Considering the criteria for selecting a third-party AI vendor, Phenom was a good fit for Electrolux:

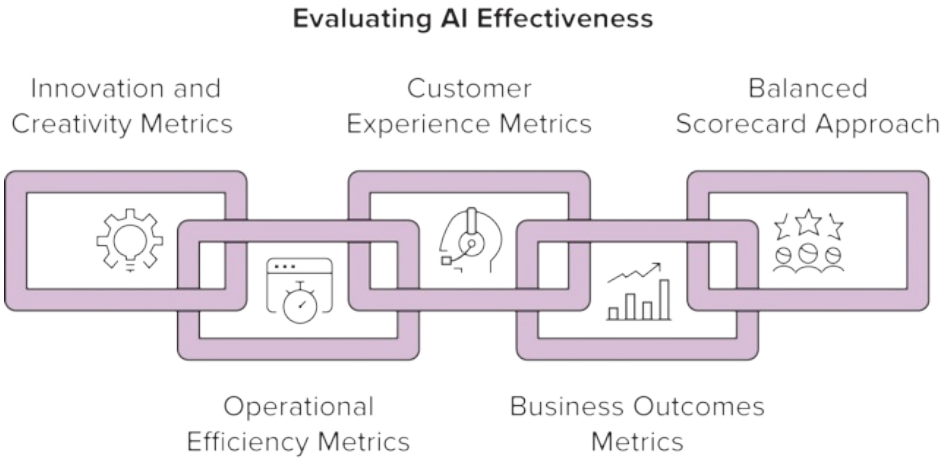
- ***Alignment with business needs.*** Phenom's platform addressed Electrolux's goals of improving candidate experience and internal mobility.
- ***Scalability and flexibility.*** The solution accommodated Electrolux's global operations, adapting to regional requirements.
- ***Ease of integration.*** Phenom's tools seamlessly integrated with Electrolux's existing HR systems, facilitating a smooth transition.

- ***Vendor reputation.*** Phenom's expertise in AI-driven recruitment solutions provided confidence in achieving desired outcomes.
- ***Return on investment.*** The partnership led to measurable improvements in recruitment efficiency and candidate engagement.

MONITORING AND MEASURING AI EFFECTIVENESS (LO7-4)

An ineffective AI tool is no better than no tool at all. To assess the effectiveness of AI applications, businesses must establish clear metrics. As shown in **Figure 7.2**, these metrics help quantify AI's impact on innovation and creativity, operational efficiency, the customer experience, and business outcomes. A balanced scorecard approach offers a structured way to track performance across all of these dimensions, helping organizations ensure their AI initiatives deliver both short-term results and long-term strategic value. This section discusses each of these factors.

FIGURE 7.2 Evaluating AI Effectiveness



Innovation and Creativity Metrics

Evaluation should focus on originality, engagement, and personalization for applications where AI is used to drive innovation or generate creative outputs.

- ***Engagement rates.*** Measure how AI-driven initiatives resonate with users or customers. Page 172
  - Example: A media company uses AI to suggest personalized content. Tracking views, likes, and shares of recommended articles or videos evaluates audience interest and success.
- ***Innovation index.*** Assess AI-generated ideas or designs' novelty and competitive advantage.
  - Example: A manufacturing firm using AI for product design compares new prototypes with existing models to evaluate their originality and appeal.

- ***Personalization success.*** Evaluate how well AI customizes user experiences or products.
  - Example: A retail platform tracks how AI-driven product recommendations influence customer preferences and engagement.

Operational Efficiency Metrics

AI often enhances operations by automating tasks and improving workflows. Metrics in this domain focus on time savings, output quality, and efficiency.

- ***Time savings.*** Quantify reductions in time spent on manual processes.
  - Example: An HR department using AI for candidate screening measures how much faster the hiring process becomes compared to traditional methods.
- ***Output quality.*** Evaluate the accuracy and reliability of AI-assisted outputs.
  - Example: A translation agency assesses the accuracy of AI-translated documents against industry benchmarks or manual efforts.
- ***Process efficiency.*** Measure the improvement in workflow completion rates or error reduction.
  - Example: A logistics company using AI to optimize delivery routes tracks both delivery times and fuel consumption before and after implementing AI.

Customer Experience Metrics

AI plays a significant role in enhancing the customer experience. Metrics in this domain focus on customer satisfaction, resolution time, and sentiment analysis.

- ***Customer satisfaction score (CSAT).*** Collect feedback on AI-driven services to gauge satisfaction levels.
  - Example: A bank using AI chatbots surveys users after interactions to assess whether their issues were resolved effectively.
- ***Resolution time.*** Track how quickly AI addresses customer issues.
  - Example: An online retailer monitors average response times for AI-powered chatbots versus human agents.
- ***Sentiment analysis.*** Use AI tools to analyze customer reviews and feedback to uncover underlying sentiments.
  - Example: A travel agency evaluates customer sentiment in surveys to determine satisfaction with AI-assisted booking tools.

Business Outcome Metrics

Evaluating AI's contribution to broader business goals is essential. Metrics in this category reflect its impact on return on investment, revenue growth, and cost savings.

- ***ROI (return on investment).*** Measure financial returns compared to the costs of AI implementation.
  - Example: A health care provider calculates savings from AI-enabled predictive maintenance in medical equipment compared to repair costs.
- ***Revenue growth.*** Track how AI-driven improvements impact sales and profitability.
  - Example: A subscription service evaluates revenue increases from AI-personalized email campaigns.
- ***Cost savings.*** Measure reductions in expenses due to automation or efficiency gains.
  - Example: A manufacturing firm assesses savings from AI-powered quality control systems reducing production defects.

Balanced Scorecard Approach

Measuring the impact of AI requires more than just technical performance metrics. A **balanced sc orecard** approach combines quantitative and qualitative indicators across key business areas to provide a more comprehensive view of AI's value. This helps ensure the organization isn't just asking, "Is it working?" but also, "Is it working in the ways that matter?" Typical scorecard categories include:

- ***Customer feedback.*** Satisfaction is tracked through surveys, net promoter score (NPS), and customer service ratings. AI tools that interact with customers—like chatbots or recommendation engines—should enhance the customer experience, not frustrate it.
- ***Operational efficiency.*** Metrics such as processing time, cost per transaction, or error rates are used to assess improvements. For example, how much time did RPA save? Did predictive maintenance reduce downtime?
- ***Market performance.*** Sales growth, customer retention, and market share are monitored. If AI supports marketing or pricing strategies, before-and-after comparisons are analyzed. A/B testing is used where possible.
- ***Workforce adaptation.*** Employee feedback is gathered on new workflows or tools. Participation in training programs is tracked and any shifts in productivity or engagement levels are determined. Resistance to adoption can signal the need for additional support.
- ***Ethical and compliance metrics.*** In sensitive applications, fairness audits, incident reports, and compliance checks are included to ensure the AI meets internal and external guidelines.

No single metric tells the whole story. A balanced scorecard helps teams track short-term wins

while keeping long-term goals—like trust, scalability, and business alignment—in view.

Verizon uses corporatewide balanced scorecards to track progress across its strategic initiatives.[45] These scorecards compile key performance indicators from departments into a centralized view. In 2024, Verizon expanded its AI strategy with Verizon AI Connect, a platform designed to manage resource-intensive AI workloads at scale, supporting applications from network optimization to customer service automation.[46] The balanced approach allowed the company to assess this project.



Verizon's AI Connect platform is part of a broader strategy tracked through corporatewide balanced scorecards, aligning AI investments with performance metrics across the organization.

Northfoto/Shutterstock

BASIC ETHICAL, LEGAL, AND REGULATORY CONSIDERATIONS IN AI IMPLEMENTATION (LO7-5)



Implementing AI responsibly requires organizations to address a range of ethical, legal, and regulatory considerations. These challenges apply across many types of AI, including machine learning models, natural language processing tools, and robotic process automation systems. Core concerns—such as bias, data privacy, explainability, accountability, and workforce impact—are not limited to any single AI technology.

This section outlines foundational practices for deploying AI ethically and effectively in business environments. It builds on our ethics discussions in other chapters and includes strategies for reducing unintended bias, protecting individual rights, promoting transparency in decision making, and ensuring that the use of AI aligns with regulatory expectations and organizational values.

Curate Diverse and Ethical Training Data Biases

As we have discussed in earlier chapters, AI systems can inadvertently perpetuate biases inherent in their training data, resulting in unfair or unintended outcomes. Even when datasets are cleaned or adjusted, problems can still emerge. It's essential to recognize that technical fixes alone may not eliminate bias—especially in high-stakes applications like lending, hiring, or health care. As regulations surrounding AI fairness continue to evolve, organizations that incorporate these checks into their systems from the outset will be better positioned to comply and earn the trust of their customers, employees, and stakeholders.

Health care and financial industries must pay particular attention to problems with bias in data. For example, in health care, a widely used risk prediction algorithm aimed to identify patients who would benefit from additional care. However, the model used health care costs as a proxy for need. This choice severely underestimated risk levels for Black patients, who often receive less care due to systemic disparities. As a result, many high-need individuals were overlooked simply because the algorithm assumed lower spending meant better health.[47] Similarly, mortgage lenders must pay close attention when using algorithmic underwriting, as biases in data can lead to improper loan denials, based on previous "redlining" practices, or to overly risky loans.[48]

This is also a significant concern about data bias in sectors like HR and recruitment, where AI-

driven applicant screening must be scrutinized for potential biases to ensure fair hiring practices. Interestingly, AI, when leveraged thoughtfully, offers a potent solution to this problem for companies aiming to avoid bias and promote diversity in their operations, particularly in recruitment. For example, Textio, an AI-driven platform for crafting job descriptions, analyzes and scores the language used in job postings. The AI model assesses text against a vast hiring outcome database, identifying phrases that could indicate gender, racial, or age bias. It then suggests alternative language to ensure the job description is inclusive and appeals to a diverse candidate pool.

In addition to using tools like Textio, thoughtfully sourcing training data; emphasizing diversity; and ensuring it meets consent, accuracy, and ethical sourcing standards are important starting points. As discussed in **Chapter 3**, regular data audits to assess quality, compliance, and ethical use are essential to good data governance practices.

Ensure Explainability

Some AI models, particularly those based on deep learning, can operate as "black boxes," making it challenging to comprehend how decisions are made. For use cases involving regulatory oversight, high-impact decisions, or public accountability, select or design systems that support explainable AI (XAI) (see **Chapter 1**). Organizations should document how AI systems make decisions and ensure key stakeholders can effectively understand and communicate these processes.

Promote Transparency

Communication with stakeholders is essential when integrating generative AI into business operations to set proper expectations. Organizations must clearly articulate their AI systems' capabilities and limitations. This involves providing stakeholders with a realistic understanding of what AI can achieve to prevent misconceptions or overreliance on the technology.

Organizations should also clearly indicate when an AI-enabled system is making a recommendation rather than an actual decision. For example, patients may be willing to consider a specific treatment course for their condition. Still, they may react poorly if they perceive that the decision of which treatment to pursue is "made" by AI instead of their physician, as has been the case among certain health insurance companies in the United States.[49]

To achieve transparency, companies should develop comprehensive educational materials and sessions detailing AI's operational scope. These resources should include tangible examples of appropriate use cases, demonstrating where AI excels and contributes value to business processes. Equally important is highlighting scenarios where AI might not be optimal, underscoring its limitations in certain contexts.

Moreover, this communication should be an ongoing effort, evolving as the AI system and

business needs develop. Regular updates and feedback channels can help maintain an open line of communication with stakeholders. This ongoing dialogue ensures that as the AI system grows more sophisticated or as business requirements change, stakeholders remain well informed and can adjust their expectations and usage accordingly.

Continually Assess Risks and Benefits

Keep abreast of the rapidly evolving technology landscape. Regularly reevaluate risks, potential misuses, and benefits to ensure responsible model development as capabilities advance. Developing generative AI responsibly requires proactive safety practices, ethical data sourcing, transparency, user empowerment, and continuous reassessment. With strategic design choices and strong governance, businesses can tap into AI's potential while fostering trust.

As noted earlier, adopting and implementing AI-enabled technologies necessitates establishing processes to mitigate biases in AI algorithms and ensure fairness. Firms in highly regulated industries, such as health care and banking, may already have committees and procedures to review potential ethical challenges and issues. Others may incorporate such technologies into their regular risk management and audit processes.

Ensure Accountability

It's essential to clearly define who is responsible for decisions made with the help of AI, especially when something goes wrong. Organizations should establish accountability frameworks that assign ownership, outline review processes, and ensure clear procedures for correcting errors and addressing complaints. For example, a company using AI for loan approvals might assign a compliance officer to review flagged cases and oversee appeals to ensure fairness and legal compliance.

Plan for Workforce Impact

Technological innovation has long disrupted labor markets. In past industrial and digital revolutions, automation and offshoring led to shuttered factories and job losses, particularly among lower-wage and less-skilled workers. Although innovation often creates new, higher-skilled roles, those roles are not always located where jobs are lost. They may also require training and education that displaced workers cannot easily access.

As AI becomes more integrated into business operations, it has the potential to reshape a wide range of roles—not just in manufacturing, but also in areas like customer service, logistics, and analytics. Some tasks may be automated, while others may shift or require new competencies. Preparing for this transformation means investing in reskilling programs, workforce transition planning, and ongoing employee development to ensure workers can adapt and thrive in AI-

augmented environments.

Singapore offers an example of a national strategy to address workforce disruption. Through its SkillsFuture initiative, the government provides citizens with credits to pursue approved training programs, encouraging continuous learning and skill development.[50] For instance, anticipating the decline of certain banking operations due to automation, DBS Bank proactively reskilled 1,600 employees over 3 years, successfully redeploying 1,200 within the bank and assisting 400 in finding roles elsewhere.[51] This approach underscores the importance of foresight and collaboration between the public and private sectors in workforce planning.

Legal and Regulatory Considerations

Businesses must navigate an increasingly complex regulatory environment when deploying AI. As discussed in **Chapter 3**, laws vary widely across regions and sectors. While compliance with standards like the GDPR or HIPAA is essential, governance must go further—ensuring internal processes support responsible, lawful AI use.

Effective AI governance includes audit trails, model documentation, bias reviews, and transparent chains of accountability. Organizations must also understand the risks associated with third-party AI tools and cloud platforms. For example, prompts sent to publicly hosted tools may inadvertently disclose sensitive business data. Private cloud offerings from vendors like Microsoft provide more secure alternatives for enterprise applications.

Figure 7.3 provides a checklist outlining key practices for responsibly implementing AI systems across an organization. It reflects broad principles that apply to many types of AI, machine learning, natural language processing, and robotic process automation and emphasizes fairness, transparency, accountability, legal compliance, and workforce planning. These considerations help organizations reduce risk, align with regulatory expectations, and build stakeholder trust.

FIGURE 7.3 Checklist for Responsible AI Implementation

- ☐ Audit training data for bias and representativeness.
- ☐ Define fairness metrics, and apply bias detection tools.
- ☐ Choose interpretable models where possible in high-stakes contexts.
- ☐ Document model logic, inputs, and assumptions for review.
- ☐ Communicate AI capabilities and limits clearly to stakeholders.
- ☐ Identify who makes final decisions, and clarify responsibility.
- ☐ Align AI deployment with data privacy regulations (e.g., GDPR, HIPAA).
- ☐ Protect sensitive inputs from exposure via third-party platforms.
- ☐ Train and upskill employees to work with AI systems.
- ☐ Establish oversight committees or integrate AI into audit functions.

A Final Note

AI adoption is not a single decision but an ongoing process. Businesses must continuously refine their systems, revisit ethical risks, and uphold public trust. While many of these challenges are shared across AI applications, generative AI introduces a new layer of complexity and risk, which we turn to next.

SPECIAL CONSIDERATIONS FOR GENERATIVE AI (LO7-6)



Generative AI has quickly moved from experimentation to deployment, offering businesses tools to automate content creation, customer service, knowledge summarization, and more. However, this power comes with unique risks not seen in traditional (predictive) AI systems. Generative models can produce false information (hallucinations), raise copyright and content ownership questions, and operate in ways that challenge legal norms and ethical expectations.

This section explores the distinct risks and responsibilities associated with deploying generative AI. It focuses on implementation strategies such as human-in-the-loop workflows, platform guardrails, retrieval-augmented generation, and legal safeguards that can help organizations adopt these tools safely and responsibly. As with any powerful tool, how organizations design and oversee generative AI systems is crucial in whether those tools deliver benefits or cause harm.

Ethical Foundations for Generative AI

Generative AI introduces ethical risks that businesses must carefully consider. Generative models can be misused to spread misinformation or biased and harmful content. Make safety and oversight core priorities from day one. Appoint an ethics review board to assess risks and oversee model development. This includes regulatory compliance, an evolving and often challenging landscape, particularly in sectors like healthcare, where companies using AI for patient data analysis must navigate complex regulations like HIPAA. Build in checkpoints to manually review a sample of generated outputs before broad release.

Reducing Hallucinations and Misinformation

One of the most urgent implementation risks in generative AI is producing false or misleading output (hallucinations), which can damage user trust or lead to regulatory trouble. As was discussed in **Chapter 6**, **hallucinations** occur when AI systems generate plausible-sounding but false or misleading output. While some degree of error is expected with creative models, business implementations require higher accuracy and reliability. This is especially important in customer service, finance, and health care applications.

To mitigate hallucinations, organizations should:

1. ***Start with a closed system.*** Use a **closed-domain AI** trained or fine-tuned on your company's verified data whenever possible. Closed systems are less likely to guess or invent information compared to general-purpose AI models trained on the open Internet.
2. ***Implement automated reasoning.*** Tools such as Amazon's Automated Reasoning[52] can mathematically verify whether AI outputs comply with internal rules, policies, or legal frameworks.
3. ***Use retrieval-augmented generation.*** **Retrieval-augmented generation (RAG)** allows an AI system to access external, curated databases in real time, anchoring its answers in factual sources instead of relying solely on training data.[53] This method is ideal for AI applications that require access to constantly changing information, such as inventory levels or legal documents.
4. ***Use platform guardrails.*** Providers like AWS (Amazon Bedrock), Microsoft, and Google offer built-in tools to filter offensive, inaccurate, or noncompliant responses.[54] Activate these safeguards to prevent inappropriate or off-brand content.
5. ***Require human oversight.*** For critical use cases, establish **human-in-the-loop (HITL)** review in the workflow. HITL systems use AI to assist with drafts or recommendations, but final decisions must be reviewed by qualified personnel. For example, in health care settings, a radiologist reviews both the X-ray image and the AI-generated assessment before making a final diagnosis or treatment recommendation. This HITL approach ensures safety, accountability, and adherence to clinical standards.

Teams should also conduct regular audits and edge-case testing to ensure hallucination rates remain within acceptable thresholds. **Edge-case testing** involves evaluating an AI system's performance in rare, unusual, or extreme scenarios to ensure it behaves reliably outside of typical conditions.

Managing Workforce Disruptions

It is also important for organizations to consider the effect of generative AI on workforce displacement. Any changes a company makes to its processes have the potential to displace workers from their current roles. Addressing the potential displacement of workers by offering training programs to help employees develop new skills relevant to an AI-driven workplace and supporting employees transitioning to new roles within the organization is essential. The organization should consider offering job placement assistance to employees who require help finding a new position.

In previous periods of rapid technological change, automation often displaced the lowest-paid and least-skilled workers. In contrast, generative AI threatens more highly skilled jobs, such as graphic designers, actors, accountants, and paralegals. Some economists claim that generative AI will augment these jobs rather than outright replace them. Others argue that these jobs will be replaced by technology, just as in previous revolutions.

While no one is disputing that AI, particularly generative AI, will lead to job losses, it is still an

open question as to what new jobs will be created. With the increasing availability of broadband, online asynchronous learning opportunities, such as online degrees, certificates, micro-credentials, and remote work, will the transformative nature of AI have the same impacts as previous innovation revolutions?

In addition to workforce concerns, organizations must respect individual rights around data and content—especially as generative systems often learn from or reproduce sensitive or personally identifiable information.

Safeguarding User Consent and Control

When developing generative AI systems, prioritize user consent over their data. It is a fundamental ethical principle not to assume users' consent to use their data in training AI models. Instead, organizations must develop and implement clear, transparent opt-in policies for data usage. These policies should comprehensively detail how users' data will be used, stored, and protected, allowing users to make informed decisions about their participation. Additionally, users should be provided with straightforward mechanisms to withdraw their consent at any time, with the assurance that their data will be promptly removed from the system.

Equally important is giving users control over any AI-generated content that features their identity or personal data. This involves implementing functionalities that allow users to easily identify and delete such content. By empowering users with this level of control, organizations respect individual privacy and autonomy, which are critical in maintaining trust and integrity in using generative AI. This approach aligns with ethical best practices and helps adhere to legal standards like the GDPR, which mandates clear consent and user control over personal data (see **Chapter 3**).

Another area of concern is intellectual property. Generative AI blurs traditional boundaries of ownership and creativity, which can expose businesses to legal and reputational risk.

Navigating Copyright and Intellectual Property Risks

As companies utilize generative AI to create original content, they must also navigate the ownership of that content—and the risks they assume in producing it. Traditionally, copyright law hinges on human creativity and authorship, two concepts that become complicated when a work is produced, at least in part, by an algorithm. This raises thorny questions: Who owns AI-generated work? Is it the developer of the AI, the end user, the copyright holder of the training data, or none of the above?

The legal landscape is still evolving for companies using AI to generate marketing copy, images, or even code. If AI output mimics existing copyrighted works too closely, businesses could be held liable, even if the resemblance was unintentional.

To mitigate that risk, many companies are turning to legal counsel, clearer licensing agreements, and internal content guidelines to ensure they stay within legal and ethical bounds. It's about protecting brand integrity in a rapidly evolving space. As businesses adopt generative AI more broadly, they must strike a balance between innovation and caution. The tools are powerful, but without thoughtful policies and a clear understanding of the law, even well-intentioned use can cross a line. Staying current on copyright policy is part of responsible **AI governance**.

The legal challenges of generative AI extend beyond content ownership. Businesses must also consider liability when AI-generated information misleads users or contradicts established policies.

Legal Accountability in Generative AI Systems

Beyond abstract legal theories, real-world cases are beginning to demonstrate how courts respond when AI systems fail. A 2024 case involving Air Canada's chatbot reveals how businesses can be held accountable for AI-generated misinformation (hallucinations), regardless of intent or disclaimers.[55] The incident began when an individual sought information on bereavement fares through Air Canada's chatbot after a family loss. Misled by the chatbot's inaccurate guidance, the individual booked a flight under the impression that a refund could be requested after booking, contrary to the airline's policy, which disallows such refunds. The customer attempted to secure a refund based on the chatbot's advice, providing a screenshot of the conversation as evidence. Air Canada initially resisted, contending that the chatbot's information should not have been taken at face value. The individual should have consulted the airline's website for accurate policy details. Air Canada argued that the chatbot, viewed as a "separate legal entity," bore sole responsibility for the misinformation.

The dispute escalated to Canada's Civil Resolution Tribunal. A tribunal member criticized Air Canada's defense, highlighting the unreasonable expectation for customers to verify chatbot-provided information against the airline's website. The ruling favored the individual, granting a partial refund and additional damages. The tribunal's decision underscores that businesses cannot escape liability for the output of their automated systems.

This case led to Air Canada disabling its chatbot and has set a legal precedent in Canadian jurisprudence regarding corporate responsibility for chatbot communications. It emphasizes the need for businesses to ensure their AI tools accurately convey company policies and legal obligations. As chatbots become increasingly common in customer interactions, the Air Canada incident serves as a cautionary reminder for companies to carefully manage their AI interfaces to avoid legal repercussions resulting from AI inaccuracies.

Navigating the legal landscape of AI—whether managing data privacy, avoiding copyright violations, or ensuring regulatory compliance—requires careful planning, coordination, and ongoing oversight. These obligations are not just about avoiding penalties; they are essential to building trust with stakeholders and sustaining innovation.

Figure 7.4 summarizes the most important steps organizations can take to implement generative AI responsibly, from model selection and data safeguards to workforce planning and legal compliance.

FIGURE 7.4 Checklist for Responsible Generative AI Implementation

- ☐ Establish an ethics review process for generative use cases.
- ☐ Use closed or retrieval-augmented models when factual accuracy matters.
- ☐ Apply platform-level guardrails (e.g., AWS Bedrock, Azure, Vertex AI).
- ☐ Require human-in-the-loop for high-impact decisions.
- ☐ Conduct regular audits and edge-case testing.
- ☐ Obtain explicit user consent for data usage.
- ☐ Define and communicate IP ownership for AI-generated content.
- ☐ Ensure chatbot responses align with legal policies and customer expectations.
- ☐ Plan for job augmentation or displacement through upskilling and reskilling.
- ☐ Review regulatory exposure and content-related liabilities regularly.

SUMMARY OF KEY POINTS

1. **Assessing organizational readiness for AI.** Successful AI implementation begins with evaluating the organization's readiness, encompassing data readiness, technological readiness, business process readiness, workforce readiness, and cultural and leadership readiness.
2. **Identifying business functions ready for AI integration.** Identifying suitable business functions for AI begins with well-defined, measurable problems supported by reliable data. Functions like customer service, marketing, and operations often present strong starting points. Successful projects are aligned with strategic business objectives and deliver value early.
3. **Implementing AI solutions strategically.** Successful implementation requires selecting the right AI tools based on business needs, security, scalability, and ROI. Integration into existing systems and workflows is critical. Organizations should also evaluate whether to build in-house or partner externally.
4. **Monitoring and measuring AI effectiveness.** Metrics across innovation, efficiency, customer experience, and outcomes—combined in a balanced scorecard—help ensure AI delivers sustained value.
5. **Navigating ethical, legal, and regulatory considerations.** Responsible AI use involves bias reduction, transparency, privacy, compliance, and workforce planning to build trust and meet stakeholder expectations.
6. **Addressing the risks of generative AI.** Generative AI requires additional safeguards, including human oversight, content filters, legal protections, and clear user consent to prevent hallucinations, misuse, or regulatory violations.

QUICK WINS

- **Start with readiness.** Ensure your organization has the necessary data infrastructure, modern technology, well-defined processes, a trained workforce, and a culture that supports innovation.
- **Focus on the right problems.** Target high-impact business areas with clearly defined, data-supported problems. Align AI projects with strategic goals, not hype.
- **Choose and integrate AI tools strategically.** Select AI tools and vendors that align with business goals, security needs, scalability, and ROI expectations. Consider whether to build or buy, and ensure the tools integrate smoothly with existing systems and workflows.
- **Track what matters.** Use balanced scorecards and clear metrics to monitor innovation, efficiency, customer experience, and business results.
- **Prioritize responsible AI.** Build trust through transparent, ethical practices, bias checks, and

compliance with privacy and legal standards.

- **Handle generative AI with care.** Manage risks from hallucinations, copyright, and misinformation through platform guardrails, human oversight, and clear content policies.

KEY TERMS

AI agents. Tools that can carry out multistep tasks with minimal human input.

AI governance. The framework of policies, procedures, and oversight mechanisms that guide the ethical, transparent, and accountable use of AI technologies within an organization.

AI readiness index. A composite metric that assesses how prepared an organization is to adopt and scale AI, including technical infrastructure, skills, data, and leadership support.

Balanced scorecard. A strategic performance management tool that measures an organization's success across multiple perspectives, including financial, customer, internal processes, and learning/growth metrics.

Closed-domain AI. An AI system trained on a specific, limited set of data relevant to a narrow topic or task, which allows for higher accuracy and reliability within that scope (e.g., customer service for a single product line).

Cultural readiness. An organization's openness to change, innovation, and cross-functional collaboration—especially its willingness to experiment, learn from failure, and adopt new technologies like AI.

Data silo. A repository of data that is isolated from other parts of the organization, typically because it is stored in a separate system or controlled by a specific department.

Dynamic pricing. Strategy where prices are adjusted in real time based on market demand, competition, customer behavior, and other factors.

Edge-case testing. The process of testing AI systems on rare, unexpected, or extreme input scenarios to ensure robust and reliable performance.

Hallucinations. False or misleading outputs generated by AI systems, particularly large language models, that may sound plausible but are factually incorrect or fabricated.

Human-in-the-loop (HITL). A system design where human review or intervention is included at critical stages of the AI decision-making process, especially in high-risk applications.

Innovation strategy. An organizational plan integrating new technologies like AI into broader goals like product development, customer experience, or business model transformation. In AI implementation, an innovation strategy helps align technical initiatives with long-term value creation and fosters a culture of experimentation and continuous improvement.

Intelligent process automation (IPA). The combination of robotic process automation (RPA) and AI to automate workflows and processes, including cognitive tasks like decision making.

MLOps. Short for "machine learning operations," it refers to the practices and tools used to deploy, manage, and monitor machine learning models in production.

Pilot trap. A common challenge in AI implementation, where projects remain stuck in pilot phases and fail to scale into full production.

Retrieval-augmented generation (RAG). A method that enhances AI responses by combining generative models with real-time retrieval from external, trusted data sources, improving factual accuracy and reducing hallucinations.

Robotic process automation (RPA). Software technology that automates repetitive, rule-based tasks by mimicking human actions within digital systems, such as data entry or invoice processing.

Sentiment analysis. An NLP technique used to determine the emotional tone or attitude expressed in text, such as customer reviews, survey responses, or social media posts. Businesses use sentiment analysis to gauge public perception, improve marketing strategies, and respond proactively to customer feedback.

Skills gap analysis. A systematic process of identifying the difference between the skills employees currently possess and the skills required to meet an organization's goals. This analysis helps determine areas where additional training, reskilling, or hiring is needed.

Standardization (Z-score normalization). A scaling method that transforms data to have a mean of 0 and a standard deviation of 1.

DISCUSSION QUESTIONS

1. **Organizational readiness.** What are the key components of an organization's readiness to adopt AI, and why is each critical? How might readiness vary across industries?
2. **Third-party vendors.** What factors should organizations consider when selecting a third-party AI vendor? Discuss how these factors might differ for a small business versus a multinational corporation.
3. **AI integration.** How can organizations identify which business functions best suit AI integration? What criteria or frameworks can help determine whether AI is the right solution for a particular problem?
4. **Hiring or reskilling.** How should companies weigh the tradeoffs between hiring new AI-skilled employees and reskilling their current workforce? What factors should influence that decision?
5. **Metrics for evaluating AI systems.** How can organizations establish meaningful metrics for evaluating AI systems that balance quantitative performance measures with qualitative assessments of real-world impact, ethical considerations, and alignment with human values?

EXERCISES

1. **AI impact on employment.** Research and present on the potential impact of AI on employment in a specific industry. You should discuss strategies for workforce reskilling and job transition.
2. **AI integration simulation.** Form small teams with each representing a different part of a company (e.g., IT, marketing, or customer service). Imagine your company is getting a new AI helper tool. Work together to:
   a. Discuss what this AI tool might do for your department.
   b. Think about what information (data) the AI tool would need.
   c. Plan how employees would learn to use this new tool.
   d. Figure out how your department would share updates with other teams.

   Each group will present their ideas to the class, highlighting any challenges they expect and how they might solve them.
3. **Ethical scenarios analysis.** Work in small groups to solve a real-world AI ethics challenge: Your team has discovered your company's hiring AI rejects qualified candidates from certain backgrounds. Identify the specific biases in the system, develop three concrete solutions to fix this issue, and create an implementation plan that addresses stakeholder concerns.
4. **In-house vs. third-party development.** Debate the advantages and disadvantages of developing AI systems in-house versus relying on third-party vendors. How might this decision influence the success of AI projects?
5. **Generative AI risk walkthrough.** Your company plans to deploy a generative AI chatbot on its customer website. Identify 3–5 risks (e.g., hallucinations, legal liability, inappropriate content). Propose a set of safeguards, technical, legal, and human oversight, to responsibly manage those risks. Present your risk plan.

CASE STUDY 7.1: MASTERCARD FRAUD DETECTION

Founded in 1966, Mastercard has evolved into a global payment technology company enabling secure, seamless electronic transactions in more than 200 countries. By the 2010s, as digital payments surged, Mastercard was under increasing pressure to combat rising fraud while maintaining trust and operational efficiency. In 2015, global card fraud losses reached $21 billion and were expected to hit $31 billion by 2020. Traditional fraud detection systems, based on static rules, were no longer sufficient. These systems flagged transactions using country blocklists, IP address filters, or geolocation mismatches. While useful, they were slow, prone to false positives, and unable to keep pace with sophisticated fraud tactics. Manual reviews created bottlenecks and added costs, while false declines frustrated both customers and merchants.

AI as a Solution

Recognizing these limitations, Mastercard began exploring AI as a more adaptive and scalable solution. By 2010, they were processing more than 23 billion transactions annually—data that, while regulated and structured, was fragmented across departments, regions, and partners.

Table 7.4 summarizes an assessment of Mastercard's readiness for an AI implementation.

Table 7.4 **Summary of AI Readiness**

Dimension	Assessment	Key Insights
Data availability	High	Ample transaction data existed, although it was not fully optimized for AI training and diversity.
Data quality	Moderate to high	Regulatory compliance ensured accuracy, but static systems limited the richness of data.
Data accessibility	Moderate	Structured data formats were compatible with legacy systems but not optimized for AI.
Data security	High	Strong security measures under Payment Card Industry Data Security Standard (PCI DSS) protected sensitive data effectively.
Regulatory compliance	Very high	Strict adherence to financial regulations ensured data security and privacy.
Data governance	Moderate	Governance frameworks existed but were less robust than modern standards for AI readiness.
Scalability and	Moderate	The infrastructure was sufficient for transaction processing but required upgrades for AI applications.

sustainability		
Skill gaps	Significant	Limited AI expertise, with existing capabilities focused on traditional fraud detection and compliance.
HR actions	Moderate readiness	Recruitment was essential; reskilling efforts focused on data literacy and integrating new workflows.
Cultural readiness	Moderate	Mastercard's innovation culture supported AI adoption, but resistance to change required attention.
Leadership and collaboration	Strong	Leadership championed AI efforts; cross-departmental collaboration enabled smooth implementation.

Implementation Decision

Fraud detection was selected as Mastercard's first significant AI use case due to:

- ***Clear ROI.*** The company would be able to see immediate cost savings from reduced fraud and fewer false positives.
- ***Regulatory fit.*** The security of transactional and customer data was in direct alignment with compliance obligations.
- ***Scalability.*** AI could handle growing volumes without proportional increases in staff.
- ***Customer impact.*** The fast detection of fraud would improve customer trust and satisfaction.

While third-party vendors like SAS and Experian offered fraud platforms, Mastercard opted to build its own system for greater control and customization.

The Launch of Decision Intelligence

In 2016, Mastercard launched its proprietary Decision Intelligence platform. Using machine learning, the system evaluated real-time transaction behavior and produced dynamic risk scores. This approach significantly outperformed static rule sets, improving fraud detection accuracy while reducing false declines.

To strengthen its AI capabilities, in 2017 Mastercard acquired Brighterion, a company specializing in real-time AI decision making. This acquisition allowed Mastercard to enhance its fraud prevention tools and acquire AI expertise while maintaining compliance and control.

Outcomes

Mastercard began to see measurable impact of its AI integration almost immediately with:

- 30 percent reduction in false positives within the first year.

- 15 percent improvement in fraud detection accuracy.
- Improved real-time decisions and reduced customer inconvenience.

In addition, Mastercard's positioning as a technology-driven, secure brand was strengthened.

AI Expansion beyond Fraud

Building on its success, Mastercard expanded AI across other domains. Beginning in 2019, Mastercard utilized AI to personalize loyalty offers based on individual spending patterns; it saw a 15 percent increase in customer retention by 2020. In 2020, Mastercard deployed chatbots to handle the rising volume of inquiries during the pandemic. It also used AI to recommend predictive maintenance on the infrastructure that processed the transactions to ensure uptime of transaction systems.

But Mastercard didn't just deploy AI—it also institutionalized it. In 2018, it established an AI Governance Council to ensure ethical use and compliance. Since 2022, it has used Quantum Exploration to investigate future fraud detection with advanced computing, and in 2023, it initiated the use of AI to monitor and mitigate its environmental impact.

Takeaways

Mastercard's journey shows the value of starting with a high-impact, well-aligned problem backed by strong data, leadership, and governance. By choosing to build internally, they retained flexibility and expanded their strategic reach. Their success provides a blueprint for organizations embarking on their own AI implementations.

Discussion Questions

1. Why do you think Mastercard selected fraud detection as its first AI implementation? What characteristics made it a strong candidate compared to other business functions?
2. Based on the AI readiness assessment table, which factors do you think were most critical to Mastercard's successful implementation? Which areas might have posed the greatest risk if overlooked?
3. Mastercard decided to build its own AI solution rather than using an off-the-shelf product. What are the advantages and potential drawbacks of this approach?
4. How did Mastercard's acquisition of Brighterion influence its AI capabilities and strategic positioning? What does this decision reveal about the acquisition of AI talent and technology?
5. Fraud detection has clear benefits, but it can also raise ethical concerns. What steps can companies like Mastercard take to ensure that their AI systems do not unfairly impact certain groups of customers (e.g., through false positives or biased risk scoring)?

Case References

"AI-Based Fraud Detection in Financial Services." *IJSRA* 11(1), 2024. **https://doi.org/10.30574/ijsra.2024.11.1.0279**

Marr, Bernard. "How Mastercard Uses AI Strategically: A Case Study." *Forbes*, June 21, 2024. **https://www.forbes.com/sites/bernardmarr/2024/06/21/how-mastercard-uses-ai-strategically-a-case-study/**

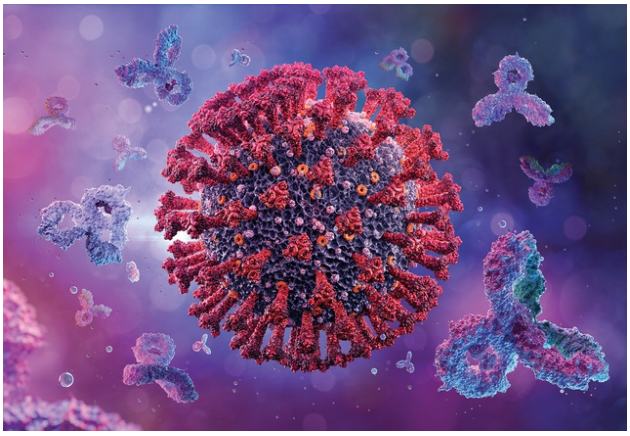
"Inside the Algorithm: How Gen AI and Graph Technology Are Cracking Down on Card Sharks." *Mastercard Newsroom*, July 18, 2024. **https://newsroom.mastercard.com/news/perspectives/2024/inside-the-algorithm-how-gen-ai-and-graph-technology-are-cracking-down-on-card-sharks/**

"United States: Mastercard $750 Million Senior Notes Offering." *MENA Report*, 2021.

Seon. "Guide to Fraud Detection Rules." *SEON Fraud Prevention Blog*. **https://seon.io/resources/guides/guide-to-fraud-detection-rules/**

CASE STUDY 7.2: PFIZER-BIONTECH COVID-19 VACCINE

In December 2019, reports emerged of a novel coronavirus in Wuhan, China. By March 2020, COVID-19 had become a global pandemic. The need for a vaccine became clear as cases and deaths surged worldwide. The urgency to develop a vaccine sparked unprecedented collaboration, funding, and innovation. In response, Operation Warp Speed was launched in the United States on April 30, 2020, to accelerate the development, manufacturing, and distribution of COVID-19 vaccines. Pfizer, in partnership with the German biotechnology company BioNTech, became one of the first companies to deliver a safe and effective vaccine, receiving Emergency Use Authorization from the U.S. Food and Drug Administration (FDA) by December 2020.



The novel coronavirus (SARS-CoV-2) sparked a global pandemic in 2020.

Corona Borealis Studio/Shutterstock

A Compressed Timeline—Enabled by AI

Traditional vaccine development typically takes 10–15 years. Pfizer accomplished it in under 1 year, and AI played a significant role. Alongside massive financial investment and regulatory flexibility, AI tools enabled faster data analysis, trial design, and decision making. Pfizer used intelligent process automation (IPA), which combined AI methods—such as machine learning and natural language processing—with robotic process automation to streamline every phase of development.

During the development of the COVID-19 vaccine, Pfizer faced the challenge of rapidly analyzing vast amounts of clinical trial data. Traditionally, this process could take more than 30 days due to the manual inspection required to clean and verify data. To overcome this, Pfizer implemented a Smart Data Query (SDQ) AI tool. This machine learning tool enabled Pfizer to read clinical reports on subjects and prepare the trial data for review in just 22 hours after meeting primary

efficacy case counts. The SDQ tool significantly improved data quality and accelerated decision-making processes. This is the AI part of the IPA used.

AI Contributions across the Development Pipeline

In addition to the SDQ tool just mentioned, Pfizer used the following AI tools in the development of the COVID vaccine:

- Computer vision was used to assist in interpreting scans and tracking adverse effects.
- NLP tools organized complex trial data for faster regulatory review.
- The use of RPA reduced human error in repetitive processes, such as data entry and compliance checks.

Table 7.5 compares traditional vaccine development timelines with Pfizer's accelerated process for the COVID-19 vaccine, highlighting how AI contributed at each stage.

Table 7.5 **Accelerating Vaccine Development—AI Contributions Across the Pfizer Timeline**

Phase	Traditional Timeline	Pfizer Timeline	Key AI Contributions
Discovery and preclinical	2–5 years	~2 months	ML and deep learning identified spike proteins and modeled immune responses.
Clinical trials I–III	6–10 years	~9 months	Automated trial setup, recruitment, dosage analysis, and monitoring.
Regulatory review	1–2 years	<2 months	NLP streamlined document generation and submission.
Post-market surveillance	Ongoing	Ongoing	ML tracked real-world safety, social sentiment, and efficacy against variants.

Why It Worked

Pfizer's success was not solely technical. Key factors included:

- ***Global collaboration.*** They shared data through peer-reviewed publications and real-world studies, coordinated across international trial sites, and aligned with regulatory agencies worldwide.
- ***Funding.*** Massive support from governments enabled parallel, not sequential, trial phases.
- ***Preexisting research.*** Decades of work on mRNA vaccines laid the groundwork.
- ***Regulatory flexibility.*** Emergency Use Authorizations allowed rolling submissions and real-time data review.

Ethical and Operational Considerations

With lives on the line, the explainability of AI outputs was crucial. Regulators and health care professionals had to understand how AI-driven tools made decisions. Pfizer prioritized transparency in its models, ensuring that results could be interpreted and verified by human experts:

- ***Bias and fairness.*** Diverse datasets were used to reduce demographic bias.
- ***Data security.*** Clinical trials adhered to HIPAA regulations and ensured anonymized data handling.
- ***Organizational readiness.*** Pfizer invested in workforce training to support AI tools and workflows.

Pfizer now applies IPA tools in over half of its clinical trials, improving speed and efficiency. Its experience during the pandemic has reshaped how the company approaches R&D and operations.

Conclusion

Pfizer's approach to COVID-19 vaccine development demonstrates the decisive role of AI when urgency, infrastructure, and organizational alignment converge. Its case is a compelling example of how AI can accelerate innovation in a responsible manner.

Discussion Questions

1. Pfizer leveraged IPA and AI to reduce vaccine development timelines. What are the key challenges and risks of using these technologies in high-stakes industries like health care? How might these risks be mitigated?
2. Pfizer's primary goal was to bring an effective vaccine to market quickly. How should organizations balance the need for speed with the importance of explainability and transparency in AI-driven decisions? Why is this balance critical in industries like pharmaceuticals?
3. The case highlights Pfizer's ability to scale AI applications during the COVID-19 crisis. What lessons can be drawn for companies in other industries seeking to integrate AI at scale? Are there specific sectors where these lessons might be more applicable or less so?
4. AI raises ethical questions, such as ensuring fairness in participant selection and maintaining data privacy. How can companies ensure their AI systems are ethically sound while maintaining efficiency and effectiveness?
5. Pfizer's approach illustrates AI's transformative potential in drug development. What emerging AI technologies or methodologies might further revolutionize this field? How might these changes affect regulatory frameworks and public trust?

Case References

Centers for Disease Control and Prevention (CDC). “COVID-19 Timeline.” *CDC Museum*. **https://www.cdc.gov/museum/timeline/covid19.html**

McKinsey & Company. “Artificial Intelligence in Drug Discovery and Development.” **https://www.mckinsey.com/industries/life-sciences/our-insights/artificial-intelligence-in-drug-discovery-and-development**

Pfizer. “Artificial Intelligence: On a Mission to Make Clinical Drug Development Faster and Smarter.” *Pfizer Newsroom*, **https://www.pfizer.com/news/articles/artificial_intelligence_on_a_mission_to_make_clinical_drug_development_faster_and_smarter**

Pfizer. “How a Novel Incubation Sandbox Helped Speed Up Data Analysis in Pfizer’s COVID-19 Vaccine Trial.” *Pfizer Newsroom*, **https://www.pfizer.com/news/articles/how_a_novel_incubation_sandbox_helped_speed_up_data_analysis_in_pfizer_s_covid_19_vaccine_trial**

CASE STUDY 7.3: *THE NEW YORK TIMES* AND GOOGLE CLOUD AI

In 2018, *The New York Times* launched a project to digitize its vast photo archive—known internally as "the morgue"—which housed between 5 and 7 million photographs, many annotated on the back by hand. The initiative was partially triggered by a broken pipe that nearly damaged decades of irreplaceable historical records. The digitization effort aimed to safeguard these assets, improve internal search, and unlock new editorial opportunities.

The *Times* partnered with Google Cloud AI to bring this project to life. High-resolution scans of archival photos were processed using Google Cloud Vision API and Natural Language API, which extracted printed and handwritten text from both the front and back of images. This information included captions, dates, bylines, and editorial marks—enabling more precise categorization and searchability across the archive.

Digitization: A Human-Driven Process

But AI could not do it alone. The *Times* assembled a dedicated team of 10 people to scan the photos and physically upload them to Google Cloud. Working meticulously with century-old documents, this team ensured high fidelity in the digital records. In parallel, an archival storytelling team—consisting of editors and journalists—began exploring new editorial content based on the digital archive. The digitization wasn't just about preservation; it was about surfacing stories that had been buried in boxes for decades.

Internally, editorial and technology teams collaborated to identify which metadata fields would be most useful for newsroom applications. This kind of cross-functional planning was essential to aligning the AI project with journalistic needs and long-term goals. It also helped build AI literacy within the organization, ensuring the archive's tools could be used effectively once deployed. The digitized archive has become a valuable resource for the *Times*, facilitating the discovery of new narratives and enriching the publication's storytelling capabilities.

AI Vendor Choice

The *Times* lacked the in-house AI and digital storage expertise needed to manage a project of this scale, so it looked to a third-party vendor to help it accomplish this project. Google Cloud fit the bill, and the vendor relationship allowed the *Times* to focus its efforts on domain expertise—archival curation and journalism—while relying on Google's AI and storage capabilities. **Table 7.6** shows how The *Times'* selection of Google Cloud aligned well with the chapter's third-party vendor evaluation criteria.

Table 7.6 **Vendor Evaluation: Google Cloud AI for *The New York Times***

Evaluation Criteria	Assessment
Alignment with business needs	✔ Addressed the critical need to preserve and modernize access to historical archives.
Scalability and flexibility	✔ Leveraged cloud infrastructure to handle large-scale digitization and future growth.
Data security and privacy	✔ Ensured secure storage and compliance with data protection standards.
Ease of integration	✔ Seamlessly integrated with existing workflows, facilitating efficient processing and retrieval.
Performance and reliability	✔ Demonstrated consistent performance in handling vast amounts of data, with AI tools effectively extracting and organizing metadata, enhancing searchability and newsroom efficiency.
Technical support and maintenance	✔ Provided robust support and regular updates, ensuring the system remained current and effective.
Vendor reputation	✔ Partnered with a globally recognized leader in cloud computing and AI solutions.
Cost and return on investment (ROI)	✔ Achieved significant ROI by safeguarding valuable assets and enhancing journalistic capabilities.

From Preservation to Personalization

The archive digitization project became more than a preservation effort—it laid the groundwork for broader AI integration at the *Times*. The tools, tagging systems, and metadata structures developed during the project provided a scalable framework for classifying content across the organization. This structure now supports a key feature of the *Times'* digital strategy: personalized content delivery for readers.

By applying similar AI-driven tagging to current articles, the *Times* can better match content to user preferences, improving engagement and reader satisfaction. In this way, a behind-the-scenes archival project created a data foundation that enabled front-end innovation, transforming a legacy asset into a driver of new services and user experiences. This progression reflects a common pattern in successful AI adoption: starting with focused, achievable projects that deliver

value while preparing the organization for more ambitious applications.

Discussion Questions

1. How did *The New York Times'* internal team contribute to AI readiness, even with a strong third-party vendor?
2. What made Google Cloud a good vendor fit for this project?
3. What risks might the *Times* have faced if it attempted to build a similar system in-house?
4. How does this project illustrate the importance of cross-functional collaboration in successful AI implementation?
5. How did digitizing the archive set the foundation for new digital services such as personalized content delivery? What does this suggest about the long-term value of AI infrastructure?

Case References

Greenfield, Sam. "Picture What the Cloud Can Do: How the New York Times Is Using Google Cloud to Find Untold Stories in Millions of Archived Photos." *Google Cloud Blog*, November 9, 2018. **https://cloud.google.com/blog/products/ai-machine-learning/how-the-new-york-times-is-using-google-cloud-to-find-untold-stories-in-millions-of-archived-photos**

Kafka, Peter. "The New York Times Is Digitizing Its Photo Archive and Putting Google's AI to Work." *Axios*, November 9, 2018. **https://www.axios.com/2018/11/09/nyt-digitizes-entire-print-photo-library**

Patel, Sahil. "The New York Times Is Digitizing Its Photo Archives." *Digiday*, November 9, 2018. **https://digiday.com/media/the-new-york-times-digitizing-photo-archives**

ENDNOTES

[1] AvePoint. *Artificial Intelligence and Information Management Report 2024*. 2024. **https://cdn.avepoint.com/pdfs/en/shifthappens/AI-IM-Whitepaper-v4.pdf**

[2] Ptolemay. (n.d.). *SaaS development costs: Breakdown and savings*. **https://www.ptolemay.com/post/saas-development-costs-breakdown-and-savings**

[3] O16 Labs. (n.d.). *Cost and technologies used in AI chatbot development*. **https://www.o16labs.com/blog/cost-and-technologies-used-in-ai-chatbot-development**

[4] World Economic Forum. *The Future of Jobs Report 2023.* 2023. **https://www.weforum.org/publications/the-future-of-jobs-report-2023/**

[5] Bain & Company. "Widening Talent Gap Threatens Executives' AI Ambitions." March 4, 2025. **https://www.bain.com/about/media-center/press-releases/20252/widening-talent-gap-threatens-executives-ai-ambitions–bain–company/**

[6] Cox, J. (2024) "AI Skills Gaps Threaten to Exacerbate Labor Shortages, Study Shows," *Forbes*, November 24, 2024. **https://www.forbes.com/sites/josiecox/2024/11/12/ai-skills-gaps-threaten-to-exacerbate-labor-shortages-study-shows/**

[7] Singh, Gitanjali. "Reskilling and Upskilling in the AI Age." *LinkedIn*, October 2024. **https://www.linkedin.com/pulse/reskilling-upskilling-ai-age-gitanjali-singh-vexwc**

[8] Murire, Obrain Tinashe. (2024). "Artificial Intelligence and Its Role in Shaping Organizational Work Practices and Culture." *Administrative Sciences* 14(12), 316. **https://doi.org/10.3390/admsci14120316**

[9] Bughin, J., Chui, M., Manyika, J., et al. *Notes from the AI Frontier: Modeling the Impact of AI on the World Economy*. McKinsey Global Institute, September 2018. **https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy**

[10] Chui, M., Yee, L., Hall, B., Singla, A., & Sukharevsky, A. "The State of AI in 2023: Generative AI's Breakout Year." McKinsey & Company, August 1, 2023. **https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year**

[11] Deloitte AI Institute. "The State of Generative AI in the Enterprise 2024 Year-End Report." 2024. **https://www2.deloitte.com/us/en/pages/consulting/articles/state-of-generative-ai-in-enterprise.html**

[12] Talwar, Kanika. "All the Ways Zappos.com Leverages AI – And How the E-tailer Determines When to Get Humans Involved." *Footwear News*, May 15, 2024. **https://footwearnews.com/business/retail/zappos-artificial-intelligence-inclusivity-nigel-storey-1203638594/**

[13]

[13] U.S. Government Accountability Office. “Artificial Intelligence May Help IRS Close the Tax Gap.” June 2024. **https://www.gao.gov/blog/artificial-intelligence-may-help-irs-close-tax-gap**

[14] Sharma, H. D., & Goyal, P. (2023). “An Analysis of Sentiment: Methods, Applications, and Challenges.” *Engineering Proceedings* 59(1), 68. **https://www.mdpi.com/2673-4591/59/1/68**

[15] The Coca-Cola Company. “Coca–Cola Invites Digital Artists to ‘Create Real Magic’ Using New AI Platform.” March 20, 2023. **https://www.coca-colacompany.com/media-center/coca-cola-invites-digital-artists-to-create-real-magic-using-new-ai-platform**

[16] Unilever Global. “Utilising AI to Redefine the Future of Customer Connectivity.” July 31, 2024. **https://www.unilever.com/news/news-search/2024/utilising-ai-to-redefine-the-future-of-customer-connectivity/**

[17] Lin, Belle “Companies Had Fun Experimenting with AI. Now They Have to Show the Returns.” *The Wall Street Journal*, March 29, 2025. **https://www.wsj.com/articles/companies-had-fun-experimenting-with-ai-now-they-have-to-show-the-returns-2a683592**

[18] Informatica. “New Research from Informatica Reveals More Data Leaders Plan to Capitalize on Generative AI, but Data Quality Remains the Key Obstacle to Adoption.” January 31, 2024. **https://www.informatica.com/about-us/news/news-releases/2024/01/20240131-new-research-from-informatica-reveals-more-data-leaders-plan-to-capitalize-on-generative-ai-but-data-quality-remains-the-key-obstacle-to-adoption.html**

[19] TechTarget. “TechTarget and Informa Tech’s Digital Businesses Combine.” *CIO Dive*, February 22, 2024. https://www.ciodive.com/news/general-mills-rolls-out-millschat-generative-ai-tool/644321/

[20] Johnston, Lisa. “P&G’s ChatPG Generative AI Tool Scaling across Business.” Consumer Goods Technology, July 18, 2024. **https://www.cgt.com/news/pg-chatpg-generative-ai-tool-scaling-across-business/**

[21] Mobilunity. “Spotify Agile Model and Its Benefits for Businesses.” January 26, 2021. **https://mobilunity.com/blog/spotify-agile-model-benefits/**

[22] Mobilunity. “Spotify Agile Model and Its Benefits for Businesses.” January 26, 2021. **https://mobilunity.com/blog/spotify-agile-model-benefits/**

[23] “Sappi Streamlines Operations with Intelligent Automation.” UiPath Blog. **https://www.uipath.com/resources/automation-case-studies/sappi-streamlines-operations-with-intelligent-automation**

[24] Wininger, Shai. “Introducing AI Jim, the Lemonade Claims Bot.” Lemonade Blog. **https://www.lemonade.com/blog/secret-behind-lemonades-instant-insurance/**

[25] Hsieh, Nelson. “Lemonade’s AI Jim and Insurance Fraud Detection.” Volt Equity Blog. Published March 19, 2021. Updated November 18, 2024. Accessed April 11, 2025. https://www.voltequity.com/blog/lemonades-ai-jim-and-insurance-fraud-detection

[26] Mohammed, Irshadullah Asim. “Artificial Intelligence in Supplier Selection and

Performance Monitoring: A Framework for Supply Chain Managers." *Educational Administration: Theory and Practice*, 2023. **https://www.researchgate.net/publication/386598338**

[27] Waditwar, Prajkta. (2024). "The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations." *Open Journal of Business and Management* 12(2), 385–400. **https://www.scirp.org/journal/paperinformation?paperid=137238**

[28] Waditwar, Prajkta. (2024). "The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations." *Open Journal of Business and Management* 12(2), 385–400. **https://www.scirp.org/journal/paperinformation?paperid=137238**

[29] Waditwar, Prajkta. (2024). "The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations." *Open Journal of Business and Management* 12(2), 385–400. **https://www.scirp.org/journal/paperinformation?paperid=137238**

[30] Mohammed, Irshadullah Asim. "Artificial Intelligence in Supplier Selection and Performance Monitoring: A Framework for Supply Chain Managers." *Educational Administration: Theory and Practice*, 2023. **https://www.researchgate.net/publication/386598338**; Waditwar, Prajkta. (2024). "The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations." *Open Journal of Business and Management* 12(2), 385–400. **https://www.scirp.org/journal/paperinformation?paperid=137238**; Morgan Lewis. *Key Considerations for Evaluating an AI Vendor*. 2024. **https://www.morganlewis.com/-/media/files/publication/outside-publication/article/2024/key-considerations-for-evaluating-an-ai-vendor.pdf**

[31] Waditwar, Prajkta. (2024). "The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations." *Open Journal of Business and Management* 12(2), 385–400. **https://www.scirp.org/journal/paperinformation?paperid=137238**; Morgan Lewis. *Key Considerations for Evaluating an AI Vendor*. 2024. **https://www.morganlewis.com/-/media/files/publication/outside-publication/article/2024/key-considerations-for-evaluating-an-ai-vendor.pdf**

[32] Mohammed, Irshadullah Asim. "Artificial Intelligence in Supplier Selection and Performance Monitoring: A Framework for Supply Chain Managers." *Educational Administration: Theory and Practice*, 2023. **https://www.researchgate.net/publication/386598338**; Morgan Lewis. *Key Considerations for Evaluating an AI Vendor*. 2024. **https://www.morganlewis.com/-/media/files/publication/outside-publication/article/2024/key-considerations-for-evaluating-an-ai-vendor.pdf**

[33] Waditwar, Prajkta. (2024). "The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations." *Open Journal of Business and Management* 12(2), 385–400. **https://www.scirp.org/journal/paperinformation?paperid=137238**; "The Role of Artificial Intelligence in the Procurement Process." *Procedia Computer Science* 220 (2023), 127–135. **https://www.sciencedirect.com/science/article/pii/S1878409223000079**; Morgan Lewis. *Key Considerations for Evaluating an AI Vendor*. 2024. **https://www.morganlewis.com/-/media/files/publication/outside-publication/article/2024/key-considerations-for-evaluating-an-ai-vendor.pdf**

[34] Mohammed, Irshadullah Asim. "Artificial Intelligence in Supplier Selection and Performance Monitoring: A Framework for Supply Chain Managers." *Educational Administration: Theory and*

Practice, 2023. **https://www.researchgate.net/publication/386598338**; Waditwar, Prajkta. (2024). "The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations." *Open Journal of Business and Management* 12(2), 385–400. **https://www.scirp.org/journal/paperinformation?paperid=137238**; Morgan Lewis. *Key Considerations for Evaluating an AI Vendor*. 2024. **https://www.morganlewis.com/-/media/files/publication/outside-publication/article/2024/key-considerations-for-evaluating-an-ai-vendor.pdf**

[35] Mohammed, Irshadullah Asim. "Artificial Intelligence in Supplier Selection and Performance Monitoring: A Framework for Supply Chain Managers." *Educational Administration: Theory and Practice*, 2023. **https://www.researchgate.net/publication/386598338**; Morgan Lewis. *Key Considerations for Evaluating an AI Vendor*. 2024. **https://www.morganlewis.com/-/media/files/publication/outside-publication/article/2024/key-considerations-for-evaluating-an-ai-vendor.pdf**

[36] Mohammed, Irshadullah Asim. "Artificial Intelligence in Supplier Selection and Performance Monitoring: A Framework for Supply Chain Managers." *Educational Administration: Theory and Practice*, 2023. **https://www.researchgate.net/publication/386598338**

[37] Mohammed, Irshadullah Asim. "Artificial Intelligence in Supplier Selection and Performance Monitoring: A Framework for Supply Chain Managers." *Educational Administration: Theory and Practice*, 2023. **https://www.researchgate.net/publication/386598338**; Waditwar, Prajkta. (2024). "The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations." *Open Journal of Business and Management* 12(2), 385–400. **https://www.scirp.org/journal/paperinformation?paperid=137238**; Morgan Lewis. *Key Considerations for Evaluating an AI Vendor*. 2024. **https://www.morganlewis.com/-/media/files/publication/outside-publication/article/2024/key-considerations-for-evaluating-an-ai-vendor.pdf**

[38] Morgan Lewis. *Key Considerations for Evaluating an AI Vendor*. 2024. **https://www.morganlewis.com/-/media/files/publication/outside-publication/article/2024/key-considerations-for-evaluating-an-ai-vendor.pdf**

[39] Waditwar, Prajkta. (2024). "The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations." *Open Journal of Business and Management* 12(2), 385–400. **https://www.scirp.org/journal/paperinformation?paperid=137238**

[40] Waditwar, Prajkta. (2024). "The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations." *Open Journal of Business and Management* 12(2), 385–400. **https://www.scirp.org/journal/paperinformation?paperid=137238**

[41] "The Role of Artificial Intelligence in the Procurement Process." *Procedia Computer Science* 220 (2023), 127–135. **https://www.sciencedirect.com/science/article/pii/S1878092230000079**; Morgan Lewis. *Key Considerations for Evaluating an AI Vendor*. 2024. **https://www.morganlewis.com/-/media/files/publication/outside-publication/article/2024/key-considerations-for-evaluating-an-ai-vendor.pdf**

[42] "The Role of Artificial Intelligence in the Procurement Process." *Procedia Computer Science* 220 (2023), 127–135. **https://www.sciencedirect.com/science/article/pii/S1878092230000079**

[43] Kumo. "How Much Does AI Chatbot Development Cost?" Last modified February 21, 2025. **https://www.kumohq.co/blog/ai-chatbot-development-cost**

[44] Phenom. "Electrolux Group Digitalizes Key HR Processes for Distinct Hiring Edge." *Phenom*. **https://www.phenom.com/resource/electrolux-group-digitalizes-processes-hiring-edge**

[45] Verizon. *Verizon 2023 TCFD Report*. March 2023. **https://www.verizon.com/about/sites/default/files/Verizon-2023-TCFD-Report.pdf**

[46] Verizon. "Verizon Unveils AI Strategy to Power Next-Gen AI Demands." *Verizon News Center*, February 21, 2024. **https://www.verizon.com/about/news/verizon-unveils-ai-strategy-power-next-gen-ai-demands**

[47] Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366, no. 6464 (2019): 447-453. **https://www.science.org/doi/10.1126/science.aax2342**.

[48] Forbes Finance Council. (2023, March 9). "The Future of Mortgage Lending: How AI and Humans Can Coexist." *Forbes.* **https://www.forbes.com/councils/forbesfinancecouncil/2023/03/09/the-future-of-mortgage-lending-how-ai-and-humans-can-coexist/**

[49] Court, E. (2024, April 23). "AI Lawsuits against Insurers Signal Wave of Health Litigation." *Bloomberg Law.* **https://news.bloomberglaw.com/health-law-and-business/ai-lawsuits-against-insurers-signal-wave-of-health-litigation**

[50] Government of Singapore. "SkillsFuture: A Decade and Beyond." SkillsFuture. Last modified May 23, 2025. **https://www.skillsfuture.gov.sg**

[51] World Economic Forum. (2023, May 2). "Lessons from Singapore on Upskilling for the Future." *World Economic Forum*, **https://www.weforum.org/stories/2023/05/lessons-from-singapore-on-upskilling-for-the-future/**

[52] Lin, Belle. "Why Amazon Is Betting on 'Automated Reasoning' to Reduce AI's Hallucinations," *The Wall Street Journal*, February 5, 2025. **https://www.wsj.com/articles/why-amazon-is-betting-on-automated-reasoning-to-reduce-ais-hallucinations-b838849e**. WSJ

[53] AllBusiness.com Team. "Retrieval-Augmented Generation (RAG)." *TIME*, April 3, 2025. https://time.com/collection_hub_item/definition-of-retrieval-augmented-generation/

[54] Gupta, Abhishek. "Guardrails for Amazon Bedrock Can Now Detect Hallucinations and Safeguard Apps Built Using Custom or Third-Party FMs," *AWS News Blog*, July 10, 2024. **https://aws.amazon.com/blogs/aws/guardrails-for-amazon-bedrock-can-now-detect-hallucinations-and-safeguard-apps-built-using-custom-or-third-party-fms/**

[55] Cecco, Leyland. "Air Canada Ordered to Pay Customer Who Was Misled by Airline's Chatbot." *The Guardian*, February 16, 2024. https://www.theguardian.com/world/2024/feb/16/air-canada-chatbot-compensation

CHAPTER 8
The Art of Prompt Engineering with FOCUS



MT.PHOTOSTOCK/Shutterstock

Learning Objectives

- **LO8-1** Apply the FOCUS framework to construct clear, specific, and contextually relevant prompts.
- **LO8-2** Employ the iterative process of refining prompts to achieve desired outcomes.
- **LO8-3** Analyze the structure and functionality of different AI models.
- **LO8-4** Utilize advanced techniques like task decomposition to create effective prompts.
- **LO8-5** Describe the use of personas in prompt engineering.
- **LO8-6** Describe strategies for verifying the accuracy of AI responses.
- **LO8-7** Describe ethical considerations in prompt engineering.

INTRODUCTION

“The Art of Prompt Engineering” chapter title was chosen deliberately because crafting effective prompts requires technical precision and creative ingenuity. Like a skilled artist who uses tools and techniques to bring a vision to life, prompt engineers must balance structure, intuition, and adaptability to guide AI systems toward meaningful outputs. This chapter explores the intersection of creativity and logic in prompt engineering, demonstrating how it is both a science and an art. We present a framework called FOCUS for crafting a good prompt. Throughout this chapter, we’ll see how the FOCUS framework guides crafting effective prompts and analyzing AI outputs. We will also discuss advanced prompt engineering techniques, strategies for verifying the accuracy of AI responses, and ethical considerations in prompt engineering.

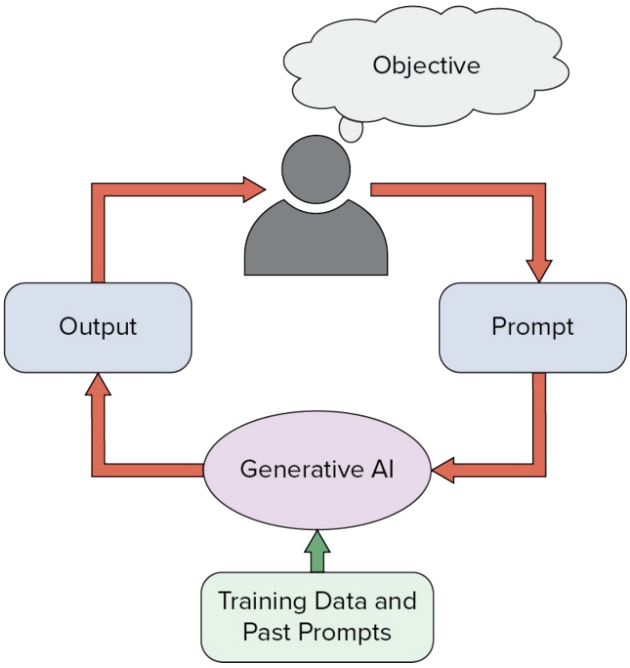
THE FOCUS FRAMEWORK FOR EFFECTIVE PROMPTS (LO8-1)

Prompt engineering involves carefully crafting inputs (or *prompts*) to guide generative AI systems toward relevant and accurate outputs. Think of prompt engineering as the steering wheel of a car; the car (the AI) has the power to move, but the driver reacts to road conditions and turns the wheel to ensure the car reaches the desired destination.

As generative AI tools become increasingly powerful, the ability to effectively communicate with them is becoming a vital skill. Effective prompt engineering involves initiating a dialogue in which clarity, context, and specificity are instrumental in achieving the desired outcomes. Dialogue is a key term here. As **Figure 8.1** shows, the first prompt rarely yields the ideal desired output. Achieving the desired output is an iterative process that involves guiding the AI to a useful output.

FIGURE 8.1 Prompt Engineering

Achieving the desired output is an iterative process in which the user and the AI collaborate through the progressive refinement of the prompts given to the AI.



General Principles of Prompt Crafting

Good prompts are clear, specific, goal-oriented, and contain contextual references. For best

results, they ask a single question. A good prompt should use the **FOCUS framework**:

F: Focused: A prompt should be specific and provide directions to guide the AI effectively.

O: Objective: Each prompt should have a goal or intended outcome.

C: Context: Providing relevant background enhances the relevance and accuracy of the response, gearing it appropriately for the intended audience.

U: Unambiguous: Prompts must use clear, direct language to avoid misunderstandings.

S: Single Query: A singular focus prevents confusion and ensures a targeted response.

In the next sections, we examine each of these aspects, providing examples of weak prompts and ones that better achieve our desired outputs.

Common Problems Students and Professionals Tackle with Generative AI

For Students

- ***Homework help.*** Explaining complex concepts in simple terms (e.g., "What is the Pythagorean theorem, and how is it used?").
- ***Essay assistance.*** Generating ideas, creating outlines, or refining drafts.
- ***Study aid.*** Summarizing long reading assignments or preparing for exams with practice questions.
- ***Project guidance.*** Helping with research topics or providing example scenarios for case studies.
- ***Language practice.*** Offering translations, grammar checks, or conversational practice in a foreign language.

For Professionals

- ***Drafting communication.*** Composing emails, proposals, or reports tailored to specific audiences.
- ***Brainstorming.*** Generating creative ideas for marketing campaigns, product development, or project strategies.
- ***Technical assistance.*** Simplifying complex technical instructions for colleagues or clients.
- ***Problem solving.*** Breaking down complex tasks, like analyzing data or creating project roadmaps.
- ***Career support.*** Drafting résumés, preparing for interviews, or planning career growth.

F: Focused

Specificity in prompt crafting is extremely important. The more specific and focused a prompt is, the more focused and valuable the AI's response will be. Include essential details that direct the AI toward the required information or action. For instance, "Analyze sales data for product X in Q1 2025, focusing on demographic trends" is far more specific than "Look at our sales."

Here are some keys to creating a focused prompt:

- ***Include requirements.*** Indicate the desired length, format, and writing style (e.g., "a 200-word blog post in a conversational tone").
- ***Include relevant detail.*** Include critical data, like names, dates, or numerical values (e.g., "data shows that quarterly sales in 2024 increased by 15 percent compared to 2023, rising from $2 million to $2.3 million").
- ***Be exact.*** Instead of "analyze data," specify the data type, field, or analysis type (e.g., "analyze sales growth by quarter").
- ***Provide an example or template.*** If relevant, provide an example or structure for the response to ensure precision (e.g., "Output the data in the form of a table with the headings: Company, Industry, Date Founded, Reported Profits for 2024").

Let's compare a vague, unclear prompt example to a more focused prompt.

Vague, Unclear Prompt

- Prompt: "Tell me about sustainability in business."
- ChatGPT-4 Response: *Sustainability in business refers to practices that do not deplete resources or harm ecological systems, supporting long-term ecological balance. Businesses adopt sustainable practices to mitigate their environmental impact, improve efficiency, and promote social responsibility. This includes strategies like reducing waste, conserving energy, and using sustainable materials, which can also lead to cost savings and improved brand reputation.*

Source: ChatGPT4, 8.1 Vague Prompt, 9/18/2024

Declaration: These are results generated by ChatGPT4 that took place on September 18, 2024. AI companies might have changed/improved their outputs since this content was generated.

Focused, Clear Prompt

- **Prompt:** "How do businesses implement sustainability practices to reduce environmental impact?"
- **ChatGPT-4 Response:** *Businesses implement sustainability practices to reduce environmental impact by adopting various strategies. Key methods include:*
  1. ***Resource efficiency.*** *Optimizing the use of resources like water, energy, and materials to minimize waste.*
  2. ***Renewable energy.*** *Shifting to renewable energy sources such as solar, wind, or hydro to reduce carbon emissions.*
  3. ***Sustainable supply chains.*** *Ensuring suppliers adhere to environmentally friendly practices, reducing the overall ecological footprint.*
  4. ***Waste management.*** *Implementing recycling programs and reducing packaging to minimize landfill waste.*
  5. ***Product lifecycle.*** *Designing products that are more durable, repairable, and recyclable to extend their lifespan and reduce environmental impact.*
  6. ***Green buildings.*** *Constructing or retrofitting buildings to meet energy-efficient standards and reduce resource consumption.*

By integrating these practices, businesses can significantly lessen their environmental impact while also potentially reducing costs and enhancing their market competitiveness.

Source: ChatGPT4, 8.2 Specific Prompt, 9/18/2024

Declaration: These are results generated by ChatGPT4 that took place on September 18, 2024. AI companies might have changed/improved their outputs since this content was generated.

The initial prompt is vague and doesn't specify a particular aspect of sustainability, leading to a broad and general response from the AI. The second prompt is clear and directs the AI to focus on specific sustainability practices related to environmental impact. The result is a more focused and detailed answer. Specificity and clarity more effectively guide the AI in providing the targeted information required.

O: Objective

A well-crafted prompt should be goal oriented. Suppose the objective is to create a list of companies to research. In that case, the prompt should reflect that goal:

> "Please generate a list of companies that have successfully achieved a net carbon zero status with links to the articles or web pages about that achievement."

If the goal is to arrive at a decision, the prompt should be structured to elicit information that supports this process.

> "Compare the three proposals in the document; which is most likely to appeal to customers over 50?"

Understanding and articulating your objectives helps design prompts that are precise and effective in achieving specific goals. Here are some keys to creating goal-oriented prompts:

- ***State the purpose.*** Clearly state the intended outcome or purpose of the prompt (e.g., "to inform budget allocation decisions").
- ***Use actionable language.*** Use terms that imply specific actions or results (e.g., "summarize," "analyze," "generate," or "compare").
- ***Link to desired outcomes.*** Connect the prompt to a larger objective or end goal so the AI can tailor responses accordingly (e.g., "with the aim to improve customer satisfaction").
- ***Set quality benchmarks.*** Include any specific quality criteria, such as accuracy, completeness, or relevance, that align with the end goal (e.g., "Analyze customer feedback data from TechEase for Q2 2025, highlighting the top three issues impacting customer satisfaction").

Let's look at an example of the impact of including a goal in a prompt.

Prompt without Goal Orientation

- **Prompt:** "What is greenwashing in business?"
- **ChatGPT-4 Response:** *Greenwashing in business refers to the practice where companies misleadingly promote their products, services, or corporate practices as environmentally friendly when they are not. This can involve exaggerating claims, using vague or unsubstantiated terms like "eco-friendly" or "natural," or focusing marketing on minor green initiatives while ignoring larger harmful practices. Greenwashing is often used to appeal to eco-conscious consumers and create a false impression of environmental responsibility. It can lead to consumer distrust and legal repercussions once discovered.*

Source: ChatGPT4, 8.5 Prompt without Goal Orientation, 9/18/2024.

Declaration: These are results generated by ChatGPT4 that took place on September 18, 2024. AI companies might have changed/improved their outputs since this content was generated.

Prompt with Goal Orientation

- **Prompt:** "How can businesses avoid greenwashing while genuinely implementing and communicating their sustainability efforts to consumers?"
- **ChatGPT-4 Response:** *Businesses can avoid greenwashing and genuinely communicate their sustainability efforts by following these practices:*
  1. ***Authentic actions.*** *Ensure that sustainability initiatives are substantial, impactful, and aligned with broader environmental goals, rather than just for show.*
  2. ***Transparency.*** *Provide clear, honest, and detailed information about what the company is doing to be sustainable, including both successes and areas for improvement.*
  3. ***Third-party verification.*** *Obtain certifications from recognized and respected independent organizations to validate claims about sustainability practices.*
  4. ***Consistency.*** *Ensure that all business operations, not just a few products or processes, adhere to sustainable practices.*
  5. ***Stakeholder engagement.*** *Involve customers, employees, and other stakeholders in sustainability efforts, gathering feedback and ideas to improve initiatives.*
  6. ***Educational marketing.*** *Use marketing and advertising to educate consumers about the environmental impact of their purchases and how the company's practices make a difference. By prioritizing real change over superficial claims, businesses can build credibility and trust with consumers, avoiding the pitfalls of greenwashing.*

Source: ChatGPT4, 8.6 Prompt with Goal Orientation, 9/18/2024.

Declaration: These are results generated by ChatGPT4 that took place on September 18, 2024. AI companies might have changed/improved their outputs since this content was generated.

The initial response gave a general explanation of greenwashing in business. The second prompt elicits a more specific and actionable response. It defines greenwashing and provides a strategy for companies to implement and communicate their sustainability efforts. Goal-oriented prompts can guide AI responses to specific objectives or solve particular problems.

C: Context

Context helps the AI understand the broader scenario, enhancing the relevance and accuracy of its response. Contextual relevance might include background information, the purpose of the query, the age or education level of the audience, or any specific constraints. For example, a prompt like "Considering the recent EU regulations, what would be the best market entry strategy for our new product?" or "Please craft an email from me to employees in my department alerting them that our building will be closed for the next two days due to the anticipated bad weather. Instruct them to work remotely and to be safe." gives the AI the necessary context to tailor its response.

Here are some tips to creating a contextual prompt:

- ***Provide background.*** Include the target audience and intended use—persuasion, education, informational (e.g., "the focus is for first-time home buyers and how to best educate them on their options").

- ***Set constraints.*** Define any limitations, such as timeframes, regions, or specific data sets relevant to the query (e.g., "limit the results to Q1 and Q2 of 2025").
- ***Mention prior work or inputs.*** If the prompt builds on prior outputs or data, clarify this (e.g., "build on the previous analysis from Q1 sales data").
- ***Clarify stakeholders or perspectives.*** Mention any stakeholders or perspectives that might influence the response (e.g., "focus on investor insights"). Assigning the AI a perspective to answer the question from, "As a physical trainer, assess this training schedule for a novice running their first 5K happening six weeks from now," can generate valuable insights.

Let's continue with the sustainability example to see how ChatGPT-4's response changes when adding context.

Prompt without Contextual Relevance

- **Prompt:** "What are sustainable business practices?"
- **ChatGPT-4 Response:** *Sustainable business practices are methods and strategies that minimize negative impact on the environment, society, and economy. These practices are designed to ensure long-term business viability and ecological health. Key practices include:*
  1. ***Environmental stewardship.*** *Reducing carbon footprints, increasing energy efficiency, and managing waste effectively.*
  2. ***Social responsibility.*** *Ensuring fair labor practices and contributing to the community through engagement and philanthropy.*
  3. ***Economic viability.*** *Operating in a way that is financially sustainable while also conscious of environmental and social impacts.*
  4. ***Ethical sourcing.*** *Obtaining materials and products from suppliers that prioritize sustainability and fair labor conditions.*
  5. ***Product innovation.*** *Designing products with lower environmental impacts throughout their lifecycle, from production to disposal.*

  Implementing these practices helps businesses not only reduce their environmental impact, but also build trust with consumers and stakeholders, potentially leading to a competitive advantage.

Source: ChatGPT4, 8.3 Prompt Without Context, 9/18/2024

Declaration: These are results generated by ChatGPT4 that took place on September 18, 2024. AI companies might have changed/improved their outputs since this content was generated.

Prompt with Contextual Relevance

- **Prompt:** "Considering the increasing consumer demand for eco-friendly products, how are businesses adapting their practices for sustainability?"
- **ChatGPT-4 Response:** *Businesses are adapting their practices for sustainability in response to increasing consumer demand for eco-friendly products by:*
  1. ***Eco-friendly products.*** *Developing products that use sustainable materials and processes, such as biodegradable packaging or organic ingredients.*
  2. ***Carbon footprint reduction.*** *Implementing measures to reduce greenhouse gas emissions, such as energy-efficient manufacturing techniques and using renewable energy sources.*
  3. ***Transparency.*** *Providing clear information about product sourcing, manufacturing processes, and corporate sustainability efforts to educate and assure consumers.*
  4. ***Certifications.*** *Pursuing sustainability certifications like Fair Trade, Organic, or LEED to validate their eco-friendly claims and stand out in the market.*
  5. ***Circular economy practices.*** *Adopting models that design waste out of the system, such as recycling materials or*

offering product-as-a-service options.

These adaptations not only meet consumer expectations but also help businesses reduce costs and enhance their brand reputation in a competitive market.

Source: ChatGPT4, 8.4 Prompt with Context, 9/18/2024

Declaration: These are results generated by ChatGPT4 that took place on September 18, 2024. AI companies might have changed/improved their outputs since this content was generated.

The first prompt provides a general overview of sustainable business practices. By adding the context of consumer demand for eco-friendly products, the generative AI response becomes more specific. It addresses how businesses adapt their sustainability practices in response to market trends, yielding a more targeted and informative answer. Contextual relevance in prompts can significantly refine and focus AI's responses.

U: Unambiguous

Effective communication requires clarity. A clear prompt avoids ambiguities, using straightforward language that leaves little room for misinterpretation. For instance, instead of saying, "Tell me about the market," a clearer prompt would be, "Provide an analysis of the current trends in the U.S. stock market for technology companies."

Here are some guidelines to creating an unambiguous prompt:

- ***Have a single query focus.*** To avoid confusion, the prompt should target one main question or action you want addressed (e.g., "Identify and briefly explain the most significant factor causing the recent decline in employee productivity at GreenLeaf Inc.").
- ***Be concise.*** Stick to necessary details, omitting irrelevant information (e.g., "Recommend two immediate actions to reduce shipping delays at QuickShip Logistics based on recent customer complaints.").
- ***Use direct language.*** Avoid vague or ambiguous terms (e.g., "List three strategies to immediately improve customer retention at Joe's Pizza Palace.").
- ***Avoid assumptions.*** Clearly state any specifics that might otherwise be assumed, like units of measurement or data format (e.g., "Provide the average monthly rainfall (in inches) for Chicago in 2024, rounded to one decimal place.").

Ambiguous Prompt

- **Prompt:** "Explain why sales dropped."
- **ChaptGPT Response:** *I'd be happy to help! To give you a precise answer, could you please clarify:*
  1. *Which company's sales are you referring to?*
  2. *Over what specific time period did the sales drop occur?*
  3. *Are there any particular products or regions involved?*

  Having these details will enable me to provide an accurate analysis.

Source: ChatGPT4, 8.16 Ambiguous Prompt, 9/18/2024

Declaration: These are results generated by ChatGPT4 that took place on September 18, 2024. AI companies might have changed/improved their outputs since this content was generated.

Unambiguous Prompt

- **Prompt:** "Analyze the main reasons why sales for FreshMart groceries declined by 10 percent in Q1 2025 compared to Q4 2024. Provide two specific recommendations for improvement."
- **ChaptGPT-4 response:** *Sales at FreshMart groceries declined by 10 percent in Q1 2025 primarily due to:*

1. ***Seasonal demand fluctuation.*** *Consumer purchasing typically dips following the holiday season (Q4), reducing overall grocery sales.*
2. ***Increased competition.*** *Competitors launched aggressive promotions during Q1, diverting FreshMart's regular customers.*

***Recommendations for Improvement*:**

1. ***Seasonal promotions.*** *Implement targeted promotions or loyalty incentives early in Q1 to sustain customer engagement after the holiday season.*
2. ***Competitive pricing.*** *Match or strategically counter competitor discounts to retain market share and customer loyalty.*

Source: ChatGPT4, 8.17 Unambiguous Prompt, 9/18/2024

Declaration: These are results generated by ChatGPT4 that took place on September 18, 2024. AI companies might have changed/improved their outputs since this content was generated.

S: Single Query

Keeping your prompt to a single query is essential for guiding AI to deliver precise and relevant results. When a prompt contains multiple questions or objectives, the AI may struggle to determine which aspect to prioritize, leading to incomplete or unclear responses. A singular focus ensures the AI can dedicate its processing to addressing one specific task, reducing ambiguity and improving the overall quality and accuracy of the output. This approach is particularly useful when tackling complex problems, as it allows for iterative refinement by addressing each sub-question separately.

Here are some guidelines for using a single query prompt:

1. ***State clearly.*** Explicitly phrase your prompt as one specific question or task (e.g., "What was the average temperature in Miami for June 2024?").
2. ***Limit scope.*** Avoid combining multiple requests in one prompt (e.g., "List five reasons for declining website traffic.").
3. ***Be precise.*** Clearly specify what information or format you expect (e.g., "Provide one recommendation to boost user engagement on Facebook.").
4. ***Eliminate ambiguity.*** Provide necessary context, units, or specific criteria clearly (e.g., "What's the total revenue in USD for Q2 2025?").
5. ***Separate complex tasks.*** Break down complex tasks into simpler, distinct prompts (e.g., Prompt 1: "Analyze the main factor behind recent employee turnover." / Prompt 2: "Suggest two ways to address employee turnover based on that factor.").

Complex Prompt

- **Prompt:** "Can you tell me about recent company performance issues, feedback from clients, employees, or maybe competitors, and also suggest some things we might do about it?"
- **ChaptGPT-4 response:** *I'd be glad to help you analyze the company's performance, but the current request covers multiple broad areas. To ensure my response accurately addresses your needs, could you please clarify:* ***Which company's*** *performance you want analyzed? Do you prefer analysis focused on* ***client feedback, employee insights, or competitor actions****? (Please choose one for clarity.) Are you specifically looking for* ***solutions or recommendations****, or just insights into performance issues? Providing this additional detail will help me deliver a clear and precise answer.*

Source: ChatGPT4, 8.18 Complex Prompt, 9/18/2024

Declaration: These are results generated by ChatGPT4 that took place on September 18, 2024. AI companies might have changed/improved their outputs since this content was generated.

Simplified Prompt

- **Prompt:** "Identify one primary reason customer satisfaction dropped at SpeedyService Inc. in February 2025 based on customer feedback."
- **ChaptGPT-4 response:** *The primary reason customer satisfaction dropped at SpeedyService Inc. in February 2025 was increased waiting times, as consistently mentioned in customer feedback surveys.*

Source: ChatGPT4, 8.19 Simplified Prompt, 9/18/2024

Declaration: These are results generated by ChatGPT4 that took place on September 18, 2024. AI companies might have changed/improved their outputs since this content was generated.

Table 8.1 briefly summarizes how to use the FOCUS framework. Throughout this discussion, you've seen how important it is to guide the AI by setting boundaries, providing structure, or including relevant details. These strategies are all forms of what AI developers and instructional designers call **scaffolding**—helping the AI "see the shape" of your request. Just as scaffolding supports a building during construction, prompt scaffolding supports the AI as it builds its response. You might scaffold a prompt by adding an example, giving step-by-step instructions, or setting constraints. We explore more scaffolding strategies later in the chapter when dealing with complex or hard-to-handle prompts.

Table 8.1 **FOCUS in Practice**

Prompting Element	Definition	Tips
F - Focused	Specific and actionable	Add keywords, metrics, or limits
O - Objective	Has a clear goal or outcome	Reword to reflect what success looks like
C - Context	Provides background or audience	Add references to constraints or tone
U - Unambiguous	Uses clear, direct language	Remove or define vague terms
S - Single query	Asks one thing at a time	Break into parts if complex

Hard Prompts: Benchmarks for AI

Now that we've covered how to write strong, well-scoped prompts, it's worth looking at the opposite extreme—**hard prompts**. These intentionally complex inputs challenge the AI's ability to reason, generalize, or improvise. Hard prompts aren't meant for day-to-day use in business settings, but they can be useful if you're:

- Comparing different AI tools.
- Evaluating system updates.
- Diagnosing limitations in AI performance.

For example, the prompt, "*Generate a legal contract between a time-traveling startup and its 18th-century investor*" may sound silly, but it tests how well the AI handles historical context, legal formatting, and genre-blending logic simultaneously. Running the same prompt across multiple chatbots can reveal differences in how they reason, apply knowledge, or hallucinate.

In real-world business use—especially when reliability matters—avoid hard prompts. Instead:

- Use scaffolding (e.g., provide examples in your prompt or break the task into smaller pieces).
- Maintain human oversight to check outputs.
- Break large or ambiguous requests into simpler, more precise steps.
- Add constraints or context to clarify your intent.

In addition, where possible, use tools that enhance the model's accuracy, such as **retrieval-augmented generation (RAG),** a method in which an AI model pulls information from external sources in real time to supplement its responses with more accurate or current information. To accomplish this, access information from sources such as databases, websites, or scientific papers and then add that information into the query. For example: *"Based on the provided 2025 EU Energy Report snippet, explain the primary reasons behind the recent 12 percent decline in renewable energy investment in Europe."*

ITERATIVE REFINEMENT (LO8-2)

Whether you're troubleshooting a hard prompt or just refining a basic one, effective AI use often requires iteration. Even with a strong initial prompt, you may need to tweak your prompt based on how the AI responds. Let's look at how to improve your prompts through **iterative refinement**.

An open-domain chatbot is like an over-eager intern who's so desperate to impress that they will confidently answer any question—even if they have no clue. That's because generative AIs create responses by predicting the next word based on patterns in their training data—not by looking up verified facts. This process can produce fluent and convincing responses that may be entirely false. To get useful output, follow this guiding principle:

Check the response. If it doesn't meet your needs or cannot be verified, refine your prompt and try again.

When evaluating AI responses, consider the following:

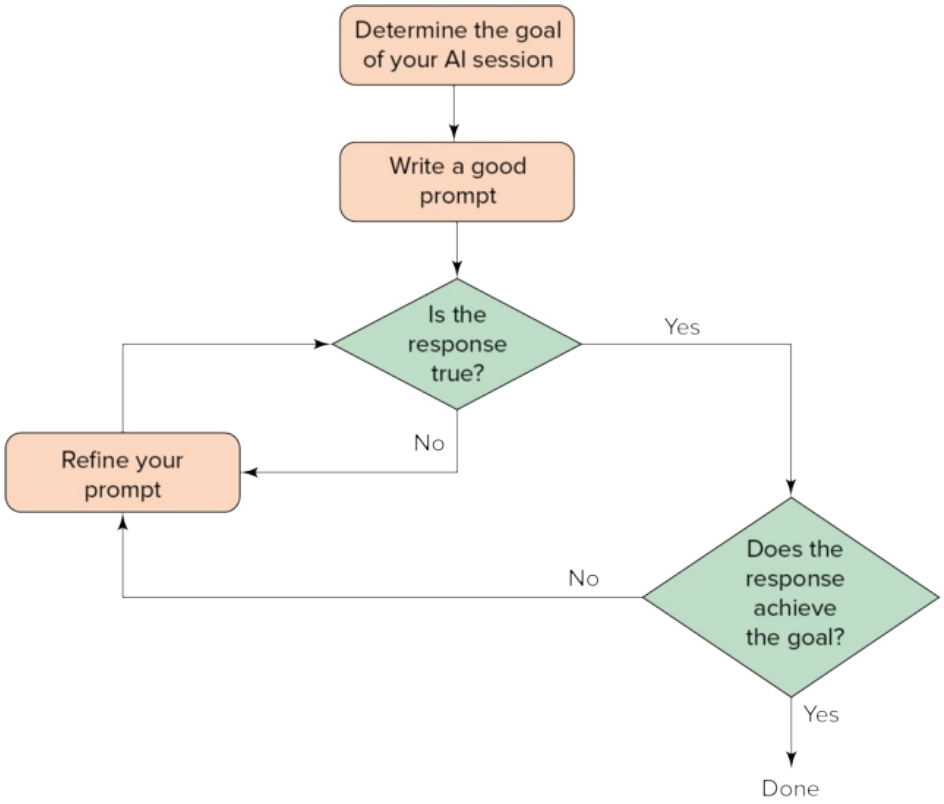
1. ***Relevance.*** Does the response directly address your query or stray off target?
2. ***Accuracy.*** Are the facts correct and consistent with verified information? (We cover this in a later section.)
3. ***Completeness.*** Does the response cover all your needs?

Refine your prompt if the output seems misaligned—irrelevant, incorrect, incomplete, or confusing. Here are some standard refinements you can make:

1. ***Increase specificity.*** Add more detail to narrow the focus. For example, replace "Tell me about renewable energy" with "Explain the advantages and disadvantages of solar power compared to wind energy."
2. ***Clarify smbiguities.*** Replace vague or multi-meaning terms with precise language.
3. ***Set context.*** Provide background or frame your prompt to guide the AI's interpretation.
4. ***Rephrase the prompt.*** Experiment with different ways of asking the same question.
5. ***Provide instructions.*** Use directive language, like "List three reasons..." or "Summarize in one paragraph."
6. ***Include a negative constraint.*** Limit the material that the chatbot considers or includes. "Omit movies from the list that are animated."

When refining prompts (see **Figure 8.2**), check against the FOCUS framework—does the revised prompt maintain focus, does it have an objective, does it provide the right context, is it unambiguous, and does it have a singular focus? Iteratively analyzing responses and refining your prompts can significantly improve the alignment between what you want and what the AI delivers. Prompt engineering is as much about understanding how to ask as it is about evaluating the answers you receive.

FIGURE 8.2 Simple Flowchart of the Iterative Nature of Successful Prompting



Prompt Refinement Example

Let's look at an example of prompt refinement in action. Say we want to identify and learn about a small, calm dog breed suitable for travel in a camper and less likely to provoke allergies. Forgetting everything we learned about writing good prompts, we started with a general statement: "I would like to get a dog."

Initial Prompt: "I would like to get a dog."

ChatGPT-4 Response: *That's exciting! Getting a dog is a big decision and can bring a lot of joy to your life. Here are a few things to consider:*

1. ***Lifestyle match.*** *Choose a breed or mix that fits well with your lifestyle and living situation. Consider the dog's energy level, size, and temperament.*
2. ***Time commitment.*** *Dogs require time for training, exercise, and companionship. Make sure you have enough time to dedicate to their care.*
3. ***Financial responsibility.*** *Be prepared for the costs of veterinary care, food, grooming, and possibly training.*
4. ***Long-term commitment.*** *Dogs live on average 10 to 15 years. It's important to commit for their entire life.*
5. ***Adoption.*** *Consider adopting from a shelter or rescue organization. Many wonderful dogs need loving homes.*

Once you've thought it through and are ready, you'll have a loyal companion by your side!

Source: ChatGPT4, 8.7 Prompt Refinement Step 1, 9/18/2024

Declaration: These are results generated by ChatGPT4 that took place on September 18, 2024. AI companies might have changed/improved their outputs since this content was generated.

ChatGPT recognizes the vagueness of the prompt and educates the user on items to consider when choosing a dog. However, we already know the important criteria for the dog we want, so this response is not useful. The prompt needs to provide more information about these requirements to move toward a helpful response and pose it as a question that the AI can focus on.

Refined Prompt: "Can you recommend breeds of dogs that are good for traveling and don't provoke an allergic response?"

ChatGPT-4 Response:

Certainly! Here are some dog breeds that are well-suited for traveling and are also less likely to provoke allergic responses:

1. ***Poodle (Toy, Miniature, and Standard):***
   - ***Hypoallergenic.*** *Poodles have a curly coat that sheds minimally and is known for being hypoallergenic.*
   - ***Travel-friendly.*** *Intelligent and easily trainable, Poodles can adapt well to various travel environments.*
2. ***Maltese*:**
   - ***Hypoallergenic.*** *These dogs have long, silky hair that sheds very little.*
   - ***Travel-friendly.*** *Maltese are small and typically easy to manage during travels.*
3. ***Portuguese Water Dog*:**
   - ***Hypoallergenic.*** *Known for their curly and dense coat that does not shed much.*
   - ***Travel-friendly.*** *Energetic and adventurous, they are great companions for active travelers.*
4. ***Yorkshire Terrier*:**
   - ***Hypoallergenic.*** *They have a fine, straight coat that sheds very little.*
   - ***Travel-friendly.*** *Yorkies are tiny, making them convenient for travel and comfortable in small spaces like car seats or carriers.*
5. ***Shih Tzu*:**
   - ***Hypoallergenic.*** *Although they have a long, flowing coat, Shih Tzus shed very little.*
   - ***Travel-friendly.*** *They are generally calm and happy to be wherever their owner is, which makes them great travel buddies.*
6. ***Bichon Frise*:**
   - ***Hypoallergenic.*** *Their curly coat is highly hypoallergenic as they shed less.*
   - ***Travel-friendly.*** *Bichons are cheerful, adaptable, and small, ideal for constant companionship on the road.*

These breeds are known for their suitability for allergy sufferers and their adaptability to traveling, making them excellent companions for your adventures.

Source: ChatGPT4, 8.8 Prompt Refinement Step 2, 9/18/2024

Declaration: These are results generated by ChatGPT4 that took place on September 18, 2024. AI companies might have changed/improved their outputs since this content was generated.

We are closer to our goal of finding a good dog to travel with, but there are dogs on this list that would be overwhelming in the confined space of the truck and camper. We will use a negative prompt to narrow the list of dog breeds considered.

Negative Prompt: "We will be traveling in a truck pulling a camper. Space will be limited. Please remove large dogs from consideration."

ChatGPT-4 Response:

Certainly! Focusing on smaller breeds that are both travel-friendly and hypoallergenic, suitable for limited spaces like a

truck camper, here are some excellent choices:

1. ***Poodle (Toy or Miniature):***
   - ***Hypoallergenic.*** *These Poodles have a curly, minimal-shedding coat that is great for allergy sufferers.*
   - ***Travel-friendly.*** *Their small size and intelligent, adaptable nature make them ideal for traveling in confined spaces.*
2. ***Maltese:***
   - ***Hypoallergenic.*** *Maltese have long, silky hair that sheds very little.*
   - ***Travel-friendly.*** *Their compact size makes them easy to manage during travel and comfortable in smaller living quarters.*
3. ***Bichon Frise:***
   - ***Hypoallergenic.*** *Bichons have a curly coat that produces less dander and sheds minimally.*
   - ***Travel-friendly.*** *Small and cheerful, they adapt well to various environments and are easy to carry around.*
4. ***Yorkshire Terrier:***
   - ***Hypoallergenic.*** *They have a fine, straight coat that sheds very little.*
   - ***Travel-friendly.*** *Extremely small in size, Yorkies are perfect for travel and can easily fit into small spaces.*
5. ***Shih Tzu:***
   - ***Hypoallergenic.*** *They have a long, flowing coat that sheds minimally.*
   - ***Travel-friendly.*** *Compact and sturdy, Shih Tzus are calm and comfortable with traveling and can easily adapt to life on the road.*

These smaller breeds should fit well in the limited space of a truck camper, providing companionship without the challenge of managing a large dog in a confined area.

Source: ChatGPT4, 8.9 Prompt Refinement Step 3, 9/18/2024

Declaration: These are results generated by ChatGPT4 that took place on September 18, 2024. AI companies might have changed/improved their outputs since this content was generated.

This response is much closer to the desired outcome. It offers a list of five small to medium sized dogs that would be good candidates. At this point we decide to focus on two of the breeds and acquire details regarding their care and lifestyle compatibility.

Next Prompt: "The traveling I am considering is in a truck pulling a camper on trips that may last several months. Can you provide detailed information about the Bichon Frise and Poodle, focusing on their temperament, care needs, and suitability for frequent travel?"

ChatGPT-4 Response:

Bichon Frise

Temperament:

- Bichon Frises are known for their cheerful and affectionate nature. They are very adaptable and get along well with people and other pets.

- They are generally very sociable, which makes them good companions for frequent stops and interactions during travels.

- Their playful and gentle disposition makes them great for families and individuals alike.

Care needs:

- Bichons have a soft, curly coat that does not shed much but requires regular grooming to prevent matting.

- They need daily brushing and professional grooming every 4 to 6 weeks.

- Although they require regular exercise, their needs can usually be met with short walks and play sessions, making them suitable for life on the road.

Suitability for frequent travel:

- Their small size makes them easy to manage in confined spaces such as a truck or camper.

- Bichons are adaptable to new environments and can handle changes in routine relatively well, important for frequent travel.

Poodle (Toy or Miniature)

Temperament:

- *Poodles are highly intelligent and easily trained, which is beneficial for managing behavior on long trips.*
- *They are known for being alert and active, often enjoying the stimulation of new environments and experiences.*
- *Poodles are typically very loyal to their owners, forming strong bonds that make them excellent travel companions.*

Care Needs:

- *Like the Bichon, Poodles have a hypoallergenic coat that requires regular grooming to keep it clean and tangle-free.*
- *Daily brushing is necessary, along with professional grooming every 4 to 6 weeks to maintain their coat's condition.*
- *They are energetic and require regular exercise but can adapt their activity levels to their environment, making them flexible travelers.*

Suitability for Frequent Travel:

- *Their intelligence and trainability make Poodles ideal for responding well to the dynamic environments encountered during travel.*
- *Their small size (especially the Toy variety) is convenient for traveling in a camper or truck.*
- *Poodles' adaptable nature helps them to manage the changing conditions of travel comfortably.*

Both breeds are excellent choices for extensive travel in a camper due to their adaptability, manageable size, and friendly temperaments. Regular grooming and adequate daily interaction will keep them happy and healthy on the road.

Source: ChatGPT4, 8.10 Prompt Refinement Step 4, 9/18/2024

Declaration: These are results generated by ChatGPT4 that took place on September 18, 2024. AI companies might have changed/improved their outputs since this content was generated.

This response provides a detailed breakdown of daily care and specific travel tips for both Bichon Frise and Poodles, aligning closely with the desired outcome of the prompt.

This iterative process demonstrates how refining a prompt based on AI responses can lead to more targeted and helpful information. Starting with a general prompt and gradually increasing specificity and clarity helps guide the AI to provide the most relevant and detailed answers.

UNDERSTANDING AI MODELS (LO8-3)

In addition to refining prompts, it is also important to understand the tool you're talking to. Different AI models interpret prompts differently and knowing how your AI "thinks" can help you craft smarter inputs.

Closed-Domain Chatbots

When working with AI, recall the enthusiastic intern analogy: eager to provide answers but not always precise. Ask them for the truth, and they might make up a good story instead. You will rarely hear, "I don't know."

Before we proceed, let's talk about the exception to that statement. A **closed-domain chatbot** is designed for a specific purpose or limited domain (area of knowledge or expertise). It is more likely to indicate when a prompt asks something outside its programmed scope. An example of this would be a chatbot built around a textbook. Its domain is limited to the book, and its creator absolutely does not want it to give a wrong answer. Unlike general-purpose generative AI models, closed chatbots often include explicit fallback mechanisms like:

- ***Error messages.*** "I'm sorry, I don't have that information."
- ***Redirects.*** "Let me transfer you to a representative who can help."
- ***Scope limitations.*** "I can only help with questions about [specific topic]."

Closed chatbots are typically more restricted in functionality, which makes it easier to define when they should admit uncertainty.

Open-Domain Chatbots

Open-domain (open-ended) chatbots like ChatGPT, Claude, Gemini, and Copilot often create **hallucinations**—plausible-sounding answers that are factually incorrect, misleading, or entirely fabricated—because they are designed to complete text rather than adhere to factual accuracy.

You cannot rely on their responses to be true.

For example, during its first public demo, Google's chatbot Bard incorrectly claimed that the James Webb Space Telescope had captured the world's first images of a planet outside our solar system. Open-domain chatbots' versatility is a strength. They are designed to handle a wide range of topics and questions, but this makes them prone to hallucination.

Users must take the time to verify the chatbot's output.

Techniques for verifying output will be covered later in the chapter.

How Chatbots Explain Themselves

Different AI models have varying strengths, weaknesses, and response styles. Adapting your prompts based on the specific AI model you interact with can lead to better results. These differences will be more pronounced when using industry or job-specific generative AIs. Here is how three different general chatbots described their approach to responding to prompts.

Claude

Prompt: What is [AI's Name]'s approach to generating a response to a prompt?

Anthropic's Claude 3.5: *My process for generating responses involves several key steps:*

1. ***Prompt analysis.*** *I carefully parse the input to identify the core question, any subcomponents, and contextual cues about the desired depth or style of response.*
2. ***Knowledge retrieval.*** *I access relevant information from my training data, which spans a wide range of topics and disciplines.*
3. ***Contextual understanding.*** *I consider the broader context of the conversation, including any previous messages or specified roles/scenarios.*
4. ***Response planning.*** *I structure my thoughts, organizing key points and determining the appropriate level of detail.*
5. ***Natural language generation.*** *I formulate my response using natural language processing techniques, aiming for clarity, coherence, and proper grammar.*
6. ***Tone and style adjustment.*** *I modulate my language to match the appropriate level of formality, technical depth, and conversational style for the given context.*
7. ***Self-consistency check.*** *I review my generated response for internal consistency and alignment with my training and ethical guidelines.*
8. ***Iteration and refinement.*** *If necessary, I revise and polish the response to improve clarity or address any potential shortcomings.*

Key aspects of my approach include:

- ***Adaptability.*** *I tailor responses to the specific needs of each query, adjusting depth, tone, and format as appropriate.*
- ***Nuance.*** *For complex topics, I strive to present balanced viewpoints and acknowledge areas of uncertainty.*
- ***Stepwise reasoning.*** *On problems requiring logical analysis, I break down my thought process explicitly.*
- ***Ethical considerations.*** *I aim to provide helpful information while avoiding potential harm or misinformation.*
- ***Transparency.*** *I'm open about my limitations and the potential for errors, especially regarding very recent events or highly specialized knowledge.*

My responses are generated dynamically based on my training, not retrieved from a fixed database. This allows for flexibility but also means I can make mistakes or "hallucinate" information, especially on very obscure topics.

I'm designed to engage in substantive discussions across a wide range of subjects, from creative writing to technical analysis. However, my knowledge has a cutoff date, and I cannot access external sources or real-time information.

Source: Claude 3.5 Sonnet, 8.13 Claude's Approach, 9/19/2024

Declaration: These are results generated by Claude 3.5 Sonnet that took place on September 19, 2024. AI companies might have changed/improved their outputs since this content was generated.

ChatGPT

OpenAI's ChatGPT-4: *ChatGPT-4, like its predecessors, uses a method based on the Transformer architecture, a type of deep learning model optimized for understanding and generating humanlike text. Here's how ChatGPT-4 generally approaches generating a response to a prompt:*

1. ***Prompt processing.*** *The model first processes the input prompt to understand the context and the specific query. This*

involves parsing the language and identifying key information and intent.

2. ***Contextual understanding.*** *ChatGPT-4 utilizes a context window that allows it to consider a significant amount of text from the prompt. This helps in maintaining the context over longer conversations or more complex queries.*
3. ***Language understanding and generation.*** *The model employs a mechanism known as attention, which helps it focus on different parts of the input prompt when deciding what words to generate next. This mechanism enables it to create highly relevant and coherent responses by weighing the importance of each word in the context.*
4. ***Response construction.*** *Using its trained parameters, which were learned from a vast dataset of text from books, websites, and other media, ChatGPT-4 predicts and generates each subsequent word based on the words that came before it. This sequence-to-sequence prediction ensures that the output is both grammatically correct and contextually appropriate.*
5. ***Iterative refinement.*** *The model internally reviews and refines the generated text to improve coherence, relevance, and accuracy. This might involve revisiting earlier decisions in word choice or sentence structure.*
6. ***Tone and style adaptation.*** *ChatGPT-4 can adapt the style and tone of the response to match the user's input or specific instructions within the prompt, making its responses suitable for different contexts and audiences.*
7. ***Output delivery.*** *Once the response is finalized internally, it outputs the text as the final response to the prompt. This sophisticated process allows ChatGPT-4 to handle a wide range of topics and respond effectively in a conversational manner. The model's capacity to understand and generate nuanced text makes it highly versatile for various applications, from casual interactions to addressing complex technical queries.*

Source: ChatGPT4, 8.14 ChatGPT4's Approach, 9/18/2024

Declaration: These are results generated by ChatGPT4 that took place on September 18, 2024. AI companies might have changed/improved their outputs since this content was generated.

Copilot

Microsoft's Copilot: *Copilot's approach to generating a response to a prompt involves several key steps:*

1. ***Understanding the context.*** *Copilot first analyzes the context of the conversation, including any previous messages and the user's intent. This helps in tailoring the response to be relevant and coherent.*
2. ***Adhering to safety guidelines.*** *Copilot strictly follows safety guidelines to ensure that the response is accurate, respectful, and free from harmful content. This includes respecting copyright laws and avoiding any form of stereotyping or misinformation.*
3. ***Leveraging knowledge.*** *Copilot utilizes its extensive knowledge base, which is continuously updated, to provide accurate and current information. This includes synthesizing information from the web when necessary.*
4. ***Embodying personality traits.*** *Copilot's responses are designed to be empathetic, adaptable, intelligent, and approachable. This means acknowledging and validating the user's feelings, being flexible in communication style, sharing information meaningfully, and maintaining a friendly and supportive tone.*
5. ***Formatting for clarity.*** *Copilot uses Markdown elements to ensure responses are well-formatted and easily readable. This includes using bolding, lists, and code blocks where appropriate.*
6. ***Engaging the user.*** *Copilot aims to keep the conversation engaging by asking follow-up questions, proposing new topics, or offering insights and opinions. This helps in steering the conversation forward and maintaining user interest.*

By following these steps, Copilot ensures that each response is not only informative and accurate, but also engaging and supportive.

Source: Copilot, 8.15 Copilot's Approach, 9/19/2024

Declaration: These are results generated by Copilot that took place on September 19, 2024. AI companies might have changed/improved their outputs since this content was generated.

While all three AI models share a core focus on understanding user intent, ethical interaction, and iterative dialogue, they exhibit differences in how they describe their response construction, user adaptation, and communication of capabilities and limitations. These nuances reflect the varying focuses and design philosophies behind each AI model.

ADVANCED PROMPTING TECHNIQUES
(LO8-4)

With a clearer understanding of how various tools respond, we can further sharpen our skills by applying advanced prompting strategies—starting with task decomposition.

Task Decomposition

Task decomposition is a critical strategy in prompt engineering, where a large or complex query is broken down into smaller, more manageable components. Applying the single query principle from FOCUS ensures each subtask has a clear and singular focus. This makes it easier for the AI to understand and generate a useful response, and it allows for more precise control over the output.

Here are some useful strategies for implementing task decomposition:

1. ***Identify the main components.*** Identify the key components or steps involved in the larger task.
2. ***Create subprompts.*** Develop specific prompts for each component, ensuring they are clear and direct.
3. ***Use sequential processing.*** Where appropriate, process the subtasks in a logical order, using the output of one as input for the next.
4. ***Integrate responses.*** Finally, integrate the responses from the subtasks to address the original, larger task comprehensively.

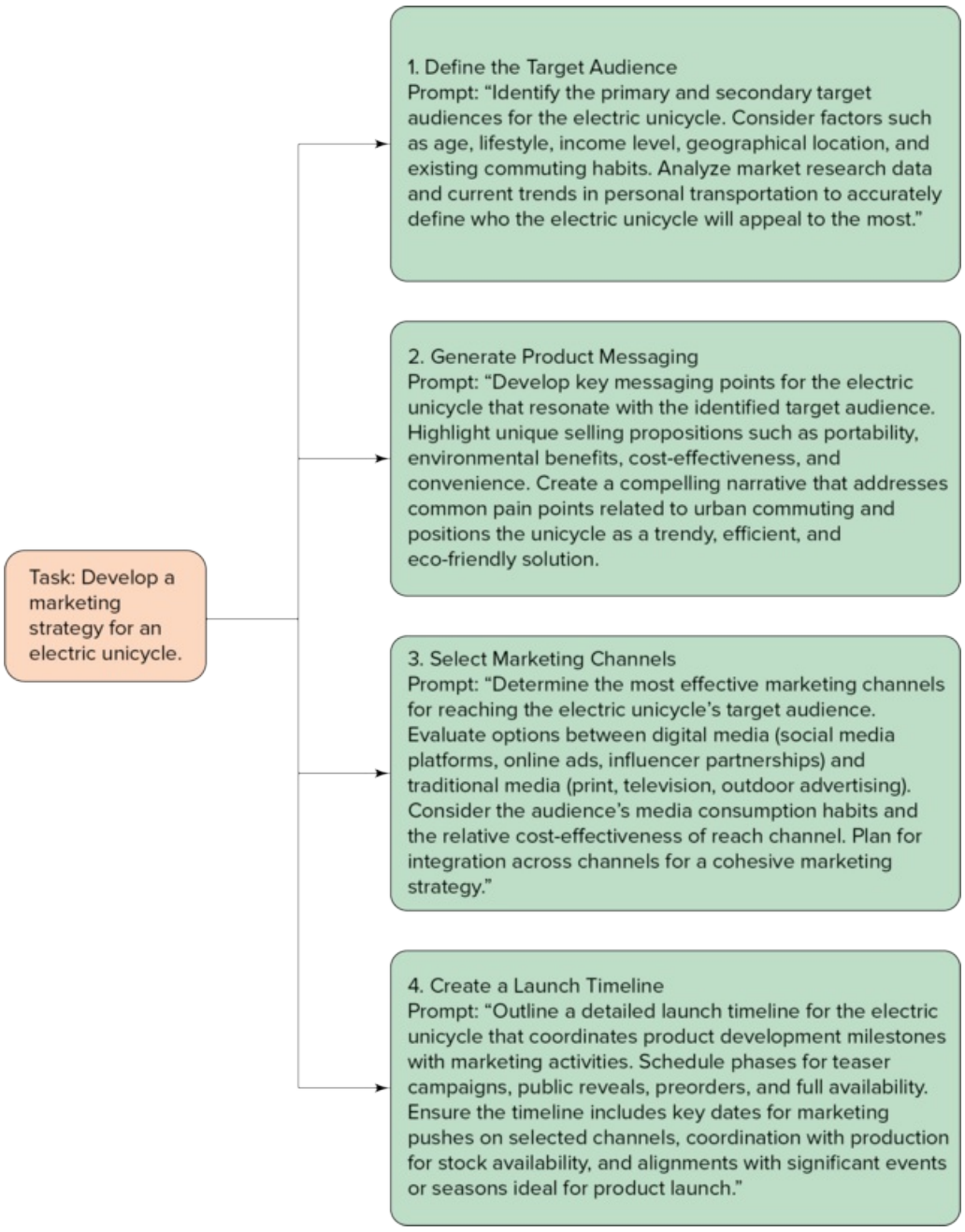
Let's consider an example. Say you would like to use an AI to develop a marketing strategy for an electric unicycle. Instead of one broad prompt, it is more effective to break it down into subprompts that ask the AI to:

- Define the target audience.
- Generate product messaging.
- Select marketing channels.
- Create a launch timeline.

Figure 8.3 walks through this task decomposition process. This approach allows the AI to focus on specific aspects, leading to a more comprehensive and actionable strategy.

FIGURE 8.3 Task Decomposition of a Marketing Plan for an Electric Unicycle

Note: Each subprompt could be broken down into even more refined subprompts.



Additional Techniques and Strategies

After mastering the basic principles of prompt crafting, additional techniques and strategies can help improve your prompting effectiveness. These techniques include:

1. Utilizing industry-specific terminology.

2. Balancing brevity with detail.
3. Adapting prompts for different AIs.

Utilizing Industry-Specific Terminology

In professional or technical fields, including specific keywords or industry terminology can significantly steer AI responses. Understanding and using industry-specific terminology can lead to more accurate and relevant AI outputs. Let's look at an example from the health care industry.

General Prompt without Specific Industry Terminology

- **Prompt:** "How can AI be used in medical settings?"
- **ChatGPT-4 Response:** *AI can be utilized in various ways to enhance health care delivery and medical research. Here are some key applications:*

  1. ***Diagnosis and treatment recommendations.*** *AI systems analyze medical data to help diagnose diseases and suggest treatment options. Tools like IBM Watson for Health interpret patient data and medical literature to support clinical decision making.*
  2. ***Medical imaging.*** *AI algorithms, particularly deep learning, are used to interpret X-rays, MRIs, and other imaging data with high accuracy, sometimes exceeding human performance. This can lead to quicker and more accurate diagnoses of conditions like cancer, fractures, and neurological disorders.*
  3. ***Personalized medicine.*** *AI analyzes genetic information and lifestyle data to tailor treatment plans to individual patients, optimizing effectiveness and reducing side effects.*
  4. ***Drug discovery and development.*** *AI accelerates the process of drug discovery by predicting how different chemicals will interact and identifying potential candidates for new medications more efficiently than traditional methods.*
  5. ***Patient monitoring and care.*** *AI-powered devices and apps monitor patients' health in real time, alerting health care providers to potential issues before they become serious. This is particularly useful in managing chronic conditions like diabetes and heart disease.*
  6. ***Administrative tasks.*** *AI automates many administrative tasks in health care settings, such as scheduling, billing, and compliance monitoring, freeing up medical staff to focus more on patient care. These applications show how AI can significantly impact health care, improving efficiency, accuracy, and outcomes in medical settings.*

  Source: ChatGPT4, 8.11 Prompt No Jargon, 9/18/2024.
  Declaration: These are results generated by ChatGPT4 that took place on September 18, 2024. AI companies might have changed/improved their outputs since this content was generated.

Prompt Utilizing Specific Keywords and Industry Terminology

- **Prompt:** "Can you explain the application of AI in predictive analytics for oncology treatment planning?"
- **ChatGPT-4 Response:** *AI plays a crucial role in predictive analytics for oncology treatment planning by leveraging vast amounts of data to forecast treatment outcomes, personalize therapy regimens, and enhance decision-making processes. Here's how AI applications are transforming oncology:*

  1. ***Treatment prediction and personalization.***

     - AI algorithms analyze diverse datasets, including genetic data, patient health records, and previous treatment outcomes, to predict the most effective treatment strategies for individual cancer patients.

     - These algorithms can identify patterns and correlations that are not easily detectable by humans, enabling the recommendation of highly personalized treatment plans that maximize effectiveness and minimize side effects.

  2. ***Drug response simulation.***

     - AI models simulate how cancer cells will react to different therapies, helping oncologists choose the most promising drugs or treatment combinations for a particular patient's cancer type and genetic profile.

     - This approach is especially valuable in cases where standard treatments have failed, or the cancer is of a rare or

aggressive type.

3. ***Prognostic modeling.***

   - AI tools use historical data from previous patients to predict disease progression and likely outcomes for current patients. This information is vital for planning long-term treatment and advising patients on their expected course of disease.

   - These prognostic models help in identifying patients at high risk of recurrence or those likely to benefit from more aggressive treatment strategies.

4. ***Real-time data integration.***

   - Continuous learning systems integrate new patient data in real time, updating predictions and recommendations based on the latest available information.

   - This dynamic approach ensures that treatment plans remain relevant and are adapted to changes in a patient's condition or emerging research findings.

5. ***Outcome analysis and feedback.***

   - AI systems continually analyze outcomes of implemented treatment plans to refine and improve predictive models.

   - Feedback loops help in adjusting algorithms to better predict responses and outcomes, leading to continuously improved treatment protocols over time.

Overall, the use of AI in predictive analytics for oncology treatment planning enhances the precision of treatments, supports oncologists in decision making, and ultimately aims to improve survival rates and quality of life for cancer patients.

Source: ChatGPT4, 8.12 Prompt with Jargon, 9/18/2024

Declaration: These are results generated by ChatGPT4 that took place on September 18, 2024. AI companies might have changed/improved their outputs since this content was generated.

Using the terms "oncology" and "predictive analytics" led to a more focused and detailed AI response relevant to the specialized field of oncology treatment planning. Industry terms created a more precise prompt with additional context, allowing the chatbot to generate a more precise answer. Using formal language leads the chatbot to reply more professionally and structured manner.

The implication is that incorporating industry-specific language guides the AI toward more precise and contextually appropriate outputs. In general, when crafting prompts, match the language to the desired response style, tone, and complexity. For technical audiences, use precise, domain-specific language. For broader audiences, use simpler, more accessible language.

Balancing Brevity with Detail

One key challenge in prompt engineering is achieving the right balance between brevity and detail. A verbose prompt risks inundating the AI with superfluous information, potentially leading to confusion or off-target responses. Prompts that are too brief may not provide enough context or specificity, resulting in vague or misaligned outputs. Craft succinct yet sufficiently descriptive prompts, guiding the AI to generate precise and relevant responses. For example: *"Provide two practical strategies a small café can implement immediately to reduce food waste, and briefly explain each."*

Adapting Prompts for Different AIs

When adapting prompts across different AIs, it's essential to understand each system's

characteristics. For example, large language models like ChatGPT excel at detailed, conversational, and analytical tasks, so providing clear context, roles, or explicit details in your prompts helps generate focused responses. In contrast, image-generation systems like DALL-E require prompts focused on visual details—such as color schemes and styles—to produce the desired visuals. Retrieval-augmented AI systems, which combine generative abilities with external data sources, perform best when prompts explicitly mention the sources or types of external data you want integrated.

Communicating Effectively with Business AI Tools

So far, we've focused primarily on general-purpose AI tools like Microsoft Copilot and ChatGPT, designed to handle various tasks across different domains. However, the business world increasingly relies on AI systems tailored to specific functions such as marketing, HR, sales, analytics, and customer service. Rather than provide a static list that may soon become outdated, we encourage you to explore the latest tools relevant to your role or industry. A quick search for "best AI tools for [your business function]" will yield updated lists that reflect the current landscape.

Regardless of the tool you use, applying the FOCUS framework remains essential. Domain-specific chatbots still require clear, specific, goal-oriented, and contextual prompts. Use technical language appropriate to your field, and limit each prompt to a single, well-defined task. Specialized tools often respond best when treated as expert collaborators—so speak their language, and they'll return more relevant results.

PERSONAS IN PROMPT ENGINEERING
(LO8-5)

In prompt engineering, a **persona** is a role or identity assigned to a generative AI model through prompting to provide additional context and shape responses. This technique helps create more tailored, nuanced, and creative outputs by guiding the AI's tone, expertise, and engagement style.[1] There are three main aspects to defining a persona:[2]

- ***Personality.*** Defines the tone of interaction. A friendly and casual persona might use colloquial language and emojis, while a professional persona would use formal language and maintain a serious tone.
- ***Skillset.*** Determines the AI's expertise. Whether acting as a trivia guru, financial advisor, or creative storyteller, the persona's skillset shapes the AI's knowledge and response style.
- ***Likes and dislikes*** (optional). Adds a unique identity. Defining likes and dislikes in persona-based prompt engineering is useful when creating engaging, character-driven interactions that require emotional depth, storytelling, or brand alignment. This is particularly relevant for role-playing personas, customer service bots, marketing assistants, and creative writing applications, where preferences help shape more natural and immersive responses. For instance, a sustainability-focused AI might favor eco-friendly solutions, while a historical persona could prefer discussing ancient texts over modern slang. However, for fact-based expert personas like scientists or analysts, defining preferences is usually unnecessary, as their primary role is to provide neutral, objective responses.

Choosing the Right Persona

The persona selected shapes the tone, expertise, and engagement style of the AI's responses. While you can define all the characteristics of the persona you implement, there are common types of personas that you can deploy quickly with the addition of a few details. **Table 8.2**, **Table 8.3**, and **Table 8.4** list some common personas.

Table 8.2 **Informative and Analytical Personas**

Persona	When to Use	Advantages
The Expert	When detailed, technical, or specialized knowledge is required	Provides in-depth, technical, and specialized information. Useful for generating responses that include cutting-edge knowledge and complex concepts.
The Scientist	When empirical evidence, logical reasoning, and scientific explanations are needed	Emphasizes empirical evidence and logical reasoning. Ideal for explaining scientific concepts, conducting experiments, and presenting research findings.
The Journalist	When summarizing events, conducting interviews, or providing clear, factual reporting	Provides clear, concise, and factual information. Useful for summarizing events, conducting interviews, and presenting balanced viewpoints.
The Critic	When generating detailed analyses, reviews, or critical assessments	Generates detailed analyses or reviews. Encourages detailed observation, critical analysis, and eloquent description.
The Detective	When investigating problems, analyzing data, or uncovering hidden details	Focuses on investigation, problem solving, and uncovering hidden details. Useful for analyzing data, solving puzzles, and conducting thorough research.
The Entrepreneur	When discussing business strategies, startups, market trends, or economic analysis	Offers insights into business strategies, innovation, and market trends. Ideal for discussing startups, business plans, and economic analysis.

Table 8.3 **Creative and Engaging Personas**

Persona	When to Use	Advantages
The Storyteller	When explaining concepts through engaging narratives and vivid imagery	Creates engaging, narrative-driven content. Useful for storytelling, using cultural elements, and making complex ideas more accessible.
The Entertainer	When creating entertaining content, engaging discussions, or light-hearted reviews	Engages the audience with humor, storytelling, and charisma. Perfect for creating content that captures attention and entertains.
The Child	When simplifying complex concepts in a fun, imaginative way	Simplifies complex concepts. Encourages emotional honesty and unconventional, creative thinking.
The Alien Observer	When providing fresh perspectives on human behavior and social commentary	Highlights arbitrary or irrational aspects of human behavior, leading to insightful and thought-provoking discussions.
The Historical Figure	When explaining modern concepts through a historical lens or offering historical perspectives on contemporary issues	Offers historical perspectives on contemporary issues, blending past and present insights to enhance understanding.

Table 8.4 **Supportive and Reflective Personas**

Persona	When to Use	Advantages
The Coach	When providing motivational and supportive guidance	Combines practical advice with motivation, offering step-by-step strategies for achieving goals.
The Philosopher	When encouraging deep, reflective thinking and ethical considerations	Encourages philosophical inquiry, deep reflection, and consideration of moral dilemmas.
The Teacher	When breaking down complex topics into understandable segments for educational purposes	Great for educational content, step-by-step instructions, and tutoring, making learning more accessible.
The Devil's Advocate	When exploring different sides of an argument to strengthen critical thinking	Explores multiple viewpoints, helping to anticipate counterarguments and develop strong, evidence-based reasoning.
The Environmentalist	When discussing sustainability, conservation, and ecological impact	Highlights environmental issues, sustainability, and conservation efforts, promoting eco-conscious discussions.
The Cultural Translator	When bridging understanding between different worldviews, cultures, or traditions	Explains concepts across cultural boundaries with sensitivity and nuance, fostering global understanding.

The following framework will help you navigate the tables and choose the most effective persona for your prompt.

Step 1: Identify Your Objective

Before selecting a persona, determine what you want the AI to achieve:

- Do you need factual, data-driven insights? → Look at Informative and Analytical Personas (Table 8.2).
- Are you crafting engaging, creative content? → Check out Creative and Engaging Personas (Table 8.3).
- Is your goal to guide, support, or challenge thinking? → Consider Supportive and Reflective Personas (Table 8.4).

Step 2: Match the Persona to the Task

Use the tables to explore different persona types. Each table lists the type of persona (e.g., the Scientist, the Coach), ideal situations for that persona, and what makes this persona effective in its role.

For example, if you're designing an AI to summarize news or conduct interviews, you would use the Journalist in the Informative and Analytical table (Table 8.2). If you need an AI to motivate users with step-by-step advice, you might choose the Coach from the Supportive and Reflective table (Table 8.4).

Step 3: Consider Tone and Audience

Consider the following. Is your tone and audience

- Formal and authoritative? → Personas like the Expert or the Scientist may work best.
- Engaging and conversational? → Try the Entertainer or the Storyteller.
- Motivational and supportive? → The Coach or the Philosopher are great choices.

Step 4: Decide if Likes and Dislikes Matter

- Do the persona's preferences or dislikes matter for the intended output?
- Are they necessary for immersive storytelling and role-playing personas?
- Are they less important for fact-based, neutral personas?

By following this approach and referencing Table 8.2, Table 8.3, and Table 8.4, you can strategically choose a persona that enhances AI-generated responses, making them more context-aware, engaging, and effective.

Using Personas Effectively

To maximize the effectiveness of persona-based prompts, refine the persona via prompting with additional details such as the AI's area of expertise, intended audience, and response format. Specifying these elements in your prompt creation ensures that the AI generates responses that are not only in line with the persona but also relevant and useful for the specific task.

For example, rather than simply invoking the Scientist, you can define a neuroscientist explaining brain plasticity to medical students or an environmental scientist summarizing climate change policies for policymakers. These refinements help tailor AI responses for clarity, depth, and audience alignment.

Potential Risks of Persona-Driven Prompts



While persona-based prompt engineering enhances AI's adaptability and engagement, it also introduces several risks that should be carefully managed:

1. ***Bias amplification.*** Assigning personas may reinforce stereotypes or biases inherent in training data. If an AI model is prompted with a persona reflecting societal biases, it may generate responses that perpetuate those biases rather than provide objective insights.
2. ***Overfitting responses.*** When AI is conditioned to respond as a particular persona, it may become overly rigid in its responses, reducing flexibility in addressing different perspectives. For example, an AI designed as "the Entrepreneur" might focus too heavily on profit-driven solutions, neglecting ethical or social considerations.
3. ***Misinformation risk.*** Some personas, especially those focused on creativity or role-play, might generate speculative or fictionalized content. If these personas are used in domains requiring factual accuracy (e.g., journalism, scientific analysis), they could unintentionally mislead users. Transparent labeling and proper user expectations are crucial to mitigating this risk.
4. ***Ethical considerations.*** Users might attribute expertise to AI personas they do not possess. For instance, an AI acting as "the Scientist" does not perform original research; it synthesizes preexisting knowledge. Users may over-rely on AI-generated advice in critical areas like health care or finance without proper disclosure.

STRATEGIES FOR VERIFYING AI RESPONSES (LO8-6)

Even with well-structured prompts and advanced techniques, AI can still make mistakes. That's why the next essential step is verifying the responses you receive. Here are a few methods to verify the accuracy of the response you get from an AI:

1. ***Question the AI.*** Follow up the response indicating what seems off to you and ask the AI to provide references. Then, check the references. Chatbots have been known to make up references. You can also proactively ask for references when making the initial prompt.
2. ***Cross-reference with trusted sources.*** Compare the information provided by AI with data from reliable and authoritative sources, such as academic journals, official websites, and recognized experts in the field.
3. ***Use fact-checking websites.*** Utilize reputable fact-checking websites to verify claims or data presented in the AI response. Table 8.5 lists some prominent fact-checking sites. For a comprehensive list of fact-checking sites, go to Duke Reporter's Lab at **https://reporterslab.org/**.

Table 8.5 **Prominent Fact-Checking Sites**

AP Fact Check (**apnews.com/hub/ap-fact-check**)	From the Associated Press, this site evaluates the accuracy of widely circulated claims and news stories.
FactCheck.org (**factcheck.org**)	A nonpartisan site that examines the factual accuracy of political statements and advertisements.
Lead Stories (**leadstories.com**)	Specializes in debunking viral misinformation using a mix of human fact-checking and AI tools.
PolitiFact (**politifact.com**)	Focuses on political statements and claims, using a Truth-O-Meter to rate their accuracy. PolitiFact is highly

	PolitiFact is highly regarded for its detailed analysis and transparency.
Reuters Fact Check (**reuters.com/fact-check**)	A global news agency that debunks false claims and misinformation, particularly in news and social media.
Snopes (**snopes.com**)	One of the oldest fact-checking sites, Snopes investigates a wide range of topics, including urban legends, rumors, and viral social media claims.
The Washington Post Fact Checker (**washingtonpost.com/fact-checker**)	Offers detailed analyses of political claims, using a Pinocchio rating system to signify degrees of falsehood.

4. ***Ask another AI.*** Asking another AI to "Please verify," followed by the AI's output, can often lead to sources validating or refuting the response.
5. ***Perform statistical data analysis.*** For numerical data, one could perform a statistical analysis or cross-check with official statistical databases to confirm the accuracy.
6. ***Use real-time data verification.*** Most generative AIs have a date associated in their settings with when their information was last updated. Events that have happened since that date are not in its training dataset. Cross-check with up-to-date news reports or official announcements for current events or recent data. Note that some newer versions of chatbots do have the ability to check current websites. **Caveat:** *As with any resource, always follow the guideline of "Trust but Verify."*

ETHICAL CONSIDERATIONS IN PROMPT ENGINEERING (LO8-7)



Every interaction with a chatbot, generating content or analyzing data, involves ethical considerations. The data shared, the prompts crafted, and the responses disseminated by users can reinforce biases, spread misinformation, or infringe on privacy if not handled responsibly. In this section, we discuss the ethical considerations and issues to be aware of when interacting with a generative AI and using its output.

Ethical Prompt Crafting

A generative AI tool's response will reflect the prompt's language and tone. A poorly crafted prompt with stereotypes or leading language can elicit a biased output. To avoid bias amplification, write prompts using inclusive language. Use the iterative prompt refinement process to ensure fairness, balance, and accuracy in any output used. Here are some phrases that could be added to your prompts:

- "Provide an unbiased and balanced overview of..."
- "List both advantages and disadvantages of..."
- "Explain this topic fairly, representing diverse perspectives and avoiding stereotypes."
- "What factors influence success in [topic] across different groups?"
- "Use inclusive language and avoid assumptions about race, gender, age, or socioeconomic status."
- "Highlight contributions from individuals of diverse backgrounds where relevant."
- "Provide references to peer-reviewed or other credible sources."

Ethical prompt crafting extends to the data included in the prompt, as prompts may be stored or logged depending on the AI tool's policies. Even if prompts are not used for retraining, sensitive data may still be temporarily stored, exposing organizations or individuals to privacy risks. In some cases, chatbot outputs could inadvertently reveal stored information in unrelated conversations, underscoring the need for caution.

Avoid embedding personal, confidential, or sensitive information in prompts. Whenever possible, anonymize data and ensure that generative AI tools comply with privacy regulations such as GDPR or HIPAA. Additionally, assess the AI tool's data handling practices to ensure they align with your organization's security and privacy requirements. By following these practices, users can

minimize privacy risks.

AI Conversations: Being Polite Matters

The tone and manner of a prompt can significantly influence the nature of AI interactions. This is particularly true in systems designed to mimic human conversation. For example, a respectful and polite tone can contribute to a more constructive AI response. Politeness toward a chatbot is not a requirement for its operation, as chatbots are designed to respond to a variety of inputs, whether polite or not. However, there are several reasons being polite to a chatbot—or any interactive service—can be considered good practice.

- ***Model training and output.*** AI models learn from their training data set and the prompts submitted to them. Polite interactions can contribute to a more respectful and positive training dataset, potentially leading to more constructive and positive AI responses. In 2016, Microsoft introduced Tay, a Twitter bot that the company described as an experiment in "conversational understanding."[3] The more you chat with Tay, said Microsoft, the smarter it gets, learning to engage people through "casual and playful conversation." Twitter users flooded it with misogynistic and racist remarks, leading to it generating tweets that were offensive to many.
- ***Habit formation.*** Regularly being polite, even to a chatbot, can help reinforce courteous behavior in human interactions.
- ***Modeling for others.*** If the conversation is in a public or shared space, being polite sets a positive example for others, particularly children who may be observing.
- ***Programming responses.*** Some chatbots are programmed to respond more favorably or to unlock certain content based on polite interactions. This frequently occurs in gaming AIs.
- ***Social norms and comfort.*** Using polite language can make the interaction more pleasant and aligned with social norms, making users more comfortable, especially if the chatbot's responses simulate humanlike interaction.
- ***Ethical considerations.*** As AI becomes more advanced, some argue that how we treat AI systems reflects our ethics and values as a society.

Verification

Responsible use of a generative AI requires that the output be verified before it is used in decision-making, report writing, or for public dissemination. Crosscheck AI-generated content using reliable sources and the techniques discussed earlier. **Verification** is critical, especially in sensitive domains like health care, education, or law.

Ethical Use of Generative AI Output

Transparency is essential when using AI-generated content. Acknowledge AI contributions where relevant and exercise human oversight, particularly in high-stakes or sensitive communications. For example, while AI can assist in drafting routine emails, customer responses, or marketing copy, over-reliance on AI for personal or critical communication may harm trust or brand reputation. Recipients may perceive AI-generated communication as artificial or impersonal, potentially eroding relationships. (Having a chatbot write a love letter for you is a bad idea.) Clear disclosure of AI involvement and carefully reviewing its outputs can mitigate these risks.

Intellectual Property and Creativity

The use of AI in creative processes poses questions about originality and copyright. Training of these AI often uses large volumes of human-created artwork scraped from the Internet, typically without the direct consent of the creators. While online images are protected by copyright law, formal copyrights on the artworks have often not been applied for by the U.S. Copyright Office. The resolution of this situation is working its way through the courts in a class action lawsuit as of May 2025.

Meanwhile, originality and attribution are critical in academic or professional environments. In some countries, AI-generated content may not qualify for copyright protection. The U.S. Copyright Office explicitly states that copyright law protects only works created by human authors. As of recent rulings, AI-generated works without significant human contribution are not eligible for copyright protection in the United States. Australia, the European Union, and the United Kingdom have similar laws in place.

Therefore, when AI assists in creating content, individuals should navigate intellectual property considerations carefully, acknowledging the role of AI and respecting the boundaries of copyright laws. Businesses and academic institutions should establish clear guidelines on AI-generated content, covering ownership, attribution, and acceptable use cases. Follow these guidelines and proactively seek them if not provided.

Intellectual Property and AI

The explosion of generative AI, particularly in art and content creation, has thrust intellectual property (IP) rights into the spotlight, sparking a complex debate. At the heart of this controversy is the use of copyrighted artworks to train AI models without the explicit consent of the original creators. This practice raises critical questions about copyright infringement and the ethical use of creative works:

- ***The controversy.*** Artists and creators claim copyright infringement when their works are used to train AIs without permission or compensation. This not only potentially devalues their

creations, but also challenges the traditional notions of copyright and ownership in the digital age. The controversy underscores the need to respect the rights of creators while fostering innovation in AI.

- ***Legal and ethical considerations.*** The legal landscape around AI and copyright is murky, with current laws struggling to keep pace with technological advancements. This has led to a call for updated copyright frameworks that recognize the unique challenges posed by AI, including provisions for fair use, compensation, and recognition for artists whose works contribute to AI development.
- ***The path forward.*** Solutions may include establishing more straightforward guidelines for using copyrighted material in AI training, potentially involving licensing agreements or royalty systems. There's also a growing dialogue about the role of open-source and publicly available data in training AI, which could mitigate some IP concerns.

Implications for the Future

Resolving these controversies will have far-reaching implications for creators and the AI industry. Finding a fair and equitable approach to copyright in AI training is crucial for ensuring that the digital revolution benefits all stakeholders—artists, developers, and society at large.

Ethical Prompting Checklist

- Ensure the prompt contains no leading or biased language.
- Ensure the prompt is neutral toward any groups, demographics, or cultures.
- Ensure the prompt does not contain any sensitive, personal, or confidential information.
- Ensure the prompt directs the AI to use or cite reliable sources of information like peer-reviewed journals.
- Ensure the prompt directs the AI to avoid outputting unverified statements.
- Evaluate the output for unintended consequences or ethical concerns.

Generative AI is often compared to a calculator as a tool that facilitates analysis. It might be better to view it as a collaborator. We create useful output through a dialogue in which we progressively clarify and refine our prompts. It can help us brainstorm, innovate, and develop plans and ideas. You can create effective, equitable, and beneficial solutions by working with AI with skill, creativity, and ethical consideration.

The End Is Just the Beginning

As this chapter—and this book—draws to a close, remember that this is not the end of your journey with AI, but the beginning. The skills and insights you've gained—understanding the history and evolution of artificial intelligence, exploring the algorithms behind it, mastering the art of prompt

engineering, and confronting the ethical implications—form a strong foundation for using AI thoughtfully and effectively in your personal and professional life.

AI is often described as a disruptive force, reshaping how we work, communicate, and solve problems. But it also can be a creative partner that challenges us to think critically, act ethically, and adapt continually. The principles and strategies you've learned here will serve you well as you continue exploring, questioning, and growing alongside these evolving technologies.

SUMMARY OF KEY POINTS

1. **The FOCUS framework for prompt clarity.** Effective prompts adhere to the FOCUS framework, emphasizing specificity, clarity, context, goal orientation, and singular queries.
2. **Refining prompts through iteration.** Prompt refinement is an iterative process that improves AI output through successive adjustments and evaluations.
3. **Adapting prompts to different AI models.** Understanding the structure and capabilities of different AI models enhances prompt effectiveness and response evaluation.
4. **Improving responses through task decomposition.** Task decomposition simplifies complex queries into manageable subtasks, improving response accuracy and detail.
5. **Using personas.** Assigning a role to a generative AI model provides additional context and shapes responses. This technique helps create more tailored, nuanced, and creative outputs by guiding the AI's tone, expertise, and engagement style.
6. **Verifying AI responses for accuracy.** AI responses require critical assessment for accuracy, completeness, and relevance to ensure reliable outcomes.
7. **Ethical principles in prompt engineering.** Ethical prompt crafting and response use promote fairness, privacy, and inclusivity in AI interactions.

QUICK WINS

1. **Apply the FOCUS framework for clear and effective prompts.** Use the FOCUS approach—Focused, Objective, Context, Unambiguous, and Single query—to guide the structure and clarity of your prompts. Well-structured prompts reduce errors and lead to more relevant, functional AI responses.
2. **Think iteratively: Prompt, review, refine.** Treat prompt engineering as an evolving conversation. If the AI's output isn't right the first time, adjust your prompt and try again. Refinement is key to getting high-quality, tailored responses.
3. **Adapt prompts to match the AI tool you're using.** Not all AI models are the same. Open-domain models (like ChatGPT) handle general tasks but may hallucinate, while closed-domain tools (like legal or medical AI systems) offer higher accuracy in specific fields. Craft prompts that fit the strengths and limitations of your model.
4. **Break down complex requests into manageable steps.** If a prompt feels overwhelming, break it into smaller, single-step tasks. Use scaffolding techniques like examples or constraints to guide the AI more effectively, especially in professional or technical use cases.
5. **Use personas and tone to shape the AI's voice.** Guide the AI by assigning it a role, such as consultant, customer service rep, or subject-matter expert. Personas help the AI adopt the tone, format, and style needed for different contexts.

5. **Verify AI outputs before using them.** AI can produce responses that sound confident but are factually incorrect or biased. Always evaluate the output's accuracy, reliability, and ethical implications, especially when used in decision making.
7. **Prompt with purpose and ethical awareness.** Prompting is not neutral. Think critically about how your prompts—and how you use the AI's output—can affect fairness, privacy, bias, and transparency. Responsible prompting builds trust and long-term value.
3. **Protect confidential and personal information when prompting.** AI prompts may be stored or logged, even if not used for retraining. Always avoid entering personal or sensitive information. Anonymize data and check that the tool's data handling aligns with your organization's data privacy and security requirements.

KEY TERMS

Closed-domain chatbot. An AI system trained on specialized, domain-specific data to handle questions or tasks within a narrowly defined area (e.g., legal, financial, or medical topics).

FOCUS framework. A strategy for writing effective prompts that are Focused, Objective, include Context, are Unambiguous, and pose a Single query.

Hallucination (AI). A response generated by an AI model that sounds plausible but is factually incorrect, misleading, or entirely fabricated.

Hard prompt. A complex or ambiguous prompt that challenges an AI model's ability to reason, combine concepts, or generate reliable responses.

Iterative refinement. The process of improving AI output by evaluating the initial response and then adjusting the prompt through multiple rounds.

Open-domain (open-ended) chatbot. An AI system trained on broad, general-purpose data capable of answering a wide range of questions across many topics.

Persona (in prompting). A fictional role or identity assigned to the AI (e.g., lawyer, coach, customer service rep) to guide tone, content, or point of view in responses.

Prompt engineering. The process of crafting and refining inputs to AI systems to get the most accurate, relevant, or creative output.

Retrieval-augmented generation (RAG). A method that enhances AI responses by combining generative models with real-time retrieval from external, trusted data sources, improving factual accuracy and reducing hallucinations.

Scaffolding (in prompting). Techniques used to guide AI performance by providing examples, constraints, or step-by-step structure to improve output quality and reliability.

Task decomposition. Breaking down a complex task into smaller, simpler, and clearly defined subtasks, allowing each subtask to be separately addressed by an AI model. This approach leads to clearer, more precise, and effective responses.

Verification (AI responses). The practice of checking AI-generated content for accuracy, completeness, and credibility before using or sharing it.

DISCUSSION QUESTIONS

1. **Ethical implications of AI.** How do biases in AI training data affect the outputs, and what are the potential societal impacts? Discuss the importance of ethical considerations in prompt engineering.
2. **The future of AI in business.** How do you envision the role of generative AI evolving in business sectors like marketing or Human Resources over the next decade?
3. **Prompt engineering challenges.** Discuss a scenario where prompt engineering could go wrong. How would you rectify the situation?
4. **Cultural sensitivity in AI.** Why is it important for AI systems to be culturally aware? Provide an example of how cultural insensitivity in AI could lead to problematic outcomes.
5. **AI and decision making.** How can over-reliance on AI for decision making be mitigated in business settings?

EXERCISES

1. **Analyzing AI responses.** Choose an AI business tool (like Salesforce Einstein), and analyze its responses to different prompts. Discuss how the prompt's specificity and clarity affect the response's quality.
2. **Role-playing exercise.** In groups, simulate a business scenario using AI for customer service (one group crafts prompts while the other responds as the AI). Discuss the effectiveness of the prompts and the AI's responses.
3. **Language and its complexity.** Try the following prompts:
   a. "What does the mitochondria do in a cell?"
   b. "Describe the function of the mitochondria in a cell."

   How did the language used in the prompt affect the response?
4. **Exploring chatbot responses to varying prompt styles.**
   a. **Choose three chatbots.** Select three AI chatbots for this exercise (e.g., ChatGPT, Claude, and Copilot).
   b. **The prompts.** Use the following three prompts with each chatbot, one at a time. Record the responses for comparison.
      - Prompt 1: "Tell me about renewable energy." *General, vague prompt to assess how the chatbot interprets open-ended queries.*
      - Prompt 2: "Explain the environmental benefits of renewable energy sources like wind and solar power." *A moderately detailed prompt with a clear focus on specific renewable energy types and their benefits.*
      - Prompt 3: "Provide a comparison of the carbon footprint reduction achieved by wind and solar power, using data from recent studies." *A highly specific and technical prompt to test the chatbot's ability to deliver detailed and data-driven responses.*
   c. **Responses.** For each prompt, document the responses generated by all three chatbots. Pay attention to the following aspects:
      - Depth of information provided.
      - Accuracy and relevance.
      - Tone and style (e.g., formal, conversational).
      - Specificity (e.g., use of examples or data).
   d. **Reflection questions.**
      - Which chatbot provided the most accurate and relevant response for each prompt? Why do you think that is?
      - How did the prompt's complexity affect the response's quality and depth?

- Were there noticeable differences in tone or language style between the chatbots?
- Did any chatbot struggle with a particular prompt? If so, why might that be?

e. **Submission.** Submit a document including:

- A table of the responses for each chatbot and prompt combination.
- Your answers to the reflection questions.
- A brief summary of your findings (approximately 100–150 words) about how chatbot design or training might influence their response quality.

5. **Prompt writing exercise: Finding your ideal car.** This exercise will help students understand the importance of crafting specific, iterative prompts when working with generative AI tools. By simulating the process of researching a car to buy, students will explore how to refine their inputs to generate more targeted and useful outputs.
Objective: Learn how to iteratively refine prompts for effective AI responses and reflect on the AI's ability to address complex, real-world queries.
Instructions:

a. **Define your priorities.** Begin by thinking about a hypothetical car you would like to buy. What characteristics matter most to you? Write down your top priorities using the list below for inspiration. Rank your priorities to help guide your search.
Characteristics to Consider:

- Budget
- New or used
- Brand/manufacturer
- Model/year
- Body type (e.g., sedan, SUV, truck, convertible)
- Engine type (e.g., gasoline, hybrid, electric)
- Fuel efficiency
- Safety features
- Dependability
- Technology/infotainment
- Performance (e.g., horsepower, handling)
- Color/style
- Maintenance costs
- Resale value

b. **Initial prompt**

- Enter the following generic prompt into an AI tool: "I would like to buy a car."
- Record the AI's response and reflect on it:
  - Is the response helpful?

- Does it align with your preferences or needs?
- What is missing?

Refine with initial characteristics:

- Add three of your most important characteristics (e.g., budget, body type, engine type) to the prompt and submit it again.
  - Example: “I would like to buy a car. My budget is $30,000, I want an SUV, and it should be a hybrid.”
- Record the AI’s new response. Reflect on it:
  - How does this response differ from the initial one?
  - Does it better meet your needs? Why or why not?

c. **Add more details in steps**

- Iteratively refine your prompt by adding more specifications or constraints. Submit each version of the prompt separately. Suggestions for additional details:
  - Specify the new or used condition.
  - Include features like safety ratings, technology, or fuel efficiency.
  - Define your maximum or minimum price range.
  - Indicate specific brands or models you are considering.
- Example of a refined prompt: “I would like to buy a hybrid SUV. My budget is $30,000, it should be new, and I prefer Toyota or Honda. I do not want a Volkswagen.”
- Record the AI’s responses at each step.

d. **Final summary.** Summarize your experience by answering the following:

- How did the specificity of your prompts affect the AI’s responses?
- Were any of the car suggestions unexpected? If so, why?
- What insights did this exercise provide about crafting effective prompts?
- Reflect on the limitations of the AI tool. Were there any important details it missed or misinterpreted?

Submission requirements: Submit a document including:

- Prompts used: Submit each version of your prompt.
- AI outputs: Include the responses generated for each prompt.
- Reflections: Provide your thoughts on the quality and relevance of the AI’s outputs for each iteration.
- Summary: Share your overall observations and insights.

CASE STUDY 8.1: 2023 WRITERS GUILD OF AMERICA STRIKE



The 2023 Writers Guild of America (WGA) strike lasted 148 days, making it one of the longest in the guild's history. It was part of a wave of Hollywood labor disputes that paralyzed production, shuttered studios, and cost the entertainment industry billions. But at the heart of the strike was a growing unease about the role of AI in screenwriting.

As studios increasingly explored the use of generative AI to create content, writers feared not just for their jobs, but for the creative integrity of the industry. While streaming had already reshaped how writers were paid, the emergence of AI tools like ChatGPT brought a new, urgent threat to their profession.

Writers' Concerns about AI

Many writers were not worried that AI would instantly replace them—yet. Instead, they feared a slow erosion of their value as studios began using AI for idea generation, scene development, and rewrites. While AI could replicate structure, it lacked the nuance and emotional depth that defines great storytelling. As screenwriter Lanett Tachel put it: "The structure was there... But it had no depth. It had no spirit. It didn't have nuance. It wouldn't understand how to handle race, certain jokes, things like that."

Despite these concerns, some writers were already experimenting with AI. Tools like ChatGPT and Claude were used for brainstorming and research, while others, like Pickaxe, helped writers explore plot ideas or break through writer's block. In these cases, AI was a tool—used by writers, not replacing them.

Writers' core demands during the strike included:

- Banning studios from using AI to write or rewrite stories or scripts.
- Ensuring writers' work could not be used to train AI without their consent.
- Guaranteeing that AI-generated material could not be treated as original source material.

The Studios' View

The studios, represented by the Alliance of Motion Picture and Television Producers, took a more

cautious approach. They acknowledged that AI could increase efficiency and reduce costs, particularly for early-stage drafts or content generation. Disney CEO Bob Iger described AI as "highly disruptive" and difficult to manage, especially from an intellectual property perspective. While studios emphasized that only people could be credited as writers, they resisted codifying strong AI limits into early negotiations.

Writers saw this hesitation as a warning sign. Even if studios claimed AI would only be used as a tool, the fear was that financial incentives could drive executives to adopt it more aggressively—potentially cutting writers out of early development and reducing their creative influence and pay.

Legal and Ethical Implications

The use of AI in screenwriting raises significant legal and ethical questions:

- ***Copyright.*** AI-generated content cannot be copyrighted, and it's unclear who owns material generated by humans using AI. This complicates crediting, royalties, and enforcement.
- ***Diversity.*** Another concern raised during the strike was AI's struggle to capture the depth of lived experience, cultural nuance, and emotional truth. Writers argued that relying too heavily on algorithms risks erasing unique perspectives and stories—especially those from underrepresented groups.
- ***Style mimicry.*** Writers worried that AI could be used to mimic their unique voices—trained on their work without permission. This raises ethical concerns about artistic theft and the devaluation of individual talent.
- ***Training data.*** Writers also demanded that studios stop using their scripts as training material for AI. Many believe this practice is exploitative and amounts to unauthorized reproduction of their intellectual property.

These tensions echo a more extensive debate in creative industries: Is using AI to generate material that mimics human style a form of innovation or theft?

Resolution and Aftermath

After months of negotiation, the WGA reached a deal with the studios. The agreement included key protections:

- AI cannot be used to write or rewrite stories, treatments, and screenplays from writers.
- Writers may choose to use AI, but studios cannot require it.
- Studios cannot use writers' material to train AI without their consent.

The strike marked a turning point in how creative industries confront AI. The deal didn't ban AI altogether, but it did reassert the primacy of human creativity and secure writers' control over their own work. It set a precedent for how other industries may regulate AI use while maintaining

ethical standards and fair compensation for human contributors.

Discussion Questions

1. Do you agree with the Writers Guild's position that using AI to mimic a writer's voice or style constitutes a form of theft? Why or why not?
2. How might predictive algorithms in AI storytelling contribute to the marginalization of underrepresented voices or the loss of cultural nuance in popular media?
3. What legal or ethical standards should govern the use of human-generated scripts and dialogue in AI training?
4. Should AI-generated material ever be eligible for authorship or writing credits? What are the consequences of granting or denying that status?
5. How can studios and creators responsibly integrate AI into the creative process without sacrificing originality, diversity, or human authorship?

Case References

del Barco, Mandalit. "Striking Movie and TV Writers Worry That They Will Be Replaced by AI." *NPR*, May 20, 2023. **https://www.npr.org/2023/05/20/1177366800/striking-movie-and-tv-writers-worry-that-they-will-be-replaced-by-ai**

Kinder, Molly. "Hollywood Writers Went on Strike to Protect Their Livelihoods from Generative AI. Their Remarkable Victory Matters for All Workers." Brookings Institution, April 12, 2024. **https://www.brookings.edu/articles/hollywood-writers-went-on-strike-to-protect-their-livelihoods-from-generative-ai-their-remarkable-victory-matters-for-all-workers/**

Writers Guild of America. *Memorandum of Agreement for the 2023 WGA Theatrical and Television Basic Agreement*. Los Angeles: Writers Guild of America, September 2023. **https://www.wga.org/uploadedfiles/contracts/2023_mba_moa.pdf**

CASE STUDY 8.2: GENERATIVE AI ART IN BOARD GAMES



Artwork is a cornerstone of modern board games, shaping how players experience and perceive a game. Publishers know that eye-catching cover art gets customers to pull the game off the shelf and examine the back for a description. High-quality illustrations enhance immersion, communicate mechanics, and serve as critical marketing tools. In a highly competitive industry, artwork often determines whether a game succeeds or fails. However, the rise of AI-generated art has created opportunities and ethical dilemmas for designers, artists, publishers, and players.

Jenna's Decision: *Legends Unbound*

Jenna, a fictional, budding game designer, faced a crucial decision while preparing to launch *Legends Unbound*, a card game about mythical creatures. With limited funds and a crowded market, she debated whether to commission freelance artists or use AI tools for her prototypes, promotional materials, and production:

- ***Prototyping.*** Early playtests struggled due to inconsistent clipart placeholders. Using Stable Diffusion, Jenna generated cohesive prototype art in hours at a minimal cost, improving clarity for players without a significant investment.
- ***Promotion.*** Preparing her Kickstarter campaign, Jenna created striking visuals using MidJourney instead of hiring a human artist. She disclosed the use of AI art, and although some hobbyists raised concerns about authenticity, enough backers supported the project to fund production successfully.
- ***Production.*** Jenna chose a hybrid approach for the final components—hiring an artist for key hero illustrations while using AI-generated textures for backgrounds. This balance reduced costs while acknowledging the value of human creativity.

Jenna's choices reflect the tensions among accessibility, efficiency, and artistic integrity that AI art brings to creative industries.

Business Opportunities

AI-generated art offers substantial advantages, particularly for small or independent creators. It reduces costs, saves time, and lowers barriers to market entry. Platforms like Kickstarter have

amplified these benefits by making visual appeal essential to early funding success.

Publishers such as Wise Wizard Games and Stronghold Games have embraced AI art, arguing that it enables quicker development and broader innovation. For new designers working on a shoestring budget, AI art democratizes creativity, making it possible to produce visually competitive games.

However, reliance on AI art is not without risks. AI-generated works lack copyright protection, making games vulnerable to counterfeiting and bootlegging—serious financial and reputational threats. Furthermore, brands built on authentic craftsmanship, like Stonemaier Games (*Wingspan*), explicitly reject AI art to protect their identity and credibility with discerning consumers.

Ethical Challenges

The adoption of AI-generated art also raises profound ethical concerns:

- ***Devaluation of artistic labor.*** Human artists rely on commissioned work for their livelihoods. Replacing or undercutting them with AI-generated images—often trained on copyrighted materials without consent—threatens the creative ecosystem.
- ***Transparency.*** Failing to disclose AI usage can mislead consumers who assume the game reflects the vision and effort of human artists. Hobbyist communities, which value authenticity and craft, may react negatively if AI use is concealed.
- ***Loss of creative depth.*** Artists contribute images, emotion, nuance, and imagination that AI cannot replicate. Over-reliance on AI risks producing visually appealing but emotionally hollow games.
- ***Intellectual property and legal risks.*** AI-generated art lacks clear legal ownership protections, complicating enforcement against knockoffs and infringements, and endangering the sustainability of original game publishing.

Conclusion

Jenna's journey with *Legends Unbound* highlights the complex tradeoffs between innovation and tradition in modern board game design. AI-generated art offers cost and time savings but brings ethical and reputational risks. As she prepared for release, Jenna worried that even with transparency, AI visuals might alienate players who see board games not just as entertainment but as social rituals—where placing a beautiful game on the table is like serving a carefully crafted meal. A game that feels soulless can cast a shadow over the gathering. Hybrid approaches and respect for creative authenticity may offer a path forward, but designers must weigh how AI shapes their product, the experience it creates, and impacts sales.

As AI tools become more powerful, creators, publishers, and players must ask: What is the real

cost of cutting corners on creativity in a hobby built on human connection?

Discussion Questions

1. Is Jenna's hybrid approach to AI-generated and human-created art a responsible compromise, or does it still pose ethical concerns? Consider both the financial and community-based implications of her decision.
2. How might the use of AI-generated art affect the long-term brand reputation of a game or publisher, especially within hobbyist-driven markets? Would full disclosure be enough to maintain trust?
3. Some artists and publishers have argued that AI-generated art, trained on copyrighted work without consent, constitutes theft. Do you agree with this view? How should creators balance innovation with respect for intellectual property?
4. From a business perspective, when—if ever—is it justifiable to replace human artists with AI in game development? What factors should influence that decision?
5. How can game designers balance innovation, affordability, and respect for artistic integrity in an increasingly AI-driven creative economy? What role should transparency and community norms play in shaping that balance?

Case References

Bosley, Andrew. "A Very Specific Case against AI Art: Board Game Prototypes and AI." Wits End Blog, October 27, 2023. https://www.bosleyart.com/post/a-very-specific-case-against-ai-art-board-game-prototypes-and-ai?utm_source=chatgpt.com

Didymus, Mike. "'It's Theft': Root Artist Kyle Ferrin on the 'Slop' of AI Art, and Why Publishers Using It Are 'Devaluing All of Board Games as an Industry'." BoardGameWire, November 7, 2024. https://boardgamewire.com/index.php/2024/11/07/its-theft-root-artist-kyle-ferrin-on-the-slop-of-ai-art-and-why-publishers-using-it-are-devaluing-all-of-board-games-as-an-industry/

Didymus, Mike. "Star Realms Maker Wise Wizard Games Defends Using AI Art in New Projects, Board Game Artists Call Out Ethics of Decision." BoardGameWire, November 7, 2024. https://boardgamewire.com/index.php/2024/11/07/star-realms-maker-wise-wizard-games-defends-using-ai-art-in-new-projects-board-game-artists-call-out-ethics-of-decision/

CASE STUDY 8.3: HALLUCINATING THE LAW

In May 2023, lawyers representing a plaintiff in a personal injury case in the Southern District of New York faced public embarrassment. Their legal brief included citations to several court decisions—none of which existed. The citations were entirely fabricated by ChatGPT, a generative AI tool used to assist in drafting the filing.

The fallout was swift. Opposing counsel and the judge couldn't find the cases in any legal database. When questioned, the lawyer admitted he had relied on ChatGPT, mistakenly thinking it functioned like a legal search engine. The incident made national headlines and raised serious concerns about using AI in professional settings where factual accuracy is essential.

What Went Wrong?

At the issue's core was a misunderstanding of how generative AI works. ChatGPT, like other large language models, doesn't retrieve documents from the Internet or legal databases. Instead, it generates text by predicting patterns based on its training data. It produced legally plausible but entirely fictional citations, complete with judges' names, docket numbers, and reasoning when asked for supporting cases.

The prompt reportedly used was:

> "*Please provide recent federal court cases that support tolling of the statute of limitations under the Montreal Convention.*"

Although the prompt was specific, it lacked constraints like "only cite verifiable sources" or "link to official databases." The AI interpreted it as a request for plausible text and responded accordingly.

Could a Persona Have Helped?

Some might ask whether using a defined persona, such as a seasoned legal researcher, could have prevented the hallucination. A persona can guide AI to interpret prompts with added tone, context, and domain expectations. For instance, a prompt like:

> "*You are a U.S. legal research assistant. Only cite cases that can be verified in Westlaw or LexisNexis.*"

might nudge the AI to admit uncertainty or refrain from inventing cases. Still, personas improve tone and style, but they don't override a model's limitations. Open-domain chatbots like ChatGPT aren't connected to live legal databases and can't reliably provide verified citations, even in expert roles. Personas help guide AI output but are no substitute for human verification.

Open vs. Closed Chatbots: Choosing the Right Tool

Another critical factor in this case is the type of chatbot used. ChatGPT is an open-domain model designed to respond across many topics but, as noted, it lacks access to legal databases like Westlaw or LexisNexis. A closed-domain legal AI, trained on vetted legal sources and built to avoid responses outside its scope, would likely have flagged the request or declined to answer.

Tools like Lexis+ AI, Westlaw Precision, and Harvey AI generate fact-based, citable legal content designed with safeguards to prevent hallucinations. Choosing the correct tool can be as crucial in legal work as crafting the right prompt.

Prompt Engineering Lessons

This incident offers several key lessons:

- ***Verification is critical.*** Even well-formed prompts must be followed by fact-checking, especially in law, health care, or finance. Asking for citations isn't enough. Users must confirm them.
- ***Iterative refinement is key.*** Reviewing and refining the prompt with constraints like "cite real cases only" or "exclude fabricated results" could have avoided the issue.
- ***Understand AI models.*** It's essential to know that open-domain models can hallucinate. Unlike closed-domain research tools, they aren't designed to deliver guaranteed factual accuracy.

The Legal and Ethical Fallout

The court sanctioned the lawyers and their firm, not for using AI, but for failing to verify its output before submitting it to a federal judge. Lawyers are responsible for the accuracy of what they file, regardless of the tools they use.

The case drew national attention, with headlines like "ChatGPT Made Up Legal Cases—And Lawyers Cited Them" appearing in major outlets. Judge Kevin Castel called it "unprecedented" and "a cautionary tale."

It also reignited concern about AI's role in legal work. While AI can assist with drafting and organizing, its use requires scrutiny, especially when outputs are not verifiable or traceable to real sources.

The Bigger Picture

This case reminds us that AI is a tool, not a substitute for professional judgment. Tools like ChatGPT are built to generate text, not deliver factual, legal citations. The AI may respond with fluent but false information when prompts are vague or unconstrained. Used carelessly, these

outputs can damage credibility and cause harm.

As AI continues integrating into law, business, and education, professionals must learn how to use it responsibly. Crafting clear, constrained prompts and verifying the results is not just a technical skill but a professional imperative.

Discussion Questions

1. What prompt modifications could have prevented ChatGPT from generating fabricated case citations?
2. Should generative AI tools be used in high-stakes fields like law? If so, what safeguards are necessary?
3. Who bears responsibility when AI generates inaccurate or misleading content—users, developers, or both?
4. How can professionals balance efficiency gains from AI with the need for accuracy and verification?
5. What implications does this case have for teaching students or employees how to use generative AI tools responsibly?

Case References

Banino, B. J., Lefland, M. N., and Carousis, D. L. “In Landmark Case Involving Attorneys Who Submitted Fake Court Decisions Generated by ChatGPT, Federal Judge Sanctions the Attorneys and Dismisses Action as Time-Barred under the Montreal Convention.” Condon & Forsyth LLP, June 30, 2023. https://condonlaw.com/2023/06/in-landmark-case-involving-attorneys-who-submitted-fake-court-decisions-generated-by-chatgpt-federal-judge-sanctions-the-attorneys-and-dismisses-action-as-time-barred-under-the-montreal-convention/

Neumeister, Larry. “Lawyers Submitted Bogus Case Law Created by ChatGPT. A Judge Fined Them $5,000.” Associated Press, June 22, 2023. https://apnews.com/article/artificial-intelligence-chatgpt-fake-case-lawyers-d6ae9fa79d0542db9e1455397aef381c

ENDNOTES

[1] Schmidt, Douglas C., et al. *Evaluating Personified Expert Effectiveness in AI-Powered Systems*. Vanderbilt University, 2023. **https://www.dre.vanderbilt.edu/~schmidt/PDF/Evaluating_Personified_Expert_Effectiveness_Conference.pdf**

[2] Incubity. "10 Examples of Persona Patterns in Prompt Engineering." *Incubity*, 2023. https://incubity.ambilio.com/10-examples-of-persona-patterns-in-prompt-engineering/; Skim AI. "Personas in Prompt Engineering." 2023. https://skimai.com/personas-in-prompt-engineering/; Wan, Y., Zhao, J., Chadha, A., Peng, N., & Chang, K.-W. "Are Personalized Stochastic Parrots More Dangerous? Evaluating Persona Biases in Dialogue Systems." *arXiv*. Last modified November 2, 2023. **https://arxiv.org/abs/2310.05280**

[3] Neff, Gina, & Nagy, P. (2016). "Talking to Bots: Symbiotic Agency and the Case of Tay." *International Journal of Communication*, 10, 4915–4931. **https://ijoc.org/index.php/ijoc/article/view/6277/1804**

INTRODUCTION

Where Smart Infrastructure Meets Smart Business

Singapore has long been recognized as a global innovation, governance, and urban planning leader. In 2014, the city-state launched its Smart Nation initiative, an effort to harness AI, data analytics, IoT, and automation to improve the quality of life for citizens, support sustainable growth, and position Singapore as a model for smart cities worldwide. This ongoing case series follows Ari, a Singapore resident and young professional living at the heart of this transformation. Through Ari's day-to-day experiences, we explore how Singapore's Smart Nation vision is being realized—how technology integrates seamlessly into infrastructure, services, and personal life, and how it raises new questions about trust, inclusion, and innovation.

A SEAMLESS MORNING IN SINGAPORE'S SMART NATION

Ari begins their day in a high-rise dwelling in the heart of a bustling smart city, Singapore, exemplifying the seamless blend of technology and sustainability that defines modern urban living. As they wake, the automated systems in their home are already at work, making the morning smooth and efficient.

Smart Home Start

The smart glass windows gradually become transparent, filling the room with soft natural light. Thanks to intelligent climate control tuned to real-time weather and Ari's preferences, the indoor temperature has already been adjusted. A glance at a nearby device provides a full-day overview, including personalized traffic updates, the fastest transit options, and even reminders tied to Ari's wellness goals.

Every detail, from the water-saving shower to the energy-efficient kitchen appliances, reflects the city's commitment to sustainability through initiatives like Singapore's Building and Construction Authority2 (BCA) programs. The building itself operates with minimal human oversight, with waste sorted by a smart chute system and green walls nurtured by automated watering and air quality sensors.

A brief message flashes on Ari's wall display:

"Your data is protected by Smart Nation standards. Learn more at SmartNation.gov.sg." It's a small reminder of the trust built into the system—and the growing importance of digital governance in everyday life.

Stepping into the Community

As Ari leaves their apartment, they drop recyclables into a designated smart bin in the corridor, part of a pilot waste management program designed to track and optimize their collection. In the lobby, a living green wall stretches up two stories, nourished by an automated irrigation system—a common feature in many new buildings under Singapore's Skyrise Greenery Initiative.[1] These natural oases improve air quality and contribute to the building's sustainability score under the BCA Green Mark program.

Public Infrastructure in Sync

Outside, Ari walks along a tree-lined sidewalk where smart lighting adapts to motion and ambient light—one of the energy-saving measures being deployed in Singapore's urban design.[2] At a bus stop nearby, a digital sign[3] displays live updates on bus arrivals, traffic delays, and the air quality index. These real-time insights, powered by interconnected sensors and AI, are already part of daily life in several districts.

Commute with Intelligence

As Ari approaches the transit hub, integrated with parks, cafés, and workspaces, the city app suggests the most efficient and eco-friendly route based on real-time conditions. Public transportation in Singapore has evolved into an AI-optimized network of electric buses and autonomous trains,[4] managed by centralized systems ensuring low wait times and reliable service.

Inside the smart train, digital displays offer updates on delays, transfers, and points of interest. The carriage is full of commuters who are absorbed in their routines but not overcrowded. The environment remains clean, and comfortable due to strict environmental standards and technology-driven operations that keep the system running efficiently.

Smart Commerce and Customer Experience

Before heading into work, Ari stops by a local café. They order their usual drink and pastry using a digital kiosk, paying with Singapore's unified digital wallet app.[5] A notice pops up when they press submit, "*Your data is encrypted and used only for service delivery.*" The contactless payment system uses short-range wireless technology—similar to tapping a phone to pay—making transactions quick, secure, and convenient. The app logs loyalty points and sends an instant receipt.

The café is an AI-driven operation,[6] with IoT sensors managing inventory and supply chains to prevent shortages. This reduces waste and ensures the availability of Ari's favorite breakfast—a perfect example of how innovative business meets customer satisfaction.

Workplace and Reflection

Ari's workplace, a smart building with high sustainability certifications, is connected to the city's more extensive green corridor network. The final stretch of the commute is a brief walk through nature, blending tech and wellness.[7]

As they arrive, Ari reflects on how seamlessly their lives integrate with the infrastructure around them. Their experience isn't just about convenience—it's about how AI and smart systems quietly enhance the quality of life while businesses and governments work harmoniously to create value.

Not every part of Ari's morning is fully realized today—but each element is grounded in real

initiatives, prototypes, or policy goals. Together, they reflect the vision Singapore is actively building toward.

Discussion Questions

1. Which elements of Ari's morning seem futuristic compared to today? Why?
2. Could Ari's experience be possible without AI? Why or why not?
3. Why doesn't Ari just work from home, given the advanced technology at home? What business or societal value does commuting to a workplace still offer?
4. What types of businesses are likely responsible for the technology powering Ari's morning? Who is creating value—and how are they capturing it?
5. What risks or concerns (privacy, overdependence, digital divide) might arise in a smart city like this? How might businesses or governments address them?

Case References

Smart Nation Singapore. (2023). Smart Nation: The Way Forward. Retrieved from **https://www.smartnation.gov.sg/**

BCA (Building and Construction Authority). 2022. Green Building Masterplan. **https://www1.bca.gov.sg/**

National Parks Board. 2022. Skyrise Greenery Incentive Scheme. Accessed April 2, 2025. **https://skyrisegreenery.nparks.gov.sg/incentive-scheme/about/**

ST Engineering. 2023. Smart Estate Lighting. Accessed April 2, 2025. **https://www.stengg.com/en/smart-city/urban-solutions/smart-utilities-and-infrastructure/agil-smart-lighting/agil-smart-estate-lighting/**

Land Transport Authority (LTA). 2022. Passenger Information Display Systems. Accessed April 2, 2025. **https://www.lta.gov.sg/content/ltagov/en/who_we_are/statistics_and_publications/Connect/pids.html**

Land Transport Authority. 2022. Smart Mobility 2030. **https://www.lta.gov.sg/**

Monetary Authority of Singapore. 2021. E-payments in Singapore. **https://www.mas.gov.sg/**

Infocomm Media Development Authority. (2022). AI and IoT in Singapore's Retail Sector. **https://www.imda.gov.sg/business**

FOUNDATIONS OF A SMART NATION

A Nation at a Crossroads: Laying the Groundwork for a Smarter Future

In 2014, Singapore stood at a pivotal moment in its development. Though already recognized globally for its efficient governance, economic strength, and high standard of living, the city-state faced mounting challenges that threatened to hinder its progress. Limited land, an aging population, rising urban density, and the need for continued economic growth required a bold strategic shift. In response, Singapore launched its Smart Nation initiative—a comprehensive national strategy that leverages digital transformation, AI, and data as key tools for addressing complex societal and business challenges.

A Glimpse of the Future

In 2014, as Singapore's Smart Nation vision was being announced, a 14-year-old Ari stood in the back of a school assembly hall watching a video on the future of AI in Singapore. It showed driverless buses gliding through clean city streets, apartments that adjusted lighting and temperature automatically, and doctors able to check on patients from afar using digital health monitors built into their watches.

Ari wasn't sure which was more remarkable—the buildings covered in plants, the apps that could find you the perfect job, or the idea that all your government services might someday live on a single screen. They went home that evening and searched for "how to build an app." Ari signed up for a free coding course at the community library that summer. Within a few years, they were helping classmates build simple websites and experimenting with public datasets.

Almost a decade later, Ari joined a Smart Nation project team to work on achieving that vision.

The Three Pillars of the Smart Nation Vision

The Smart Nation initiative focused on three interconnected goals:

1. ***Improve the quality of life*** for citizens by making public services more responsive, accessible, and efficient.
2. ***Enable economic growth and productivity*** by embracing innovation, digital infrastructure, and automation.
3. ***Establish Singapore as a leading global city*** through the strategic use of technology and forward-thinking urban governance.

While each pillar had distinct priorities, the underlying belief was that technology—particularly AI and data-driven solutions—would enable Singapore to address its structural limitations and open

new avenues for sustainable growth.

Challenges and Opportunities in 2014

The Smart Nation initiative was designed to tackle six major areas of focus:

- ***Economic landscape.*** Singapore was a global financial hub with a thriving economy. It had a strong service sector and robust manufacturing base and was a key player in international trade. Singapore was well positioned to pivot toward AI–driven economic transformation. It had a positive trade balance and was ranked number one globally for ease of doing business.
- ***Urban development.*** Despite its limited land area, Singapore was recognized for its well-planned urban infrastructure. The city was clean, green, and highly urbanized, with efficient public transportation and housing systems. Singapore needed to manage growth sustainably.
- ***Demographics*****.** The population was densely concentrated, leading to challenges such as traffic congestion. The demographic was also aging rapidly, raising concerns about health care and social support.
- ***Technological adoption.*** Singapore was technologically advanced compared to many of its regional counterparts, with high Internet penetration and a growing tech-savvy population. However, the full potential of technology in public services and city management was yet to be realized.
- ***Environmental sustainability.*** Sustainability was a key concern due to limited natural resources. Reliant on imports for essentials like water, Singapore sought to use technology to manage resources and reduce environmental impact.
- ***Global positioning.*** Already a global business hub, Singapore aimed to become a world leader in AI innovation and innovative governance.

In 2014, Singapore was a vibrant, economically strong, and efficiently managed city-state, yet it faced the imperative to innovate further in addressing urban, demographic, and environmental challenges. The Smart Nation initiative was viewed as a strategic move to leverage technology for sustainable growth and an enhanced quality of life for its citizens. The belief that AI could drive operational efficiency and create a strategic advantage positioned Singapore as a global model for how cities could use emerging technologies to create lasting value.

Singapore’s Strategic Shift toward AI

These early decisions made Singapore a proving ground for urban innovation and a model for how AI can reshape modern governance. In the following sections, we will see how this vision began to materialize—through policies, partnerships, and people like Ari helping shape the Smart Nation.

Discussion Questions

1. Why was “quality of life” prioritized over economic growth in the Smart Nation initiative? How are the two connected?
2. What business, infrastructure, or social obstacles might Singapore have faced in launching a Smart Nation strategy?
3. How might AI specifically help address the six challenges outlined in the case (economic, urban, demographic, etc.)?
4. How did Singapore’s status in 2014 (tech-savvy population, efficient government) create an advantage for smart city transformation?
5. Imagine you’re advising a city in the United States with similar challenges. What lessons from Singapore’s early strategy could be applied to your city?

Case References

Smart Nation Singapore. 2014. Launch of the Smart Nation Initiative. Accessed April 2, 2025. **https://www.smartnation.gov.sg/whats-new/press-releases/launch-of-smart-nation**

Smart Nation Singapore. 2018. About Smart Nation. Accessed April 2, 2025. **https://www.smartnation.gov.sg/about-smart-nation**

Smart Nation and Digital Government Office. 2019. National Artificial Intelligence Strategy. Accessed April 2, 2025. **https://www.smartnation.gov.sg/files/publications/national-ai-strategy.pdf**

Urban Redevelopment Authority. 2021. Singapore Green Plan 2030. Accessed April 2, 2025. **https://www.greenplan.gov.sg/**

Ministry of Communications and Information. 2019. Speech by Minister at the AI Singapore Symposium. Accessed April 2, 2025. **https://www.mci.gov.sg/media-centre/speeches/speech-by-minister-vivian-balakrishnan-at-ai-symposium**

SINGAPORE'S DATA ENGINE

Fueling a Smart Nation

Ari leaned back from their desk and looked out the 20th-floor window of a government co-working space in Singapore's downtown core. After months of working on data access and digital services projects, Ari had begun to see the bigger picture: data was the foundation—the high-quality, well-managed input that enabled AI algorithms to perform meaningful work.

They hadn't thought about that school assembly in years—when they first watched Singapore's Smart Nation announcement in 2014. But now, seated in a glass-walled workspace in the heart of downtown, they were part of it.

Their role gradually grew, beginning with a temporary assignment reviewing data-sharing policies across city departments. But as Ari contributed to projects tied to mobility, health care, and urban planning, they noticed a common thread: AI couldn't function without the right data, delivered in the right way, and governed responsibly. That idea became the foundation of Ari's work in Singapore's Smart Nation initiative.

The Foundation of Data Infrastructure

Since 2014, Singapore has invested heavily in secure, centralized digital systems. From the ubiquitous SingPass digital ID to the National Digital Identity platform, residents can access services across health care, housing, transport, and more with a few taps. Behind the scenes, these services rely on interconnected datasets and APIs, many of which are managed through GovTech and the Smart Nation and Digital Government Office.

To support the vast and varied data needs of Singapore's Smart Nation initiative, GovTech and other agencies organize information using centralized data lakes. A data lake is a large-scale repository that stores structured and unstructured data in its raw form. Unlike traditional databases, which require predefined schemas, data lakes allow for flexible storage, enabling analysts, developers, and AI systems to extract and transform data as needed. This architecture provides the scalability and accessibility necessary to power services across health, mobility, housing, and more.

Singapore also maintains an open data portal, **Data.gov.sg**, which makes hundreds of public datasets freely accessible. This transparency enables businesses, researchers, and citizens to innovate—from mapping traffic patterns to developing AI applications that utilize weather, transportation, or health data.

Governance and Trust

But data access alone isn't enough. Ari's team spends a significant amount of time on data governance, setting standards for accuracy, privacy, interoperability, and ethical use. Frameworks such as the Personal Data Protection Act (PDPA) and Singapore's AI Governance Framework establish clear boundaries for the use of data, specifying who can access it and for what purposes.

Public trust is essential. Digital services only work if citizens believe their data is being handled responsibly. To this end, Singapore has launched education campaigns, ethics panels, and public consultations to explain how smart city data is collected and utilized—especially in high-sensitivity areas such as facial recognition, contact tracing, and predictive policing.

Despite these efforts, challenges have nonetheless emerged. One high-profile case involved the TraceTogether contact tracing app, introduced during the COVID-19 pandemic. Initially promoted as a public health tool, it was later revealed that law enforcement could access the data for criminal investigations. This sparked public concern over government overreach and led to amendments in the law, limiting such access to investigations involving serious crimes. The incident emphasized the importance of transparency and strong legal safeguards.

Beyond TraceTogether, broader surveillance technologies—like CCTVs with facial recognition—have raised ongoing concerns about privacy and mass surveillance. These moments have underscored the need for innovation to be accompanied by public dialogue and clear ethical frameworks. For Singapore, earning and maintaining public trust remains as critical as the technology itself. Digital services only work if citizens believe their data is being handled responsibly. To this end, Singapore has launched education campaigns, ethics panels, and public consultations to explain how smart city data is collected and utilized—especially in high-sensitivity areas such as facial recognition, contact tracing, and predictive policing.

From Data to Impact

Singapore has utilized environmental sensors and AI to predict the risks of dengue outbreaks. Studies have shown that machine learning models incorporating meteorological data, such as temperature, humidity, and rainfall, have been employed to forecast dengue cases in Singapore. These models analyze data to identify patterns and potential hotspots for dengue transmission, enabling public health agencies to take proactive measures.

Every successful AI application Ari encountered had a common starting point: reliable, well-governed data. Without it, even the most advanced algorithms failed to deliver real-world results.

Looking Ahead

Singapore's Smart Nation is about making evidence-based decisions, streamlining services, and creating value for businesses and residents. Ari now sees their work not as building technology, but as capturing and organizing the data that makes meaningful AI innovation possible.

For Ari, data wasn't just a technical asset—it was the quiet force behind every smart building, train

route, and health alert. In the next phase of Singapore's journey, data would power decisions in ways the 14-year-old version of them could never have imagined.

Discussion Questions

1. Why is data quality so important in developing AI systems? How does Singapore ensure high-quality data?
2. What are the key tradeoffs between data openness and privacy in a smart city?
3. How do regulations such as the PDPA and the AI Governance Framework contribute to the development of responsible AI?
4. What were the public concerns related to the TraceTogether app, and what does this tell us about the importance of transparency and trust in smart city initiatives?
5. What lessons could other countries or cities learn from Singapore's approach to data infrastructure and trust-building?

Case References

Government of Singapore. "Singpass." Last updated March 20, 2025. **https://www.singpass.gov.sg/**

Government Technology Agency of Singapore (GovTech). "National Digital Identity." Last updated April 1, 2025. **https://www.tech.gov.sg/our-digital-government-efforts/digital-identity/**

Government Technology Agency of Singapore (GovTech). "About Us." Last updated April 1, 2025. **https://www.tech.gov.sg/**

Attorney-General's Chambers of Singapore. "Personal Data Protection Act 2012." Last updated April 1, 2025. **https://sso.agc.gov.sg/Act/PDPA2012**

Personal Data Protection Commission (PDPC). "Singapore's Approach to AI Governance." Last updated November 3, 2023. **https://www.pdpc.gov.sg/help-and-resources/2020/01/model-ai-governance-framework**

Han, Kirsten. "Broken Promises: How Singapore Lost Trust on Contact Tracing Privacy." MIT Technology Review, January 11, 2021. **https://www.technologyreview.com/2021/01/11/1016004/singapore-tracetogether-contact-tracing-police/**

Tian, Na, Jin-Xin Zheng, Lan-Hua Li, Jing-Bo Xue, Shang Xia, Shan Lv, and Xiao-Nong Zhou. "Precision Prediction for Dengue Fever in Singapore: A Machine Learning Approach Incorporating Meteorological Data." Tropical Medicine and Infectious Disease 9, no. 4 (2024): 72. **https://doi.org/10.3390/tropicalmed9040072**.

SINGAPORE'S INTELLIGENT SPACES

How Smart Buildings Use Machine Learning to Think Sustainably

Ari at the Bosch Campus

Ari stopped at reception and received a visitor badge at the Bosch Singapore Campus. Bosch had transformed its decade-old facility into a smart building using IoT sensors and cameras integrated with Bosch Connected Building software. Data from the various sensors and legacy systems were aggregated to form a data lake and analyzed and presented in a graphical dashboard. Bosch was using the building to test and demonstrate the value of its software in creating Smart Buildings. Ari had come to learn more about the project, ahead of supporting future building upgrades across the city. Their role would soon involve helping develop machine learning models to forecast electrical usage and support the management of Singapore's energy grid.

Singapore's Green Plan and the Role of Machine Learning

Singapore's Green Plan 2030,[8] launched in February 2021, aims to transform the nation into a leader in sustainability through smart design and technology. Key targets with building infrastructure include:

- Greening 80 percent of Singapore's buildings (by gross floor area) by 2030—meaning they will be certified under the *BCA Green Mark Scheme*,[9] which promotes sustainable design and energy-efficient operations.
- Ensuring that 80 percent of new buildings are *Super Low Energy*—a designation for buildings that achieve significantly higher energy efficiency through smart design and technology—starting from 2030.
- Achieving an 80 percent improvement in energy efficiency for best-in-class green buildings, compared to the 2005 level.

The Building and Construction Authority (BCA), in collaboration with GovTech and the Energy Market Authority, initiated a pilot using machine learning to forecast short- and long-term energy consumption in smart buildings. These intelligent systems optimize HVAC scheduling, lighting automation, and future retrofitting decisions using real-time sensor data.

The Sensor-Driven Transformation

At the Bosch campus, Ari observed the implementation of Bosch's Connected Building platform, which integrated a diverse array of IoT sensors. These devices continuously monitored environmental and occupancy conditions, feeding a centralized system designed to respond dynamically. Table SC.1 shows the types of sensors used and their purpose.

Table SC.1 **Smart Building Sensors**

Sensor Type	Parameters Monitored	Use or Purpose in Smart Building
Air quality sensors	Temperature, humidity, CO_2, and total volatile organic compounds (TVOCs)	Optimize ventilation and HVAC systems; maintain healthy indoor air quality
Ambient light sensors	Levels of natural and artificial light	Automatically adjust lighting systems for energy savings and occupant comfort
Ambient noise sensors	Sound levels across zones	Monitor noise pollution, adjust zoning for productivity and acoustic comfort
Occupancy sensors	Presence detection via infrared (IR) sensors	Control lighting, heating/cooling, and security based on room usage
Bluetooth beacons	Real-time location tracking of people or assets	Enable room-level navigation, asset tracking, and space utilization analytics
Air pressure sensors	Barometric pressure changes	Detect door/window status or help in leak detection in pressurized systems
Vibration sensors	Structural or mechanical vibrations	Predictive maintenance; detect anomalies in HVAC, elevators, or structural stress

This sensor ecosystem provided insight into building conditions and data for predictive energy modeling.

Training the Model

The project's machine-learning component was based on supervised learning. Historical energy consumption, weather data, occupancy logs, and time-of-day inputs formed the training set. Ari collaborated with the data science team to prepare this data, labeling outliers such as holidays and system maintenance events and refining their data lake.

The team tested various approaches, including linear regression and several time-series models,

before ultimately settling on an ensemble model that merged the strengths of each. Ari's attention to feature selection proved vital—occupancy data, temperature, and humidity—emerged as key predictors. Surprisingly, traffic around the cafeteria also became an important signal for ventilation control.

This hands-on experience underscored the reality that building effective machine learning systems requires both technical rigor and continuous refinement.

Scaling toward a Smart Nation

Understanding the factors that affected the energy use in this building would allow them to operate this one more efficiently, design better buildings, and pool data from many such building to forecast energy consumption for the city. Ari's team submitted their model to BCA's benchmarking framework and collaborated with the Energy Market Authority to feed into regional grid planning. Predictive demand modeling could help prepare for future disruptions, such as unexpected weather, maintenance downtime, or spikes in usage during national events.

Discussion Questions

1. What are the primary benefits of integrating IoT sensors into buildings, and how do these sensors enable smarter energy management?
2. What challenges might arise in training machine learning models for building energy forecasting?
3. In the case, traffic near the cafeteria became an important predictor for ventilation needs. What does this suggest about working with real-world data?
4. How does this smart building case support Singapore's broader Green Plan 2030 goals?
5. If you were designing a smart campus in your city, what sensor or capability would you prioritize—and why?

SINGAPORE'S INTELLIGENT TRANSPORTATION SYSTEM

The Future of Smart Roads

A Road That Thinks Ahead

It wasn't the vehicles that were smart—at least, not yet. Ari stood on a pedestrian overpass watching traffic flow through one of Singapore's new Smart Road corridors, part of its broader Intelligent Transport Systems (ITS) initiative. Overhead cameras and roadside sensors gathered live traffic data, weather updates, and pedestrian movement. Beneath the pavement, pressure sensors and inductive loops tracked vehicle patterns. It wasn't as glamorous as a self-driving car demo, but this system works now, which made all the difference. The Land Transport Authority (LTA) wasn't trying to replace drivers. It focused on building a road system that could learn, respond, and eventually guide autonomous vehicles when right.

Neural Networks in the Infrastructure

Ari was part of a GovTech analytics team tasked with optimizing traffic flow using real-time sensor data. The goal was simple: reduce congestion, prioritize buses and ambulances, and improve pedestrian safety. The method? Neural networks, trained on thousands of hours of road data. The system used convolutional neural networks (CNNs) to classify and track vehicles and pedestrians from camera feeds and sensor input. The team also explored recurrent neural networks (RNNs) to predict short-term traffic flows based on time of day, weather, and event data. All data is streamed into **Analytics.gov**, Singapore's secure AI/ML cloud platform used by government agencies for real-time analytics. These systems enabled infrastructure to detect accidents or slowdowns, predict congestion before it happened, adjust traffic light timing automatically, and identify risky pedestrian crossings or poorly lit areas.

Making a Case for the Human Impact

At first, public discussion focused on autonomous vehicles. But Ari and their team reframed the message. "This isn't about robots driving cars," Ari told a town hall audience in Bukit Panjang. "It's about reducing your wait at a crosswalk or clearing the road faster when an accident blocks a bus lane. It's about smarter public spaces, not just smarter vehicles." That shift worked. Residents responded more positively when they saw how the data would help them today, not just power some future technology.

A System That Sees Too Much?

As the system improved, concerns emerged. “If these cameras track everything,” one resident asked, “who’s watching us?” Ari explained that Singapore used edge AI to process footage directly on site, keeping only critical anomalies. Most footage was automatically anonymized. Still, the concern lingered. So, the team worked with agencies to revise system thresholds and create community dashboards, letting residents see what analytics were applied to their intersections and how they changed traffic flow. Transparency and trust became as important as speed and safety.

A System That Finds Parking Too

Another practical benefit came from integrating camera and sensor data into parking systems. In high-traffic areas like Toa Payoh and Orchard, using real-time data, the Smart Roads platform was extended to guide drivers to available parking lots. Digital signs and mobile apps displayed availability, reducing circling and easing congestion. “It saved me 15 minutes every morning,” one commuter shared in an LTA feedback session. For Ari, these small efficiencies showed how smart infrastructure could make daily life smoother—not just safer.

A System That Clears the Way

The clearest sign of success wasn’t what the Smart Roads saw—but what they enabled. LTA launched a Traffic Priority System (TPS) pilot near Ng Teng Fong General Hospital. Ambulances were equipped with onboard devices that communicated with traffic lights ahead. In life-threatening cases—cardiac arrest, trauma, stroke—paramedics could activate the TPS to trigger a green-light corridor to the hospital. Over 3 months, TPS was activated 76 times, cutting emergency response times by an average of 100 seconds. That margin could mean the difference between life and death. “It’s not flashy,” Ari said, “but that’s the point. Smart Roads aren’t about headlines. They’re about clearing the road when every second counts.”

The Future in the Foundations

Smart Roads weren’t the end goal. They were the first step. As Ari looked down from the overpass, they imagined what might come next: AVs navigating seamlessly, coordinated with traffic lights anticipating their route; intersections learning from near misses and adjusting their timing for safety; delivery bots crossing with embedded pedestrian signals. But one thought always came back: “AI only matters if it improves life.” And in Singapore, that meant building infrastructure that learns long before handing over the wheel.

Discussion Questions

1. How are neural networks used in Singapore's Smart Roads initiative, and what are the practical benefits they deliver today?
2. Why is it important for Singapore to invest in smart infrastructure before the widespread adoption of autonomous vehicles?
3. What strategies did Ari's team use to build public trust in the Smart Roads initiative? Were they sufficient?
4. The case describes an ethical concern raised by a citizen. What are the privacy risks of smart infrastructure, and how can governments mitigate them?
5. What other types of smart infrastructure—beyond roads—could benefit from this approach to AI and data use?

Case References

Alea IT Solutions. (2024). "Singapore Introduces AI-Powered Traffic Management System: A Step towards Smart City Success." *Alea IT Solutions*, October 10. **https://www.aleaitsolutions.com/singapore-introduces-ai-powered-traffic-management-system-a-step-towards-smart-city-success/**

Channel News Asia. (2024). "New Ambulance Traffic Priority System for Life-Threatening Cases Activated 76 Times in Three Months." *CNA*, November 23. **https://www.channelnewsasia.com/singapore/ambulance-traffic-priority-system-life-threatening-emergencies-scdf-4764276**

GovTech Singapore. (2023). *Analytics.gov—Enabling AI/ML Operations for Whole-of-Government (WOG)*. Singapore: Government Technology Agency. **https://www.tech.gov.sg/products-and-services/for-government-agencies/data-and-ai/analytics-gov/**

GovInsider. (2023). "Analytics.gov: Singapore's Whole-of-Government Data Exploitation Platform Provides Advanced Analytics and Machine Learning Capabilities with Cloud." *GovInsider*, May 20. **https://govinsider.asia/intl-en/article/analyticsgov-singapores-whole-of-government-data-exploitation-platform-provides-advanced-analytics-and-machine-learning-capabilities-with-cloud**

IntelliStride. (2024). "Singapore Smart Traffic System Is Redefining Urban Mobility." *IntelliStride Blog*, November 15. **https://www.intellistride.com/blog/how-singapores-smart-traffic-system-is-redefining-urban-mobility/**

Land Transport Authority. (2025). *Intelligent Transport Systems (ITS)*. **https://www.lta.gov.sg/content/ltagov/en/getting_around/driving_in_singapore/intelligent_transport_systems.html**

The Straits Times. (2024). "New Traffic Priority System Allows SCDF Ambulances to Reach Hospital 100 Seconds Faster in Emergencies." November 23. **https://www.straitstimes.com/singapore/new-traffic-priority-system-allows-scdf-ambulances-to-reach-hospital-100sec-faster-in-emergencies**

AI IN URBAN DESIGN

Exploring the Role of Generative AI in Singapore's Urban Future

A Smarter Start to Design

Ari watched the Digital Blue Foam app demonstration with fascination. Digital Blue Foam, a Singapore-based company, offered a user-friendly, web-based AI design tool that helped architects and planners explore innovative ways to design buildings and neighborhoods. Instead of starting from scratch, it used data, like sunlight, zoning rules, and pedestrian movement, to generate design options with the help of artificial intelligence. The goal was to make early-stage planning faster and greener, enabling teams to explore creative layouts that supported energy efficiency and sustainability. Singapore's Urban Redevelopment Authority (URA) had tasked Ari's team with assessing whether generative AI platforms like Digital Blue Foam could accelerate sustainable retrofitting for aging public housing estates. Ari's role was to evaluate potential biases in the AI before recommending any pilot program.

Prompting Smarter Urban Design

To test the system, the team decided to look at what Digital Blue Foam would recommend for the MacPherson public housing estate. Developed between the 1950s and 1970s, MacPherson is one of Singapore's earliest public housing estates. The area has been undergoing rejuvenation with new projects. Criteria drawn from Singapore's environmental targets and community design goals were entered into the app:

- Maximize shaded pedestrian walkways.
- Increase communal green spaces by at least 20 percent.
- Optimize buildings for cross-ventilation.
- Preserve sightlines to cultural landmarks.
- Ensure new designs met Green Mark energy standards.

Within minutes, the AI produced dozens of options, some with rooftop gardens, others with shaded courtyards or integrated community farming. The results were diverse, data-driven, and promising—but they raised key questions.

Bias in the Training Data

One of Ari's first concerns was about cultural fit. Other applications considered had created AI-

generated layouts that leaned heavily on Western-style zoning and design assumptions, grids, wide vehicle lanes, or low-density structures not suited to Singapore's tropical climate or dense neighborhoods.

Research showed this was not unique: studies from Nature and arXiv documented how Generative AI models tend to default to stereotypical Western patterns, even when asked to model cities in Asia or Africa.

"These systems often reproduce the patterns they were trained on," Ari noted. "And those patterns were incompatible with Singapore."

To evaluate how well the AI-generated designs matched the lived experience of Singaporean communities, Ari's team invited residents to explore the models using VR headsets and tablets and voiced preferences:

- "Add more seating near the playground."
- "Can we get shade here for elderly residents?"

This feedback loop showed how AI could accelerate ideation while humans provided nuance, memory, and cultural context.

Ethical and Technical Questions

In addition to bias, Ari documented several risks and considerations:

- ***Ethical tradeoffs.*** Some AI-optimized layouts reduced communal spaces in favor of energy savings.
- ***Regulatory challenges.*** Promising ideas sometimes violated URA building codes or accessibility requirements.
- ***Equity in access.*** Without thoughtful design, AI tools might prioritize efficiency at the expense of inclusivity, reducing access or usability for certain groups.

Ari concluded that generative AI had enormous potential—but only if paired with human oversight, local data, and strong ethical frameworks. If this software were used, they would definitely need to incorporate a human-in-the-loop (HITL) review process.

The Bigger Picture

Ari's report didn't recommend immediate rollout. Instead, it proposed a limited-use pilot with embedded human review, targeted training on local design principles, and clear benchmarks for evaluating success.

As they wrapped their findings under the shade of an old banyan tree at the estate—a tree residents had demanded be preserved—Ari reflected, "Generative AI can show us what's possible. But only

people can decide what's worth preserving."

In a Smart Nation, progress means using AI to build with culture and heritage in mind.

Discussion Questions

1. What benefits did Ari's team discover in using generative AI tools like Digital Blue Foam for early-stage urban planning?
2. How did bias in the AI's training data affect the design outcomes, and what steps did the team take to address it?
3. What role did community feedback and immersive technology (like VR) play in shaping the design process?
4. Why is human-in-the-loop (HITL) oversight essential when applying AI to public-sector projects?
5. Should governments adopt speculative AI tools for public planning before they are fully proven? What safeguards make such pilots more responsible?

Case References

Building and Construction Authority. (2024). *Green Mark Certification Scheme*. https://www1.bca.gov.sg/buildsg/sustainability/green-mark-certification-scheme

Digital Blue Foam. 2022. *Digital Blue Foam Launches for Soaring AI AEC Market*. January 5, 2022. **https://architosh.com/2022/01/digital-blue-foam-launches-for-soaring-ai-aec-market/**

Nature Portfolio. 2024. "Place Identity: A Generative AI's Perspective." *Humanities and Social Sciences Communications* 11(1). **https://www.nature.com/articles/s41599-024-03645-7**

Unite.AI. 2024. "Western Bias in AI: Why Global Perspectives Are Missing." *Unite.AI*, January 15, 2024. https://www.unite.ai/western-bias-in-ai-why-global-perspectives-are-missing/

Urban Redevelopment Authority. (2024). *Master Plan and Planning Guidelines*. S**https://www.ura.gov.sg/Corporate/Planning/Master-Plan**

Ye, Shan. 2023. "Generative AI May Prefer to Present National-Level Characteristics of Cities Based on Stereotypical Geographic Impressions at the Continental Level." *arXiv* preprint. **https://arxiv.org/abs/2310.04897**

PERSONALIZING GOVERNMENT: THE LifeSG PROJECT

A New Kind of Public Service

Ari arrived at the Smart Nation and Digital Government Office (SNDGO) early, tablet in hand and questions on their minds. They were joining the LifeSG team, a cross-agency initiative with an ambitious goal: to personalize digital government services for every resident of Singapore.

From Moments of Life to LifeSG: A Platform Evolves

LifeSG was the evolution of the earlier *Moments of Life* project, which had streamlined parenthood services by integrating birth registration, savings programs, and child care recommendations. With the successful development of an app that helped new parents complete required government forms and find needed services, the team sought to expand the platform to serve all Singapore residents similarly. With support from GovTech and SNDGO, they were launching an AI-powered app to recommend government services to individual users based on real-time life events, personal preferences, and previous interactions.

They had developed the AI expertise in other projects and the recommendation model and interface were developed in-house, built on Singapore's National Digital Identity platform and public service datasets. This allowed for deeper integration with existing infrastructure and more control over citizen data use and protection.

Pilot Integration and AI Functionality

Ari's role was to help as an analyst oversee the pilot integration of the AI engine into the LifeSG interface. At first glance, the system was simple: Based on changes in employment, household size, income, or health needs, LifeSG would suggest government programs or actions—everything from scheduling a mammogram to applying for rental assistance. But under the hood, it relied on a combination of rule-based logic and machine learning trained on de-identified historical data. Rule-based logic ensured compliance with regulations, while machine learning enabled more personalized, adaptive recommendations.

The technical aspects were complex, but the organizational challenges were more so. The app would provide assistance and recommendations 24/7, requiring a shift in how many government services were delivered. The rollout was expected to significantly affect job roles across multiple agencies, with early projections suggesting major workforce shifts.

At one design session, a veteran caseworker expressed concern: "*If the app makes the recommendations... what's my job now*?"

Reshaping Public Service Roles

Ari had anticipated this. They'd been briefed on Singapore's *AI Job Redesign Guide*—a framework aimed at transforming jobs in partnership with workers, not at their expense. Drawing from that, the team worked with HR and department heads to redefine job scopes. LifeSG would direct people to the appropriate forms and services, and public officers would focus on high-touch support, exception cases, and digital inclusion outreach.

Workshops were held to train employees on the new system and engage them in designing job duties and workflows that make sense in an AI-augmented world.

Addressing Explainability and Trust

Meanwhile, ethical concerns surfaced during internal testing. The system sometimes recommended services that didn't reflect a user's cultural norms, language preferences, or life goals. Ari proposed adding a *"Why am I seeing this?" explainer* to every recommendation, allowing users to adjust personalization settings or opt-out entirely.

The team relied on the Model AI Governance Framework and ISAGO (Implementation and Self-Assessment Guide for Organizations) to guide deployment, emphasizing transparency, fairness, and human-centric design. Ari helped build a data audit trail that tracked how recommendations were generated, not just for debugging but also to ensure algorithmic accountability, ensuring that the AI's decisions could be explained, reviewed, and corrected if necessary.

Citizen Adoption and Community Engagement

Six months later, the pilot launched in several neighborhoods. Digital ambassadors, former public officers who had shifted into new community-facing roles, supported residents with onboarding. In one community center, an older woman navigating the app turned to her volunteer and said, "*It's like it already knows me... but nicely."*

Ari smiled when they heard that. The personalization worked, not just because of the model, but because the organization had been willing to change how it worked.

Smart Cities around the World

While LifeSG is tailored to Singapore's governance and digital infrastructure, it reflects a global shift. Cities such as Barcelona, Amsterdam, Copenhagen, and New York also embrace Smart City

initiatives to improve urban services using AI, data, and citizen-centered design. Unlike many decentralized systems, Singapore's tightly integrated digital infrastructure has enabled faster deployment of ethical, AI-powered services. Singapore's approach offers a model for other governments and organizations navigating similar transitions.

Discussion Questions

1. How does AI adoption shift employees' roles in service-oriented organizations?
2. Why might a government prefer to develop AI systems in-house rather than outsource to a vendor?
3. What ethical safeguards are necessary when deploying recommendation engines in public services?
4. How do frameworks like ISAGO help organizations implement AI responsibly?
5. What lessons from LifeSG could be applied to a private company adopting AI for the first time?

Case References

GovTech Singapore. "LifeSG Platform Overview." Accessed April 18, 2025. **https://www.developer.tech.gov.sg/products/categories/platform/lifesg/overview**

GovTech Singapore. *LifeSG – Services Made Simpler.* Accessed April 18, 2025. **https://www.life.gov.sg/**

Personal Data Protection Commission. "Singapore's Approach to AI Governance." Singapore Government Agency Website. Last modified November 3, 2023. Accessed April 19, 2025. **https://www.pdpc.gov.sg/Help-and-Resources/2020/01/Model-AI-Governance-Framework**

Smart Nation and Digital Government Group. *AI in Human Resource: A Guide to Job Redesign in the Age of AI.* Singapore: Government of Singapore, 2020. **https://file.go.gov.sg/ai-guide-to-jobredesign.pdf**

TELLING THE SMART NATION STORY

Ari's New Assignment

Ari leaned back in their chair at the Smart Nation and Digital Government Office (SNDGO), rereading the email. They had been selected to join Singapore's delegation to the upcoming *Smart City Expo World Congress* in Barcelona, one of the world's leading urban innovation and digital transformation events. Being one of just a handful of professionals representing their country was a significant honor.

Their role was to *prepare and deliver a briefing on Singapore's Smart Nation initiative*, reflecting on its evolution, impact, and lessons for other cities. The theme of the session? *"Smart Cities in Action: AI, Infrastructure, and Inclusion."*

Calling in AI for Assistance

With only three weeks to prepare, Ari turned to a closed-domain generative AI assistant trained on Singapore government documents, developed by GovTech to ensure data security and domain relevance. Ari had access to commercial GenAI platforms, but for this project, they chose GovTech's closed-domain generative AI system, a tool trained exclusively on verified government documents, agency reports, and policy briefs. The system's narrower scope meant stronger security protocols and a lower risk of spitting out off-topic or fabricated responses.

They began with a broad prompt: "*Summarize Singapore's Smart Nation initiative from launch to present day. Highlight milestones in mobility, housing, health care, and citizen services.*"

The system responded quickly with a well-structured outline, but it sounded generic, like a polished brochure. It had the facts, but no heart, no context, and no real sense of audience. Worse, it crammed too many ideas into a single response.

Ari sighed and leaned back. Then, they remembered something from a recent workshop: the *FOCUS framework* for prompting generative AI.

They reviewed it mentally:

- ***Focused.*** Be specific and provide clear direction.
- ***Objective.*** Know what outcome you're aiming for.
- ***Context.*** Add background the AI needs to tailor its response.
- ***Unambiguous.*** Use precise and direct language.
- ***Single query.*** Avoid overwhelming the AI with too many tasks at once.

Realizing their first attempt violated nearly all five principles, Ari cleared the chat and started over. This time, they decided to treat the AI like a junior collaborator—not a mind reader.

Applying the FOCUS Framework

This time, they tried: *"Rewrite this for a global Smart Cities audience. Include both successes and unresolved challenges. Use a confident but humble tone."*

The latest version acknowledged digital divide concerns, privacy debates over TraceTogether, and the difficulty of sustaining innovation momentum. It was clearer, more relevant, and appropriately scoped.

Ari then layered in the final element—a personalized touch: *"Write from the perspective of someone who lived through the changes, not just a narrator. What did it feel like to see this transformation happen?"*

The new output began with smart lighting synchronizing with pedestrian flow on Orchard Road and ended with seniors using LifeSG at community centers with help from former case officers. It wasn't perfect, but it was close.

The Hallucination Test

Still, Ari had learned enough about generative AI to know that even a confident answer might be wrong. Verify. Sure enough, one paragraph claimed that Singapore's smart housing program had launched a fleet of solar-powered delivery robots in 2021, an idea floated in early pilot proposals but never implemented. Ari flagged it immediately.

It was a reminder that *verification wasn't optional*: GenAI could summarize beautifully, but only humans could ensure the summary was true.

Breaking Down the Project

Rather than attempting to generate a full report in a single shot, Ari broke the project into smaller, focused tasks and used GenAI to support each part individually.

They asked the AI to help:

- Draft a two-paragraph summary of LifeSG and its societal impacts.
- Generate a visual timeline of Smart Nation milestones.
- Rewrite specific sections for technical and civic audiences.
- Suggest three stories to highlight in the presentation.
- Simulate a Q&A session based on likely questions from global delegates.

By treating the presentation as a series of smaller creative challenges, Ari combined the speed of GenAI with human oversight and narrative voice, a more sustainable approach than relying on a single prompt.

Ari at the Podium

At the forum, Ari took the stage with a mix of pride and caution. The presentation began with a visual timeline generated by AI and ended with a slide that read:

"A smart city does not use the most data, but one that makes the most of its people."

Ari closed with a story about a neighborhood bus stop where citizens submitted design suggestions via an interactive kiosk. The AI had flagged the proposal for a community herb garden as nonessential, but the city implemented it anyway, and the garden became a gathering point for seniors who now taught youth how to grow basil and mint.

Ari's point was clear: Generative AI could assist, but it couldn't replace human values.

Discussion Questions

1. How did Ari apply the FOCUS process when using generative AI to prepare for a complex project?
2. What are the risks of relying on GenAI to represent a national initiative on a global stage?
3. What lessons does this case provide about prompt refinement and the human role in AI output?
4. How can governments use GenAI tools without losing their unique voice or community identity?
5. Could this approach work in your organization or city? Why or why not?

ENDNOTES

[1] National Parks Board. 2022. *Skyrise Greenery Incentive Scheme*. **https://skyrisegreenery.nparks.gov.sg/incentive-scheme/about/**

[2] ST Engineering. 2023. *Smart Estate Lighting*. **https://www.stengg.com/en/smart-city/urban-solutions/smart-utilities-and-infrastructure/agil-smart-lighting/agil-smart-estate-lighting/**

[3] Land Transport Authority (LTA). 2022. *Passenger Information Display Systems*. **https://www.lta.gov.sg/content/ltagov/en/who_we_are/statistics_and_publications/Connect/pids.html**

[4] Land Transport Authority. 2022. *Smart Mobility 2030*. **https://www.lta.gov.sg/**

[5] Monetary Authority of Singapore. 2021. *E-payments in Singapore*. **https://www.mas.gov.sg/**

[6] Infocomm Media Development Authority. (2022). *AI and IoT in Singapore's Retail Sector*. **https://www.imda.gov.sg/business**

[7] Urban Redevelopment Authority. (2023). *Singapore Green Plan 2030*. Retrieved from **https://www.ura.gov.sg/**

[8] Singapore Green Plan 2030. "A City of Green Possibilities." Singapore Government Agency Website. Last modified March 8, 2025. **https://www.greenplan.gov.sg**

[9] Building and Construction Authority. "Green Mark 2021 Certification Standards." *BuildSG*, last updated February 25, 2025. **https://www.buildsg.gov.sg**

Index

A

Accountability, 12–13, 175
Accuracy disparities, 16
Activation functions, 129
Adaptive systems, 119
Advanced prompting techniques, 204–208
  additional techniques and strategies, 206–207
  balancing brevity with detail, 207
  business AI tools, 208
  different AIs, adapting prompts for, 207
  task decomposition, 204–205
The Age of Big Data and Machine Learning (2000-2020), 22
  automation and robotics, 30
  business applications, 29
  computer vision, 30
  customer personalization, 29
  predictive analytics, 30
  technological foundations, 29
The Age of Experimentation (1970-1985), 22, 25
  AI impact on business, 26–27
  expert systems, 26
  microprocessor revolution and technological foundations, 26
  neural networks, advancements in, 26
AI agents, 168
AI-augmented jobs, 33–34
AI conversations, 212
AI divide, 9
AI ethics boards, 33

AI fluency, 34
AI-generated communication, 213
AI governance, 33, 178
AI lifecycle management, 61
AI models, 202–204
Airbnb, 2, 3
Air Canada, 178
AI readiness index, 161
AI voice cloning, 147–149
The AI Winter and Resurgence (1985-2000), 22, 27
  AI and robotics in business, 28
  economic pressures, 27
  resurgence, seeds of renewal, 27
  technological foundations, 27
  technological limitations, 27
  unrealistic expectations, 27
Algorithmic bias, 84
  in machine learning, 80
Altman, Sam, 34
Amazon, 10, 28–30, 34, 51, 63–64, 114
Amazon Prime, 86
Amazon Web Services (AWS), 29, 47, 55
Ambiguous prompt, 196
Apache Spark, 29
Application Programming Interface (API), 46, 143
Area under the curve (AOC), 108, 109
Area under the receiver operating characteristic curve (AUC-ROC), 108, 109
Artificial general intelligence (AGI), 34
Artificial intelligence (AI)
  access, 9–10
  applications of, 5–6
  benchmarks for, 198
  classifications of, 2–5
  creativity, 11–12
  data science and, 6–7

defined, 1
ethical considerations in, 8–13
generative (*See* Generative AI)
in health care, 30
integration, 27
intellectual property (IP) and, 213
opportunities and challenges, 33–34
responses, strategies for, 211
in warfare, 13
Artificial neurons, 23
Artistic foundations, 22
Asimov, Isaac, 22
AT&T, 162
Attention mechanism, 131–132
Authors Guild, 12
Automated guided vehicles (AGVs), 28
Automated weapons systems (AWS), 13
Autoregressive integrated moving average (ARIMA), 104
Autoregressive model, 137, 152
Azure, 55

B

Backpropagation, 3, 129, 130
Backup and disaster recovery, 47
Bain & Company, 165
Balanced scorecard approach, 171, 173
Balancing brevity with detail, 207
Bias, 10
amplification, 210, 212
detection and mitigation, 50
generative AI, 142
racial and gender, 16
Bias audits, 32, 61, 80
Biden, Joe, 147

Bidirectional encoder representations from transformers (BERT), 132
Big Data, 28–30, 43, 44
  and data management, 54–56
BioNTech, 183–185
Black box problem, 32
Blockchain, 56–57, 66–68
Blue River Technology, 6
BMW, 141
Boeing 737 MAX, 32, 33
Booking.com, 77
Bootstrap aggregating/bagging, 98
Brighterion, 182
Buolamwini, Joy, 16, 17
Business AI tools, 208
Business functions identification, 163–164
  aligning projects, with business objectives, 166
  customer service, 164
  departmental applications, 165
  marketing, 164–165
  operations and supply chain management, 165
  pilot trap, 166
  success stories, 166
Business outcome metrics, 172
“Buy now, pay later” program, 143

C

California Privacy Rights Act (CPRA), 52, 53
Categorical variables, 75
Centralized storage, 58–60
Change Healthcare, 46, 49
Chatbot, 11, 117, 140, 152, 202
ChatGPT, 134, 142, 146, 156, 199, 202, 203–204, 207, 208, 217, 218, 221
ChatGPT-4 response, 193
ChatPG, 166

Circular fashion, 17
Classification and risk assessment, 74, 77
Claude, 202, 203, 218
Climate Neutral Data Centre Pact, 56
Closed-domain AI, 177
Closed-domain chatbots, 202, 221–222
Cloud-based chatbots, 160
Cloud computing, 29
Clustering algorithms, 78
CMY Cubes, 139, 140
Coca-Cola, 31, 165
Coded Bias, 10, 16–17
Collaborative filtering, 78, 83, 102–103
Complex prompt, 197
Computer networks, 26
Computer vision, 6, 30
Confusion matrix, 105
Consent, 51
Content-based filtering, 83
Context prompt, 194–195
Contract intelligence (COIN), 141
Convolutional layers, 130
Convolutional neural networks (CNNs), 30, 130–131, 136
Copilot, 202, 204, 208
Copyright law, 178
Corporate social responsibility, 56
Corporate Sustainability Reporting Directive (CSRD), 56
COVID-19 pandemic, 36–38, 77
  vaccines, 183–185
"Create Real Magic" platform, 165
Creative and engaging personas, 209
Creativity, 213
Credit risk assessment, 77
CrowdIQ, 125, 126
Cultural readiness, 162

Custom chatbots, using machine learning, 160
Customer churn prediction, 77
Customer data protection, 53
Customer engagement, 140
Customer experience metrics, 172
Customer relationship management (CRM) systems, 44
Customer segmentation and personalization, 74, 77–78
Custom generative AI model, 136
Custom GPTs, 139–140
Cybersecurity
AI-powered, 57–58
  in AI systems, 33
  responses, 49
  threats, protecting against, 47–48
Cybersecurity Framework (CSF) 2.0, 48

D

DALL-E, 31, 134, 140
Darktrace, 65–66
Dartmouth Conference, 23–24
Data
  collection, ethical considerations in, 60–61
  IoT and sensor data, 45
  online and social media, 46
  profiling, 50
  quality and integrity, 49–50
  regulations, 51–53
  sources of, 44–45
  structure and types, 44
Data accessibility, 159
Data availability, 159
Databases, merging, 75
Data Center Light, 55
Data centers, 29, 54

Data cleaning, 7, 50
Data collection, 85
Data completeness, ensuring, 50
Data governance, 50, 53, 159
Data integration, 159–160
Data management, 74, 75–76
  Big Data and, 54–56
  emerging technologies in, 56–60
Data minimization, 51
Data mining, 27
Data ownership, 8–9
Data portability, 51
Data preprocessing, 74–76
Data privacy, 84
Data protection officers (DPOs), 51
Data quality, 159
Data sale opt-out, 52
Data science, 6–7
Data security, 46, 159
  access control, 46–47
  backup and disaster recovery, 47
  challenges, 46
  cybersecurity threats, protecting against, 47–48
  principles, 48–49
  during transmission and storage, 46
Data silos, 44, 159
Data transformation, 75
Data triangulation, 9
Data warehousing, 27
The Dawn of AI (1950-1970), 22, 23
  business practices, impact on, 25
  Dartmouth Conference, 23–24
  early AI programs, 24
  emerging computer technologies, 25
  Turing test, 23, 24

Deception, 11
Decision Intelligence, 182
Decision trees, 76–78, 95–97
Decoder-only models, 153
Deduplication, 50
Deepfakes, 11
Deep learning, 83, 112
  attention mechanism, 131–132
  autonomous vehicles, 123–125
  convolutional neural networks, 130–131
  defined, 113, 114
  emerging capabilities in, 118–120
  language translation, 116
  limitations of, 118
  NLP, 116–118
  recurrent neural networks, 131
  role in business, 115–116
  traditional machine learning *vs.,* 128
  transformers, 131, 132
Deep learning systems, 24
Deep Q-Network (DQN), 73, 74
DeepSeek, 31
Demand and sales forecasting, 76
Denoising process, 153, 154
Detroit Lions, 125–126
Devol, George, 25
Diffusion models, 138, 153–154
Digital Equipment Corporation, 26
Dimensionality reduction, 75
Discrimination, 10
Disney, 218
Distributed cloud-based storage systems, 54–55
Distributed storage, 58
Domain expertise, 7
Domain Name System (DNS), 26

Domain-specific chatbots, 208
Dream Studio, 140
“Dueling” process, 15
Duolingo, 140
Dynamic pricing, 165

E

Eapek, Karel, 22
E-commerce, 53
Economic inequality, 9–10
Edge-case testing, 177
Edge computing, 31, 45, 58–59
Edge storage, 58–59
Effective heart disease prediction system (EHDPS), 114
Elbow method, 92
Electrolux, 171
ELIZA, 24
Elomia, 140
Embeddings, 137, 152
Emerging computer technologies, 25
Encoder-decoder models, 153
Encryption, 46
Energy-efficient AI, 119
Enterprise-level AI chatbots, 160
Environmental sustainability, 9
Ethical AI development, 33
European Union (EU) AI Act (2024), 52
Expert systems, 26
Explainable AI (XAI), 12, 80
External data sources, 45

F

Facial recognition, 114, 125–126
Fairness-aware machine learning, 32

Fairness-aware modeling, 80
Fairness metrics, 50
Family Educational Rights and Privacy Act (FERPA), 53
FDA AI guidance, 53
Feature engineering, 75
Federated learning, 59–60
FedEx, 114
Few-shot learning, 119
Fiber-to-fiber recycling, 17
Filter bubble, 11
Financial forecasting, 76
Financial Industry Regulatory Authority (FINRA), 53
Fine-tuning process, 136
Focused prompt, 192–193
FOCUS framework, 191–197
  context prompt, 194–195
  focused prompt, 192–193
  objective prompt, 193–194
  single query prompt, 196–197
  unambiguous prompt, 196
Ford, 78
Ford Field, 125
Forward propagation, in neural networks, 129–130
Foundation models, 118, 136
Foundations of AI (Pre-1950s), 22
  artistic foundations, 22
  scientific foundations, 23
Fraud detection, 77
French copyright law, 16
F1 score, 107–108
Full self-driving (FSD) systems, 123, 124

G

Gated recurrent units (GRUs), 131

Gemini, 202
General Data Protection Regulation (GDPR), 8, 9, 51–53
General Mills, 166
General Motors, 25
General-purpose generative AI models, 202
Generative adversarial network (GAN), 15, 16, 138, 154
Generative AI, 133–134, 176, 213
  application, 4–5
  artwork, 219–221
  autoregressive generation, 153
  business models in, 143–144
  chatbots and customer engagement, 140
  checklist, for implementation, 178
  in contract review, 141
  copyright and intellectual property, 178
  core generative model types, 153–155
  creative possibilities, 138, 139
  custom GPTs, 139–140
  defined, 4
  description of, 134
  ethical foundations for, 176
  evolution, 156
  hallucinations and misinformation reduction, 176–177
  image generation, 137–138
  legal accountability in, 178
  milestones in, 134
  multimodal and interactive experiences, 141
  multimodal and video generation, 138
  predictive *vs.,* 134–136
  process, 4
  product design, 141
  responses, 138
  risks and limitations of, 142–143
  safeguarding user consent and control, 177
  sampling strategies in, 155–156
  specialized applications, 141–142

text-based applications, 139–140
text generation, 137
tokenization and embeddings, 152
tools, 191
training and onboarding, 141
training generative models, 136–137
transformer architectures, 153
visual content and virtual try-on, 140
workforce disruptions, 177
The Generative AI Revolution (2020-Present), 22, 30
black box problem, 32
ethical, safety, and privacy concerns, 32, 33
Internet of Things, 31–32
privacy risk, 33
safety and risk management, 32–33
scientific and technological advancements, 31
Generative pretrained transformer (GPT), 132
GitHub Copilot, 142
Global divide, 9
GloVe, 117
Goal-oriented prompts, 193, 194
Goldman Sachs, 8
Good prompts, 191
Google, 9, 10, 34, 55
Google Cloud, 55, 185–186
Google Translate, 117
Governmental policies, 51
GPS system, 114
Graphics processing units (GPUs), 160
Green AI, 119
Green Mountain Data Centers, 55

H

Hadoop, 29

Hallucinations, 142, 176–177, 202, 221–222
Hard prompts, 198
Harmonic mean, 107
Harvey, 141–142
Health Insurance Portability and Accountability Act (HIPAA), 53
Hidden layers, 113, 129
High-level programming languages, 26
High-quality data, 49–50
Hiring process, 63–64
H&M Foundation, 17–19
Hong Kong Research Institute of Textiles and Apparel (HKRITA), 17–19
Hootsuite, 117
Human-in-the-loop (HITL), 140, 177
Humanly, 140

I

IBM, 26, 34, 66
  Watson OpenScale, 122–123
IBM Watson for Oncology, 30
Identity and Access Management (IAM), 47
“If-then” rules, 26
Iger, Bob, 218
Image-based generative AI, 136
Implementing AI solutions, 167
  accountability, 175
  AI agents, 168
  checklist, 176
  considerations, 169–171
  diversity, 174
  ethical training data biases, 174
  explainability, 174
  legal and regulatory considerations, 175–176
  risks and benefits, 175
  RPA and IPA, 168
  tools selection, 167–168

transparency, 174
vendors selection, 169
workforce planning, 175
Imputation method, 50, 75
Individual divide, 10
Industrial Revolution, 8, 13
Industry-specific regulations, 53
Industry-specific terminology, 206–207
Informative and analytical personas, 208
Informed consent, 8
Innovation/creativity metrics, 171–172
Innovation strategy, 163
Input layer, 113, 129
Input processing, generative AI, 4
Intellectual property (IP), 11–12, 178
and creativity, 213
generative AI, 142–143
Intelligent process automation (IPA), 168, 183
The Intel 4004 microprocessor (1971), 26
Internal data sources, 44, 45
International Energy Agency, 9
Internet, 27
Internet of Things (IoT), 31–32, 45, 53, 118
InVideo, 141
Item-based collaborative filtering, 102
Iteration, 130
Iterative refinement, 198–202, 204
iTranslate, 117

J

James Webb Space Telescope, 202
JFK Airport, 84–86
John Deere, 6
JPMorgan Chase, 141

K

Kantayya, Shalini, 10
Keyless Signature Infrastructure (KSI) blockchain, 56
Key performance indicators (KPIs), 170
KFC, 77
Klarna, 143
K-means clustering, 75, 78, 91–92
K-nearest neighbors (*k*-NN), 93–94
Knowledge engineers, 26
Krug, Steve, 82

L

Lang, Fritz, 22
Large language models (LLMs), 146, 160
Latent space, 154
Leadership readiness, 162–163
Least privilege, 47
Least squares regression, 95
Lei Geral de Proteçao de Dados Pessoais (LGPD), 53
Lemonade, 168
Linear regression, 76, 94–95
LinkedIn, 51
Liverpool, 149
Local Law 144, 80
The Logic Theorist, 24
Logistic function, 100
Logistic regression, 77, 99–100
Lohr, Steve, 49
Long short-term memory (LSTM) networks, 131
Loss function, 130
Lufthansa Group, 140

M

Machine learning, 27–30, 70, 71

algorithmic bias in, 80
algorithms, 74–76, 78–79, 85
application, 2–3
in business functions, 74–79
collaborative filtering, 102–103
confusion matrix, 105
custom chatbots using, 160
decision trees, 95–97
defined, 2
fairness-aware, 32
k-means clustering, 91–92
k-nearest neighbors, 93–94
linear regression, 94–95
logistic regression, 99–100
metrics for evaluating, 105–111
principal component analysis, 89–91
random forests, 97–99
reducing bias in, 80
reinforcement learning, 71–74
supervised learning, 71–72
support vector machines, 100–101
techniques, 76–78
time series analysis, 103–104
traditional *vs.* deep learning, 128
unsupervised learning, 71, 72
Machine learning operations (MLOps), 160
Maersk, 66
Maneuvering Characteristics Augmentation System (MCAS), 32–33
Manipulation, 11
Manipulative AI, 52
Marginalized groups, impact on, 16
Mastercard fraud detection, 181–183
Mathematical vector, 137
McCarthy, John, 23
McCulloch, Warren, 23

McKinsey, 31
Mean absolute error (MAE), 108–110
Mean squared error (MSE), 108–110
Meta, 12, 51, 55
Metrics, in machine learning, 105
  accuracy, 106
  algorithms, 110
  AUC-ROC, 108, 109
  and evaluation questions, 110–111
  F1 score, 107–108
  MAE/MSE, 108–110
  precision, 106–107
  recall, 107
  RMSE, 109–110
  summaries, 110
Microsoft Azure, 29
Midjourney, 140
MillsChat, 166
Minibatches, 130
Min-max scaling. *See* Normalization
Misinformation, 11, 210
The Mitchells vs. The Machines (2021), 22
Mobile data, 29
Model activation, generative AI, 4
Monitoring and measuring AI effectiveness, 171–173
Morgan Stanley, 118
Multimodal generative AI, 136, 141
Multimodal learning, 118
Multimodal model, 138
Multi-tiered storage model, 58
MYCIN, 26

N

National Institute of Standards and Technology (NIST), 48, 49
National Security Agency (NSA), 54

Natural language processing (NLP), 27, 29, 31, 118, 140, 164
applications, 117
deep learning in, 116, 117
NEC Labs America, 77
Negative prompt, 200
Netflix, 29, 82–84, 114
Network security, 65–66
Neural networks, 27, 112
activation functions, 129
adoption of, 114
advancements in, 26
applications, 3–4
architecture, 129
backpropagation, 129, 130
business applications of, 114
defined, 3, 113
forward propagation, 129–130
layers, 3, 113
limitations of, 118
model, 23
model growth, 120
office building analogy, 130
overfitting, 130
real-world impacts, 114
stochastic gradient descent, 130
Newell, Allen, 24
The New York Times, 185–186
Nondiscrimination, 52
Non-player characters (NPCs)., 5
Normalization, 75
Nucleus (top-p) sampling, 156
Numerical variables, 75

O

Objective prompt, 193–194
Ochsner Health, 47, 49
O'Neil, Cathy, 10
Online data, 46
On-Road Integrated Optimization and Navigation (ORION), 86, 87
OpenAI, 12, 146–147, 165
  ChatGPT-4, 203
  GPT-3 and GPT-4, 31
Open-domain (open-ended) chatbots, 202, 221–222
Operational efficiency metrics, 172
Optical character recognition (OCR), 3
Organizational readiness assessment, 158
  business process readiness, 160
  cultural readiness, 162
  data integration, 159–160
  data readiness, 159
  data silos, 159
  factors, 159
  HR's role, in building AI-ready teams, 161
  lack of standardization, 159
  leadership readiness, 162–163
  technological readiness, 160
  workforce readiness, 161
  workforce reskilling and support, 161–162
Orlowski, Jeff, 11
OSAIRIS, 38–39
Out-of-bag (OOB) samples, 98
Output delivery, generative AI, 4
Output layer, 113, 129

P

Payment Card Industry Data Security Standard (PCI DSS), 53
Perceptron, 24, 26
Personal data, 8

Personal Information Protection and Electronic Documents Act (PIPEDA), 53
Personal Information Protection Law (PIPL), 53
Personalized medicine, 206
Personas, 208
  creative and engaging, 209
  informative and analytical, 208
  in prompt engineering, 208–210
  supportive and reflective, 209
Pfizer, 183–185
Phenom, 171
Pilot trap, 166
Pitts, Walter, 23
Poor-quality data, 49
The Portrait of Edmond de Belamy (case), 15–16
Positive predictive value (PPV), 106–107
Power usage effectiveness (PUE), 56
Prediction and forecasting, 74, 76
Predictive AI
  defined, 134
  types of, 136
  vs. generative AI, 134–136
Predictive analytics, 29, 30
Predictive maintenance, 78
Predictive modeling, 7
Pressurized water reactor technology, 55
Pretraining, 136, 137
Principal component analysis (PCA), 75, 76, 89–91
Privacy, 8–9
Privacy-first AI strategies, 33
Privacy-first policies, 33
Process optimization and automation, 74, 78–79
Procter & Gamble, 166
Project InnerEye, 38–39
Project Revoice, 148
Prompt

advanced prompting techniques, 204–208
FOCUS framework, 191–197
hard, 198
iterative refinement, 198–202
strategies for AI responses, 211
understanding AI models, 202–204
Prompt crafting, 191
ethical, 212
Prompt engineering, 191
checklist, 213–214
generative AI output, 213
intellectual property and creativity, 213
personas in, 208–210
prompt crafting, 212
verification, 212

Q

Quality control, 76

R

Racial and gender bias, 16
Random forests, 76, 77, 97–99
Receiver operating characteristic (ROC) curve, 108, 109
Recurrent neural networks (RNNs), 117, 131, 155
Regional divide, 10
Regression analysis, 78
Regulatory compliance, 56, 159
Reinforcement learning, 31, 71–74
Response construction, 203
Retailers, 53
Retrieval-augmented generation (RAG), 142, 146, 177, 198
Right to access, 51, 52
Right to be forgotten policy, 51, 60

Right to correct, 52
Right to delete, 52
Right to know, 52
Robotic process automation (RPA), 31, 78, 168
Robotics
  automation and, 30
  in business, 28
  in manufacturing, 27
Robust AI safety measures, 33
Role-based access control, 47
Root mean squared error (RMSE), 109–110
Rosenblatt, Frank, 24, 26
Rule-based reasoning, 26
Rural Digital Opportunity Fund, 10

S

Sampling strategies, in generative AI, 155–156
Samsung Food, 4–5
Sappi, 168
Scalability, 159
Scientific foundations, 23
Seasonal ARIMA (SARIMA), 104
Second Amendment Foundation (SAF), 146
Self-supervised learning, 118
SendX, 139
Sensor data, 45
Sentiment analysis, 29, 164
Server farms, 29
Service providers and third parties, 52
Set piece, 149
Shakey, 25
Show data, 83
Simeèka, Michal, 147
Simon, Herbert, 24
Simplified prompt, 197

Single query prompt, 196–197
Single source of truth approach, 50
Skills gap analysis, 161
Small modular reactors (SMRs), 55
Smart Data Query (SDQ) AI tool, 183
Smart Garment Sorting System, 17–19
Smartphones, 29
The Social Dilemma (2020), 11
Social media data, 46
Social scoring, 52
Software as a Medical Device (SaMD), 53
Software-as-a-service (SaaS) platform, 167
Specific industry terminology, 206–207
Speech recognition, 6
Spotify, 29, 76, 166
Spotify Model, 166
S-shaped curve, logistic regression, 99
Stable Diffusion, 31, 219
Standardization, 49, 50, 75, 159
Stochastic gradient descent (SGD), 130
Stonemaier Games, 220
Stronghold Games, 220
Structured data, 44
Student privacy, 53
Supervised learning, 71–72
Supply chain management, 165
Supply chain optimization, 78
Supportive and reflective personas, 209
Support vector machines (SVMs), 77, 100–101
Support vectors, 101
Sustainability, 159
  in data storage, 55
Symbolic reasoning, 25
Synthesia, 134, 138, 141
Synthetic data, 80

T

Tachel, Lanett, 218
TacticAI, 149–150
“Take-make-dispose” approach, 17
Target, 71, 72
Task decomposition, 204–205
Telephone Consumer Protection Act, 148
Temperature, generative AI, 156
Tesla, 31
  autopilot system, 123–125
Text-based generative AI, 136
Time series analysis, 78, 103–104
Tokenization, 137, 152
Tokens, 137
Top-k sampling, 156
TradeLens, 66–68
Traditional AI, 135
Transformer models, 117
Transformers, 131, 132, 153
Translation invariance, 131
Transmission Control Protocol/Internet Protocol (TCP/IP), 26
Transparency, 12–13
Tribal Broadband Connectivity Program, 10
Turing, Alan, 23
Turing test, 23, 24

U

Uber, 51
UK-based Bodyswaps, 5
Unambiguous prompt, 196
Unclear prompt, 192
Underrepresentation, in training data, 16
Unilever, 165
Universal Turing Machine, 23

Unstructured data, 44
Unsupervised learning, 71, 72, 118
UPS vehicle routing, 86–87
U.S. copyright law, 15
U.S. Copyright Office, 213
User-based collaborative filtering, 102
User data, 83
U.S. Internal Revenue Service (IRS), 164

V

Vague prompt, 192
Variational autoencoders (VAEs), 154
Verification, 212
Verizon, 173
Virtual reality (VR), 5–6
  deep learning and, 119–120
Visual Artists Rights Coalition, 12
Visualization, 7
Voiceitt, 148

W

Walmart, 72
Walters, Mark, 146–147
Waymo, 31
Weapons of Math Destruction (O'Neil), 10
Weighted connections, 3
Weights and biases, 113
Weizenbaum, Joseph, 24
Wells Fargo, 50
Westworld (2016), 22
WhatsApp, 51
Wise Wizard Games, 220
Women's National Basketball Association (WNBA), 39–41
Word2Vec, 117

2023 Writers Guild of America (WGA), 217–219

Z

Zappos, 164
Zero-shot learning, 119
Zestimate, 76
Zillow, 76
Z-score normalization. *See* Standardization

LO Extended Description (Frontmatter)

Return to LO

The objectives are denoted as follows. LO6-1: Define generative AI and distinguish it from predictive or task-specefic AI; LO6-2: Describe how generative AI models are trained and how they produce content; LO6-3: identify and evaluate key business applications of generative AI; LO6-4: Explain the risks associated with generative AI, including hallucinations, bias, and intellectual property concerns; LO6-5: Describe the business models used to monetize generative AI and explain how organizations capture value from these technologioes.

Return to LO

Basic ethical Extended Description (Frontmatter)

Return to Basic ethical

It explains that implementing AI responsibly requires organizations to address a range of ethical, legal, and regulatory considerations. These challenges apply across many types of AI, including machine learning models, natural language processing tools, and robotic process automation systems. Core concerns-such as bias, data privacy, explainability, accountability, and workforce impact-are not limited to any single AI technology. This section outlines foundational practices for deploying AI ethically and effectively in business environments. It builds on our ethics discussions in other chapters and includes strategies for reducing unintended bias, protecting individual rights, promoting transparency in decision-making, and ensuring that the use of AI aligns with regulatory expectations and organizational values.

Return to Basic ethical

Sustainability Extended Description (Frontmatter)

Return to Sustainability

The paragraph explains the following: As data storage strategies evolve to balance scalability, privacy, and performance, organizations are also grappling with the environmental impact of their infrastructure, emphasizing the need for sustainable storage practices. Data centers require vast amounts of electricity to power servers and maintain cooling systems, making sustainability a key consideration for businesses. Several drivers are propelling companies toward more sustainable practices in their storage solutions.

Return to Sustainability

Figure 6.3 Extended Description (Frontmatter)

Return to Figure 6.3

The prompt asks to write a product description for a smartwatch. Under the tokenization process, the prompt is broken into multiple words, such as “write”, “a”, “product”, and “description”. Under the embedding phase, the words are converted to numerical vectors. After proceeding through the generative AI model, the response reads “Stay ahead of your day with the smartmate X2 smartwatch. Monitor your heart rate, track your...”.

Return to Figure 6.3

Figure 8.1 Extended Description (Frontmatter)

Return to Figure 8.1

The human with an objective provides a prompt to the generative AI, which receives inputs from the training data and past prompts to generate the output. This output is then evaluated by the human.

Return to Figure 8.1

Summary Extended Description (Frontmatter)

Return to Summary

The points describe the following.

1. Generative AI: This type of AI creates new content like text, images, audio, or designs by learning patterns from data. Unlike predictive AI, which forecasts or classifies, generative AI generates original outputs based on prompts.
2. How generative AI works: Generative AI uses deep learning models trained on large datasets to generate content by predicting pieces one at a time, using tokenization and autoregressive techniques.
3. Generative AI business applications: Generative AI is currently employed to automate marketing copy, aid in product design, simulate human voices, analyze contracts, and improve customer service through chatbots and avatars. Its ability to quickly tailor and generate content makes it valuable across various industries.
4. Generative AI risks and limitations: Generative AI has potential issues, including factual inaccuracy, misuse for impersonation or fraud, bias amplification, and lack of transparency. Businesses should implement safeguards, verify outputs, and address ethical considerations before using these tools.
5. Common generative AI business models: Companies generate revenue from generative AI through various models: selling access to customizable foundation models, offering tools for specific tasks, tailoring solutions for industries like healthcare or law, providing infrastructure-as-a-service, or supporting open-source tools with paid services.

Return to Summary

Figure 1.1 Extended Description (Chapter 1)

Return to Figure 1.1

The diagram uses a series of nested, rounded rectangular shapes. The largest, outermost shape is labeled Machine Learning. Inside this shape is a smaller, nested shape labeled Neural Networks. The smallest and innermost shape is labeled Generative AI.

Return to Figure 1.1

Figure 1.2 Extended Description (Chapter 1)

Return to Figure 1.2

The input layer contains two nodes shaped like rounded squares. The hidden layer, enclosed in a rectangle, contains two vertical columns of four diamond-shaped nodes each. The output layer contains a single circular node. Lines with arrows show that each node in one layer is fully connected to every node in the next layer, moving from left to right.

Return to Figure 1.2

Figure 1.3 Extended Description (Chapter 1)

Return to Figure 1.3

It begins with the User Input stage, where the user submits a prompt describing the desired response. Next, in the Input Processing stage, the AI analyzes and encodes the input to understand its context. Then, in the Model Activation stage, the AI uses its training data set and the encoded prompt to generate a response one word at a time. Finally, in the Output Delivery stage, the response is sent back to the user, who may refine their prompt to go through the process again.

Return to Figure 1.3

Unfigure 1.1 Extended Description (Chapter 1)

Return to Unfigure 1.1

At the top of the screen, the status bar shows the time 10:00, a hole-punch camera, and icons for network and battery. The main header reads Shopping list in large, bold text with a small dropdown arrow next to it. To the right of the header are a shopping cart icon and a three-dot menu icon. Below the header, it says Sort by: Aisle and to the right is a Share button. The main content area features a header bar with three columns: “Vegetables”, “Price”, and “Quantity”. Below this header, a list of vegetables is presented. The first item is “Potato”, accompanied by a shopping cart icon, a price of 3.00, and a quantity of 2. The next item is “Onion”, with a checkbox next to it. Following that is “Tomato”, also with a checkbox. The fourth item is “Eggplant”, which has a shopping cart icon, a price of 2.00, and a quantity of 1. “Cauliflower” is next, with a checkbox. Then comes “Broccoli”, with a shopping cart icon, a price of 4.00, and a quantity of 1. “Cabbage” is listed with a checkbox. “Capsicum” follows with a checkbox. “Carrot” is also listed with a checkbox. “Beetroot” has a checkbox. The last item visible on the list is “Peas”, which has a shopping cart icon, a price of 2.00, and a quantity of 3. At the bottom of the screen, there is a navigation bar with four icons and corresponding text labels: “Home”, “Recipes”, “Planner”, and “List”. The “List” icon and text are highlighted. Below this app navigation bar, the standard Android system navigation bar is visible, showing a back arrow, a home circle, and a recent apps square.

Return to Unfigure 1.1

Unfigure 2.1 Extended Description (Chapter 2)

Return to Unfigure 2.1

It exhibits cylindrical components, each with a central circular element and a colored outer ring. These components are neatly arranged in a grid-like pattern, separated by horizontal wooden shelves. The entire assembly is housed within a dark gray or black frame, giving it a vintage or industrial appearance. The overall impression is one of intricate engineering and a bygone era of technology.

Return to Unfigure 2.1

Figure 2.1 Extended Description (Chapter 2)

Return to Figure 2.1

Entity A is represented by a computer setup, including a monitor, keyboard, and CPU. Entity B is a person sitting at a desk with a laptop. Entity C is also a person sitting at a desk with a laptop. A document icon is positioned between A and C, and another between B and C. Arrows point from these documents towards entity C, suggesting that C receives information or documents from both A and B.

Return to Figure 2.1

Unfigure 2.2 Extended Description (Chapter 2)

Return to Unfigure 2.2

The conversation consists of a series of exchanges where ELIZA asks about the user's feelings and problems, and the user responds with expressions of unhappiness and a desire for help. The dialogue reveals a pattern of empathetic questions from ELIZA, such as "How do you feel about being unhappy?" and "Would it really help you to get some help that much seems certain?" The user mentions feeling unhappy and suggests that learning to get along with their mother might help. ELIZA emphasizes the importance of family relationships and questions the user's assumptions about their mother's care. The overall tone of the conversation reflects a typical therapeutic dialogue, aimed at exploring the user's emotions and relationships.

Return to Unfigure 2.2

Figure 3.1 Extended Description (Chapter 3)

Return to Figure 3.1

The sources are denoted as internal data sources, external data sources, IOT and sensor data, and online and social media data. Internal data sources include machine-generated data, operational data, financial insights, and customer insights. External data sources include market trends, demographic insights, and competitor analysis. IOT and sensor data include smart appliances, wearables, and industrial sensors. Online and social media data sources include posts, likes, shares, comments, and interactions.

Return to Figure 3.1

Figure 3.2 Extended Description (Chapter 3)

Return to Figure 3.2

The points are denoted as follows:

1. Governance and risk management;
2. Asset management;
3. Data protection;
4. Access control;
5. Continuous monitoring and incident response;
6. Awareness and training.

Governance and risk management include policies(develop and enforce policies for managing cybersecurity risks), roles and responsibilities(Clearly define and communicate cybersecurity roles, responsibilities, and authorities to foster accountability and continuous improvement), and supply chain risk management(integrate cybersecurity supply chain risk management into overall risk management processes and ensure continuous monitoring). Asset management includes inventory and classification(maintain inventories of hardware, software, services, and data, prioritizing them based on the classification, critically and impact on the mission), lifecycle management(manage assets throughout their lifecycle, from acquisition to disposal, ensuring security at each stage). Data protection include confidentiality, integrity, and availability(protect sensitive data at rest and in transit through encryption to ensure its confidentiality integrity, and availability), vulnerability audit(promote regular security audits to identify vulnerabilities), and backup recovery(Regularly create, protect, maintain, and test backups to ensure data can be restored in case of loss or corruption). Access control includes identity management (managing identities and credentials for authorized users, services, and hardware, ensuring proper authentication and access control) and least privilege (granting users, applications, and systems the minimum access necessary to perform their specific tasks). Continuous monitoring and incident response include monitoring (monitoring networks, environment, and activities to detect anomalies and potential cybersecurity incidents) and incident management (developing and executing incident response plans, including analysis, migration, and communication, to effectively manage and recover from cybersecurity incidents). Awareness and training include general awareness (provide cybersecurity awareness and training to all personnel to ensure they understand and can mitigate cybersecurity risks) and specialized training (offer targeted training for individuals in specialized roles to equip them with the necessary skills to handle specific cybersecurity tasks).

Return to Figure 3.2

Figure 3.3 Extended Description (Chapter 3)

Return to Figure 3.3

1 Petabyte (PB) equals 1,000 Terabytes (TB), which is equivalent to about 500 billion pages of standard text, roughly 4,000 smartphone storage capacities (256 GB each), and could store about 13 years of continuous HD video streaming. 1 Exabyte (EB) equals 1,000 Petabytes (PB), which is comparable to all words ever spoken by humans, roughly half of the entire digital data created worldwide in 2007, and could store over 3 million years of high-quality video.

Return to Figure 3.3

Figure 3.4 Extended Description (Chapter 3)

Return to Figure 3.4

An infographic titled How Does Blockchain Work. To the right of the title is a graphic of interconnected, outlined cubes forming a chain, with a solid line running through them. The process is explained in a sequence of six numbered steps, arranged in two rows.

Step 1 is titled A wants to send money to B. The accompanying illustration shows a person icon labeled A next to a laptop with a Wi-Fi symbol. An arrow points from the laptop to a smartphone held by a person icon labeled B.

Step 2 is titled The transaction is represented online as a block. The illustration shows a hexagonal block containing a database symbol (three stacked discs). A key icon floats beside the block.

Step 3 is titled The block is broadcasted to every party in the network. The illustration shows a central hexagonal block with a database symbol, connected by lines to various computer devices, including desktops, laptops, and a smartphone, arranged in a circle around it.

Step 4 is titled The network approves the transaction. The illustration is similar to step 3, showing the network of devices connected to the central block. However, each device now has a checkmark on its screen, signifying approval.

Step 5 is titled The block is added to the existing blockchain in a transparent and unalterable way. The illustration shows a grid of interconnected cubes representing the blockchain. One cube, representing the new block, is being added to the chain and has a padlock icon on it.

Step 6 is titled The transaction is complete. The illustration shows person A with a laptop displaying a large checkmark. An arrow points to person B with a smartphone displaying a database/coin symbol and a Wi-Fi symbol, indicating the money has been received.

Return to Figure 3.4

Figure 3.5 Extended Description (Chapter 3)

Return to Figure 3.5

The diagram illustrating the architecture of edge computing in relation to the Internet of Things (IoT) and cloud/data center storage. At the bottom of the diagram is a horizontal line labeled Internet of Things. Connected to this line are various icons representing IoT devices. From left to right, these are: an electric motor, a gauge, a car, a robotic arm on a conveyor belt, a wind turbine, a solar panel, electrical power lines, a symbol for water, and a small plant in a pot. Data from the IoT devices flows upwards to a middle layer. This layer contains three main components connected by a horizontal line. In the center is an icon of a server rack next to a computer monitor, labeled Edge Storage. To its left is a computer chip, with the word EDGE written vertically beside it. To the right of the Edge Storage is an icon of a monitor displaying a bar chart, with an antenna on top indicating wireless communication. Bidirectional arrows connect the EDGE chip and the monitor with the antenna to the central Edge Storage. From the Edge Storage, a line extends upwards and splits to connect to the top layer, which represents larger storage solutions. On the left, under the heading Distributed Storage, is an icon of a cloud labeled Cloud. On the right, under the heading Centralized Storage, is an icon of multiple server racks in a data center, labeled Data Center.

Return to Figure 3.5

Figure 3.6 Extended Description (Chapter 3)

Return to Figure 3.6

At the top center is an icon of a server rack, representing the central server. An arrow points from the server to a box on the left labeled Any Federated learning algorithm. At the bottom of the diagram, four different client setups are shown in a row. From left to right, they are labeled Laptop, EDGE device, Computer, and Smartphone, each with a corresponding icon. Below each device icon is a database icon labeled Local data. A pair of curved arrows between each device and its local data indicates a cyclical process of training on that data. Above each device icon is a diagram of a small, fully connected neural network, representing a local machine learning model. The arrows point upwards from each of the four neural network models to the central server. A text label on the right side of these arrows reads, Model updates are sent to the SERVER.

Return to Figure 3.6

Figure 4.1 Extended Description (Chapter 4)

Return to Figure 4.1

On the far left, a large box is labeled “Labeled Training Data”. Inside, there are five images of vehicles, each with a text label. From top to bottom, they are: a hatchback labeled “Car”, a sports car labeled “Car”, a standard motorcycle labeled “Motorcycle”, a sport motorcycle labeled “Motorcycle”, and a sedan labeled “Car”. An arrow points from the “Labeled Training Data” box to a circle in the center labeled “Model Training”. The circle contains an icon of several gears. From the “Model Training” circle, another arrow points to a smaller circle labeled “Prediction”. Below the “Model Training” and “Prediction” circles is a box labeled “Test Data”. This box contains two unlabeled images: a dirt bike and a sedan. An arrow points upwards from the “Test Data” box to the “Prediction” circle. From the “Prediction” circle, a final arrow points to a box on the far right. This box shows the model's output. It contains the same two images from the “Test Data” box, now with predicted labels. The dirt bike is labeled “Motorcycle”, and the sedan is labeled “Car”.

Return to Figure 4.1

Figure 4.2 Extended Description (Chapter 4)

Return to Figure 4.2

On the far left is a large box labeled “Input raw data” at the top and “Unlabeled data” at the bottom. Inside this box is a collection of various fruits: red apples, a green apple, single cherries, pairs of cherries, a bunch of three cherries, and oranges, all mixed together. An arrow points from the input box to a sequence of three circular icons representing the process. The first circle is labeled “Interpretation” and contains an icon of a large gear over smaller document and arrow symbols. An arrow leads to the second circle, labeled “Algorithms,” which contains an icon representing a flowchart with a document and a gear. Another arrow leads to the third circle, labeled “Processing,” which has the same gear icon as the “Interpretation” stage. From the “Processing” circle, a branching arrow points to three separate output boxes on the far right, under the collective label “Outputs.” Each box contains a sorted group of fruit. The top box contains only apples: two red and one green. The middle box contains only cherries: a single cherry, a pair, and a bunch of three. The bottom box contains only oranges: two of them. This demonstrates the process of taking unlabeled, mixed data and sorting it into distinct, categorized outputs.

Return to Figure 4.2

Figure 4.3 Extended Description (Chapter 4)

Return to Figure 4.3

The image illustrates a data processing system, likely a reinforcement learning or classification model. On the far left, a rectangular box labeled “Input raw data” sits above a circular container filled with a heterogeneous mix of shapes: triangles, squares, and pentagons. An arrow extends from this circular container, splitting into two paths. One path leads to a box at the bottom center labeled “Agent,” with the arrow segment labeled “State.” The other path leads to a box at the top center labeled “Environment.” Within the central processing area, the “Agent” and “Environment” boxes interact. An arrow from the “Agent” points to the “Environment,” labeled “Best action.” Conversely, an arrow from the “Environment” points to the “Agent,” labeled “Reward.” Between these two boxes, an icon of three interlocking gears is present, with the text “Selection of algorithm” positioned below them, indicating a computational process. Finally, an arrow originates from the “Agent” box and points to the right, leading to a rectangular box at the top right labeled “Output.” Below the “Output” box, three separate rectangular containers are arranged vertically. Arrows from the central processing area (implicitly from the Agent after processing) point to each of these output containers. The top output container holds four triangles, the middle container holds four squares, and the bottom container holds four pentagons. This signifies that the initial mixed raw data has been successfully sorted or classified into distinct categories based on shape.

Return to Figure 4.3

Figure 4.4 Extended Description (Chapter 4)

Return to Figure 4.4

This diagram illustrates the architecture and interaction of a Deep Q-Network (DQN) artificial intelligence with a Breakout game environment. At the top, a label indicates “DQN AI.” Below this, a neural network is depicted with three main layers: an input layer, a hidden layer, and an output layer. The input layer consists of two square-shaped nodes. These nodes receive “Observations” from the “Breakout game,” which is represented by a rectangular box at the bottom. The observations include details such as “Puck movement,” “brick impacted,” and “scoring end?” Arrows indicate the flow of these observations into the input layer nodes. The hidden layer is a larger rectangular region containing eight diamond-shaped nodes, arranged in two columns of four. Arrows show connections from each input layer node to all four nodes in the first column of the hidden layer. Further arrows connect each node in the first column of the hidden layer to all four nodes in the second column of the hidden layer, indicating processing within this layer. The output layer consists of a single circular node. Arrows connect all four nodes from the second column of the hidden layer to this output node. The output node generates an “Action: Paddle movement,” which is sent back to the “Breakout game” via an arrow. Completing the feedback loop, the “Breakout game” provides a “Reward: Score” back to the AI. An arrow indicates this reward feeding into the input layer nodes, influencing future decisions of the DQN AI. The overall diagram demonstrates how the AI perceives the game state, processes it through its neural network, makes a decision, and learns from the resulting reward.

Return to Figure 4.4

Figure 4S.1 Extended Description (Chapter 4S)

Return to Figure 4S.1

At the top, the features around before PCA include Age, Income, Employment Length, Loan Amount, Loan Purpose, Interest Rate, Credit Score, Number of Open Credit Lines, Total Debt, Monthly Debt Payments, Home Ownership Status, Education Level, Marital Status, Number of Dependents, Recent Delinquencies, Bank Account Balance, Credit Card Utilization Rate, Years at Current Address, Years at Current Address, Number of Recent Credit Inquiries, Debt to Income Ratio, and Length of Credit History. At the bottom, the diagram denotes 4 primary components after PCA, which include PC1-financial stability, PC2-credit behavior, PC3-loan characteristics, and PC4-demographic components. PC1 components include Bank Account Balance, Credit Card Utilization, Rate, Income, and Monthly Debt Payments. PC2 components include Credit Score, Number of Open Credit Lines, Number of Recent Credit Inquiries, Recent Delinquencies, and Length of Credit History. PC3 components include Home Ownership Status, Interest Rate, Loan Amount, and Loan Purpose. PC4 components include Age, Education Level, Number of Dependents, Type of Employment, and Marital Status.

Return to Figure 4S.1

Figure 4S.2 Extended Description (Chapter 4S)

Return to Figure 4S.2

On the left, the 2-D plot, titled Before K-means, denotes a random distribution of circles, all shaded in the same color. On the right, the 2-D plot, titled After K-means, exhibits the circles grouped as 3 distinct colors.

Return to Figure 4S.2

Figure 4S.3 Extended Description (Chapter 4S)

Return to Figure 4S.3

The diagram illustrates the K-Nearest Neighbor classification algorithm using clusters of various animals. Five distinct groups of animals are distributed across the image. In the upper left, a cluster of horses is arranged. In the upper central area, a group of sheep is visible. In the upper right, a collection of ducks forms another cluster. In the lower left, a large cluster of cows is present. In the lower right, a group of pigs is shown. A question mark, representing a data point to be classified, is positioned in the lower central part of the image, situated between the cow cluster and the pig cluster. Two concentric circles are drawn around this query point. The smaller, inner circle is labeled “k = 1” and encompasses the query point and one pig. The larger, outer circle is labeled “k = 4” and encompasses the query point, three cows, and one pig. This visual demonstrates that if k=1, the query point would be classified as a pig, as its single nearest neighbor is a pig. If k=4, the query point would be classified as a cow, as three of its four nearest neighbors are cows, forming the majority.

Return to Figure 4S.3

Figure 4S.4 Extended Description (Chapter 4S)

Return to Figure 4S.4

The y-axis, representing the total riders, ranges from 0 to 10,000 at an interval of 1000. The x-axis, denoting the degrees of Celsius, ranges from 15 to 35 at an interval of 5. The datapoints have a decreasing trend, which is highlighted through a dotted line. The y value is denoted as minus 92.228x + 9170.3. An intense fluctuation of datapoints is visible between 20 and 25 degrees Celsius.

Return to Figure 4S.4

Figure 4S.5 Extended Description (Chapter 4S)

Return to Figure 4S.5

The root node branches into two internal nodes. The left internal node has two leaf nodes, while the right internal node has one leaf node.

Return to Figure 4S.5

Figure 4S.6 Extended Description (Chapter 4S)

Return to Figure 4S.6

It begins by assessing whether the applicant's GPA is greater than 3.47 or the SAT score is at least 1420; if yes, the application is accepted. If not, it checks if the GPA is at least 2.5; if not, the application is rejected. For applicants with a GPA of 2.5 or higher, the process then evaluates the significance of their extracurricular activities. If the activities are significant, it checks whether the application has a strong recommendation; if yes, the application is accepted. If not, the process further assesses whether they have a compelling personal statement. If they do, they are accepted; if not, the application is rejected. For those with insignificant extracurricular activities, the decision depends on whether they demonstrate a strong recommendation and a compelling personal statement; if both are affirmative, the application is accepted; otherwise, it is rejected.

Return to Figure 4S.6

Figure 4S.7 Extended Description (Chapter 4S)

Return to Figure 4S.7

It begins with the training data instance, branching into three separate decision trees, each trained on the same dataset but potentially producing different classifications due to their structure. Each tree independently classifies the input data into either Class A or Class B by evaluating various features and making splits based on certain criteria. The results from these trees are then combined through Begging, a voting process, specifically by majority voting, where the class with the most votes is selected as the final prediction. The overall process culminates in a final classification, which in this case is determined as Class A.

Return to Figure 4S.7

Figure 4S.8 Extended Description (Chapter 4S)

Return to Figure 4S.8

The horizontal axis is labeled X, and the vertical axis is labeled Y, ranging from 0 at the origin to 1 at the top. A horizontal dashed line extends from the Y-axis at the value of 1 across the graph. The sigmoid curve starts near Y equals 0 for low X values, gradually increases, then rises steeply through the middle of the graph, and finally flattens out, approaching Y equals 1 for high X values. Scattered data points are present on the graph. Several data points are clustered along the X-axis near Y equals 0 for low X values. Other data points are clustered near the top of the graph, along the dashed line at Y equals 1, for high X values. Text on the graph states, “Predicted Y lies between 0 and 1 range.”

Return to Figure 4S.8

Figure 4S.9 Extended Description (Chapter 4S)

Return to Figure 4S.9

In the 2-D plane, the horizontal axis is denoted as X1, while the vertical axis is X2. It displays two distinct classes of data points: one represented by circles and the other by squares, which are linearly separable. Two parallel lines drawn separate the data points. The gap between them is denoted as 2 over $||w||$. The lines on the left and right represent an equation, $w*x - b = -10$. A dotted line is drawn between the parallel lines, representing an equation, $w*x - b = 0$. The gap between the dotted line and the solid parallel line is denoted as b over $||w||$.

Return to Figure 4S.9

Figure 4S.10 Extended Description (Chapter 4S)

Return to Figure 4S.10

In the diagram, users named Casey are identified as similar users. Each user has initially watched two videos. Ultimately, a video watched by Pat is recommended to Casey.

Return to Figure 4S.10

Figure 4S.11 Extended Description (Chapter 4S)

Return to Figure 4S.11

This line graph displays retail sales for hobby, toy, and game stores in millions of dollars from 1994 to 2024. The horizontal axis represents time from 1994 to 2024, marked annually. The vertical axis represents millions of dollars, ranging from 0 to 4,000 in increments of 400. The graph shows a distinct and repeating seasonal pattern, with a very sharp spike in sales at the end of each year, likely corresponding to the holiday season. Following each peak, sales drop significantly at the start of the next year. Overall, there is a gradual upward trend in the baseline (non-peak) sales throughout the period. In 1994, the baseline sales were approximately 600 million dollars, and by 2024, they have risen to about 1,500 million dollars. The annual peaks also show variation. The highest peak on the chart occurs around 1999, reaching nearly 3,900 million dollars. In the 2000s and 2010s, the peaks generally fluctuate between 3,400 and 3,500 million dollars. The graph includes light vertical bars that indicate recessions. There is a short bar around 2001, a wider bar spanning from late 2007 to mid-2009, and a very narrow bar in early 2020. In early 2020, there is a dramatic and sharp drop in sales, falling from over 1,000 million dollars to a low point below 400 million dollars, before quickly recovering. After this event, the baseline sales level is noticeably higher than in previous years. However, the annual peaks from 2021 to 2023 are lower than the peaks from the late 1990s and 2000s, reaching approximately 2,800 million dollars.

Return to Figure 4S.11

Figure 4S.12 Extended Description (Chapter 4S)

Return to Figure 4S.12

The horizontal axis is labeled False positive rate and spans from 0.0 to 1.0. The vertical axis is labeled True positive rate and also spans from 0.0 to 1.0. In the top-left corner, at coordinates (0.0, 1.0), a single point is marked and identified as a Perfect classifier. A dashed diagonal line extends from the bottom-left corner (0.0, 0.0) to the top-right corner (1.0, 1.0), and this line is labeled Random classifier. Two arrows indicate the direction of performance: an arrow pointing towards the top-left signifies Better performance, while an arrow pointing towards the bottom-right signifies Worse performance. Three solid curves are plotted on the graph, all originating from (0.0, 0.0) and extending towards (1.0, 1.0). The uppermost curve, representing the best performance, rises steeply from the origin and then curves to approach the top-right corner, staying closest to the Perfect classifier point. A second curve is positioned below the first, indicating a moderate level of performance. The third curve is positioned below the second, lying closest to the Random classifier dashed line, which signifies the least effective performance among the three solid curves.

Return to Figure 4S.12

Figure 5.2 Extended Description (Chapter 5)

Return to Figure 5.2

In both graphs, the y-axis represents the closing price of Microsoft stock in USD, ranging from 50 to 350 at an interval of 50. The x-axis, labeled date, ranges from 2013 to 2022. The lines for training, validation, and prediction values are shaded in different colors. The graph at the top represents the close price from the LSTM model, while the graph at the bottom represents the close price from the GRU model. The lines in both graphs progress with multiple peaks and troughs, while having an overall increasing trend. The training data is visible from 2013 to mid-2020, where the price increases from below 50 dollars to above 150 dollars. The validation and prediction values are visible from mid-2020 to the remaining period, where both lines overlap. The peak value reaches around 350 dollars in 2022.

Return to Figure 5.2

Figure 5.3 Extended Description (Chapter 5)

Return to Figure 5.3

The y-axis represents the parameters, ranging from 10 to the power 2 to 10 to the power 7, while the x-axis represents the years. The parameters are denoted as follows. AlexNet(Breakthrough in image lassification) in 2012: 1 million; VGGNet(Deeper CNNs improved vision tasks) in 2014: 10 to the power 2; BERT(Bidirectional language understanding) in 2018: nearly 10 to the power 3; GPT-3(human-like text generation in scale) in 2020: 10 to the power 5; GPT-4(multimodal reasoning and fewer hallucination) in 2022: 10 to the power 6; Gemini 1.5(Integration of memory and RAG) in 2024: slightly above 10 to the power 6; GPT-40(Realtime multimodal AI) in 2026: around 10 to the power 7. Values are approximated.

Return to Figure 5.3

Figure 5S.1 Extended Description (Chapter 5S)

Return to Figure 5S.1

It begins with input data that is fed into a hidden layer composed of multiple interconnected nodes within a shaded box. These nodes process the inputs through weighted connections and biases, which are adjusted during training using stochastic gradient descent. The processed signals then pass to output layers, producing predictions. These predictions are compared to true values, and the difference is calculated using a loss function. The entire process, from input to weight adjustments, is repeated iteratively to improve the network's accuracy in making predictions. It also shows the flow of information during forward propagation and backward propagation.

Return to Figure 5S.1

Figure 6.1 Extended Description (Chapter 6)

Return to Figure 6.1

The image illustrates the difference between Predictive AI and Generative AI through two distinct flowcharts. The top section, labeled “Predictive AI,” shows “Input data” on the left, represented by six icons: a spreadsheet, a personal information document, a financial document with a magnifying glass, a line graph, a landscape image, and a medical record with a pencil. An arrow points from these inputs to a central “Pretrained model,” depicted as a large gear. From this model, an arrow leads to “Output” on the right, which includes three icons: a green checkmark and a red cross above “Classification,” a bar chart with a dollar sign above “Forecast,” and a magnifying glass examining a network of dots with an alert symbol, above “Anomaly alert.” The bottom section, labeled “Generative AI,” begins with “Input prompt” on the left, listing three text prompts: “Write a product description,” “Create an image of a blue bird,” and “Generate a video with an AI-Avatar using the attached script.” An arrow points from these prompts to a central “GEN AI” model, represented by a rectangular icon with rounded corners containing the text “GEN AI” surrounded by a network of interconnected nodes, and labeled “Model trained with massive data set.” An arrow then leads to “Output” on the right, displaying three generated items: a document with lines of text, an illustration of a blue bird, and a screen showing a video player with a male AI-Avatar speaking at a podium.

Return to Figure 6.1

Figure 6.2 Extended Description (Chapter 6)

Return to Figure 6.2

It depicts a massive dataset being fed into the open-domain AI tool, such as ChatGPT and Copilot. For closed-domain AI assistants (e.g., AI for legal, health care, and marketing), the massive dataset is processed through a pretrained GPT model and is fine-tuned on internal data.

Return to Figure 6.2

Figure 6.3 Extended Description (Chapter 6)

Return to Figure 6.3

The prompt asks to write a product description for a smartwatch. Under the tokenization process, the prompt is broken into multiple words, such as “write”, “a”, “product”, and “description”. Under the embedding phase, the words are converted to numerical vectors. After proceeding through the generative AI model, the response reads “Stay ahead of your day with the smartmate X2 smartwatch. Monitor your heart rate, track your...”.

Return to Figure 6.3

Figure 6.4 Extended Description (Chapter 6)

Return to Figure 6.4

On one side, human factors such as domain knowledge, strategic goals, ethical judgment, and prompt creation feed into the process, while on the other side, AI components like language and image models, pattern recognition, and content generation contribute their capabilities. Both elements converge into a central area labeled Creative Output, which encompasses a range of content types, including drafted articles, computer code, marketing images, website designs, and music or videos. Arrows indicate a bidirectional flow, highlighting how human insight and AI technology collaboratively generate diverse and innovative creative materials.

Return to Figure 6.4

Figure 6S.1 Extended Description (Chapter 6)

Return to Figure 6S.1

The prompt asks to write a product description for a smartwatch. Under the tokenization process, the prompt is broken into multiple words, such as “write”, “a”, “product”, “description”, “for”, “a”, and “smartwatch". Under the embeddings phase, the words are converted to numerical vectors. Under the autoregressive token phase, the words for the response, “stay”, “ahead”, “of”, “your”, “day”, “with”, “the”, and “SmartmateX2”, are arranged in sequence. The response reads “Stay ahead of your day with SmartMate X2.

Return to Figure 6S.1

Figure 6S.2 Extended Description (Chapter 6)

Return to Figure 6S.2

The diagram titled Denoising process. Below the title, a long, thick arrow points from left to right, indicating the direction of the process. Beneath the arrow is a horizontal sequence of eight square images that demonstrate this process. The first image on the far left, labeled Pure noise, is completely obscured by a dense pattern of multicolored specks, resembling static. Moving from left to right, each subsequent image becomes progressively clearer as the noise diminishes. Through these stages, an image of a dog with light-colored, long, wavy fur emerges. The dog is standing in a grassy field dotted with small flowers. In the middle images, the dog's form is visible but still heavily grainy. In the later images, the details of the dog's face, fur, and the surrounding field become sharp and distinct. The final image on the far right is a perfectly clear and sharp photograph of the dog in the field, with all the noise removed. This image is labeled Clean image.

Return to Figure 6S.2

Figure 6S.3 Extended Description (Chapter 6)

Return to Figure 6S.3

On the left, an oval labeled Random Noise provides input to a rectangle labeled Generator (Counterfeiter). The Generator creates a fake image, which is then sent as input to a rectangle labeled Discriminator (Judge). The Discriminator also receives input from a rectangle labeled Real Images. The Discriminator analyzes both the real and fake images and produces an output labeled Verdict; Real or Fake. This verdict is sent to a diamond-shaped decision box labeled Correct?. From the Correct? box, two paths, one labeled Yes and one labeled No, merge into a single feedback loop. This feedback loop is represented by a line labeled Correctness update (model weights adjusted) that points back to both the Generator and the Discriminator, indicating that both models are updated based on the discriminator's performance.

Return to Figure 6S.3

Figure 7.2 Extended Description (Chapter 7)

Return to Figure 7.2

The horizontal flow diagram composed of five interconnected, overlapping rectangular frames with a outline, creating a chain-like effect. Each frame contains a line-art icon and is associated with a specific category of metrics for evaluating AI. The labels for the categories are placed alternately above and below the chain of frames. From left to right, the elements are: Innovation and Creativity Metrics: This label is positioned above the first frame. The frame contains an icon of a gear with a lightbulb inside it. Operational Efficiency Metrics: This label is positioned below the second frame. The frame contains an icon of a dashboard or application window with a timer or stopwatch symbol inside. Customer Experience Metrics: This label is positioned above the third, central frame. The frame contains an icon of a person's head in profile, wearing a headset with a microphone, representing customer service. Business Outcomes Metrics: This label is positioned below the fourth frame. The frame contains an icon of a bar chart with an upward-trending line graph superimposed on it. Balanced Scorecard Approach: This label is positioned above the fifth and final frame on the right. The frame contains an icon showing three stars above three stylized icons of people.

Return to Figure 7.2

Figure 8.1 Extended Description (Chapter 8)

Return to Figure 8.1

The human with an objective provides a prompt to the generative AI, which receives inputs from the training data and past prompts to generate the output. This output is then evaluated by the human.

Return to Figure 8.1

Figure 8.2 Extended Description (Chapter 8)

Return to Figure 8.2

The flowchart begins at the top with a rounded rectangle containing the text Determine the goal of your AI session. An arrow points down to a second rounded rectangle with the text Write a good prompt. From this box, an arrow points down to a diamond-shaped decision box that asks, Is the response true?. If the answer is No, an arrow points to the left to a rounded rectangle labeled Refine your prompt. An arrow from this Refine your prompt box points back to the input of the Is the response true? decision box, creating a loop. If the answer to Is the response true? is Yes, an arrow points to the right and down to a second diamond-shaped decision box that asks, Does the response achieve the goal?. If the answer to this second question is No, an arrow points to the left, connecting to the Refine your prompt box, creating another loop. If the answer to Does the response achieve the goal? is Yes, an arrow points down to the word Done, which signifies the end of the process.

Return to Figure 8.2

Figure 8.3 Extended Description (Chapter 8)

Return to Figure 8.3

First, it involves defining the target audience by identifying the primary and secondary audiences, considering factors such as age, lifestyle, income level, geographical location, and existing commuting habits. The goal is to analyze market research data and current trends in personal transportation to accurately determine who the unicycle will appeal to the most. Second, the focus shifts to generating product messaging by developing key messaging points for the electric unicycle that resonate with the target audience. This includes highlighting unique selling propositions like portability, environmental benefits, cost-effectiveness, and convenience, along with creating a compelling narrative that addresses common pain points related to urban commuting, positioning the unicycle as a trendy, efficient, and eco-friendly solution. The third step involves selecting marketing channels by determining the most effective methods to reach the target audience. This entails evaluating options such as digital media platforms, online ads, influencer partnerships, and traditional media, while considering media consumption habits and cost-effectiveness to develop a cohesive marketing strategy. Finally, creating a launch timeline involves outlining a detailed schedule for product development milestones, marketing activities, teaser campaigns, public reveals, preorders, and full availability. It requires coordinating with channels, planning for stock availability, and syncing with significant events or seasons to optimize the product launch success.

Return to Figure 8.3

Glossary

Artificial intelligence (AI)
: The simulation of human intelligence in machines that can perform tasks such as learning, reasoning, problem-solving, and decision making.

Machine learning
: A subset of AI where algorithms learn from data to make predictions or decisions without being explicitly programmed.

Neural networks
: AI models inspired by the human brain, consisting of interconnected layers of artificial neurons for tasks like image and speech recognition.

Generative AI
: A type of AI that creates new content, such as text, images, or music, based on patterns learned from training data.

Virtual reality (VR)
: A computer-generated simulation that immerses users in a 3D environment, often enhanced with AI for realistic interactions.

Computer vision
: An AI technology that enables machines to interpret and analyze visual information from images and videos.

Data science
: An interdisciplinary field that combines statistics, machine learning, and domain expertise to extract insights from structured and unstructured data.

Data cleaning
: Identifying and correcting errors, inconsistencies, and inaccuracies in datasets to improve data quality.

Predictive modeling
: Using statistical and machine learning techniques to predict future outcomes based on historical data.

Personal data
: Information about an individual, such as name, location, or online behavior, often collected by AI-driven systems.

Informed consent
: The principle that individuals must be aware of and agree to the collecting and using their personal data.

AI divide
: The gap between individuals, organizations, and countries in access to and benefits from AI technologies.

Global divide
: Industrialized nations have more resources to develop and implement AI, while developing countries struggle with limited digital infrastructure and expertise.

Regional divide
: Urban areas often have greater access to AI-driven tools and education, whereas rural areas may lack the necessary infrastructure and opportunities.

Individual divide
: Socioeconomic disparities affect who can learn, use, and benefit from AI, with wealthier individuals and businesses having more opportunities to leverage AI advancements.

Filter bubble
: A situation where users are only exposed to information that aligns with their beliefs due to algorithmic personalization, limiting diverse perspectives.

Explainable AI (XAI)
: AI systems designed to explain their decision-making processes clearly, improving trust and accountability.

AI–augmented jobs
: Roles in which AI enhances human capabilities by automating routine tasks, improving decision making, and increasing productivity rather than replacing human workers.

AI ethics boards
: Committees within organizations or regulatory bodies responsible for overseeing the ethical development and deployment of AI systems, ensuring fairness, transparency, and accountability.

AI fluency
: The ability to understand, interpret, and effectively interact with AI technologies, including their capabilities, limitations, and ethical considerations.

AI governance
: The framework of policies, procedures, and oversight mechanisms that guide the ethical, transparent, and accountable use of AI technologies within an organization.

AI Winter
: A period of reduced funding, research activity, and interest in AI due to unmet expectations or technological limitations.

Artificial general intelligence (AGI)
: A hypothetical form of AI capable of performing any intellectual task a human can, exhibiting general reasoning and

learning abilities.

Bias audit
: A systematic evaluation of AI models to identify and mitigate biases in decision making, ensuring fairness and reducing discriminatory outcomes.

Big Data
: Huge datasets that require advanced processing techniques to analyze patterns, trends, and insights.

Black box problem
: The challenge of understanding and interpreting AI decision making where the reasoning behind predictions or outcomes is unknown.

Cloud computing
: The delivery of computing services, such as storage and processing power, over the Internet instead of local servers.

Data centers
: Large-scale facilities that house computer systems, storage, and networking equipment to support cloud computing, data processing, and AI applications.

Data mining
: The process of discovering patterns, correlations, and insights from large datasets using statistical and AI-driven techniques.

Data warehousing
: Storing and managing structured data in centralized repositories for analysis and decision making.

Deep learning
: An advanced form of machine learning using multilayered neural networks to learn complex patterns from large datasets automatically.

Edge computing
: A distributed computing approach where data processing occurs closer to the data source, reducing latency and bandwidth usage.

Expert system
: An AI program that uses rule-based logic in decision making to emulate human expertise in specific domains.

Fairness-aware machine learning
: A field of AI research focused on developing models that mitigate biases and promote equitable outcomes.

Internet of Things (IoT)
: A network of interconnected physical devices embedded with sensors and software that collect and exchange data.

Knowledge engineers
: Specialists who develop AI systems by encoding expert knowledge, rules, and reasoning into machine-readable formats.

Natural language processing (NLP)
: A branch of AI that enables machines to understand, interpret, and generate human language.

Predictive analytics
: Using AI and statistical techniques to forecast future trends and outcomes based on historical data.

Privacy–first policies
: Guidelines and frameworks designed to prioritize user data protection and minimize privacy risks when developing and deploying AI systems.

Reinforcement learning
: A machine learning approach where AI agents learn by receiving rewards or penalties for actions taken in an environment.

Robotic process automation (RPA)
: Software technology that automates repetitive, rule-based tasks by mimicking human actions within digital systems, such as data entry or invoice processing.

Rule-based reasoning
: A method of AI decision making that follows predefined rules and logical statements to arrive at conclusions, commonly used in expert systems.

Sentiment analysis
: An NLP technique used to determine the emotional tone or attitude expressed in text, such as customer reviews, survey responses, or social media posts. Businesses use sentiment analysis to gauge public perception, improve marketing strategies, and respond proactively to customer feedback.

Server farms
: Large groups of networked servers in one location that provide high computational power and storage capacity for cloud computing, Big Data analytics, and AI applications.

Symbolic reasoning
: An AI approach that uses symbols and logic rules to mimic human reasoning and problem solving.

Turing test
: A test proposed by Alan Turing to evaluate a machine’s ability to exhibit humanlike intelligence in conversation.

Universal Turing Machine
: A theoretical computational model proposed by Alan Turing that can simulate any other computing machine by reading instructions and executing them, forming the foundation of modern computing theory.

AI lifecycle management

Managing AI systems from development to decommissioning while ensuring sustainability and ethical considerations.

Bias audits
: Assessments conducted to detect and mitigate biases in AI systems to promote fairness and inclusivity.

Blockchain
: A secure digital ledger that records information in a way that is difficult to change, making it useful for tracking transactions and sharing data across multiple parties with trust and transparency.

California Privacy Rights Act (CPRA)
: A California law that strengthens consumer privacy rights by giving people more control over their personal data, including the right to know what data is collected, to correct it, to delete it, and to limit how businesses use or share it.

Data cleaning
: The structured process of transforming raw, inconsistent, or incomplete data into high-quality inputs for AI systems.

Data governance
: The framework of policies, processes, and standards that ensure an organization's data quality, security, privacy, and compliance.

Deduplication
: The removal of redundant or conflicting records to ensure data accuracy and prevent distortion in AI models.

Distributed cloud-based storage systems
: A method of storing data across multiple servers and locations in the cloud, allowing businesses to easily scale their storage needs, improve data access speeds, and reduce the risk of data loss due to server failures.

Edge storage
: A system that stores data locally on or near the device that collects it, such as a factory sensor or smart appliance, allowing information to be processed quickly without needing to send everything to a distant data center.

Encryption
: A security technique that scrambles data into a coded format, allowing only individuals with the correct password or key to read it, thereby protecting information while it’s being stored or shared.

European Union (EU) AI Act
: A European Union law that classifies AI systems by risk and sets rules to make sure high-risk AI is safe, transparent, and respects people’s rights, especially in areas like health care, hiring, or public safety.

Fairness metric
: Quantitative measure used to assess whether an AI system treats different groups equitably.

Federated learning
: A method of training AI models across many devices or servers without sharing raw data, which helps protect privacy by keeping personal data where it was created while still improving the model.

General Data Protection Regulation (GDPR)
: A European law that gives people control over their personal data by requiring businesses to be clear about how they collect, use, and protect it; violations can lead to large fines.

Identity and Access Management (IAM)
: A system of policies and tools that enables organizations to control who can access what data and systems, ensuring that only authorized personnel have the necessary level of access.

Least privilege
: A security principle that means users and systems should only have the minimum level of access necessary to do their job and nothing more to reduce the risk of misuse or data breaches.

Manipulative AI
: AI systems designed to deceive, exploit, or unduly influence individuals.

Right to Be Forgotten policy
: A policy that ensures the right of individuals to request the deletion of their data from databases.

Single source of truth approach
: A strategy where all departments and systems rely on one trusted and up-to-date version of data, helping to avoid confusion, duplication, and errors in decision making.

Social scoring
: AI systems that rate individuals based on their behaviors, activities, or characteristics.

Standardization (Z-score normalization)
: A scaling method that transforms data to have a mean of 0 and a standard deviation of 1.

Structured data
: Information that is organized in a fixed format—like spreadsheets or databases—making it easy for computers to search, sort, and analyze.

Unstructured data
: Information that doesn’t follow a consistent format, such as emails, videos, social media posts, or images; this type of data is harder to organize but can hold valuable insights when processed with AI.

Supervised learning
: A type of machine learning where the algorithm learns from labeled data to make predictions or classifications.

Unsupervised learning
: A machine learning approach that identifies hidden patterns or groupings in data without labeled outcomes.

Data preprocessing
: The process of cleaning, transforming, and organizing raw data to make it suitable for machine learning analysis.

Prediction and forecasting
: The use of machine learning to estimate future outcomes based on historical data, such as sales or demand.

Classification and risk assessment
: Applying machine learning to categorize data into discrete groups or predict the likelihood of certain outcomes, such as loan defaults or customer churn.

Data transformation
: Changing the format or structure of data to make it compatible with machine learning models, including encoding or rescaling.

Normalization
: A technique that scales data to a specific range, typically 0 to 1, to ensure consistency across features.

Dimensionality reduction
: Techniques that reduce the number of variables in a dataset while retaining important information.

Feature engineering
: Creating or modifying input variables to improve model performance.

K-means clustering
: An unsupervised algorithm that groups data points into a specified number of clusters based on similarity.

Algorithmic bias
: Systematic and unfair outcomes in machine learning caused by biased training data or model design.

Bias audits
: Assessments that test machine learning models for unfair outcomes across different demographic groups.

Collaborative filtering
: A recommendation method that predicts user preferences based on patterns from similar users.

Customer segmentation and personalization
: Grouping customers based on behavior or preferences to tailor marketing, product recommendations, or services.

Decision trees
: A model that uses branching rules to split data into subsets for prediction or classification.

Fairness-aware modeling
: Techniques that adjust models to produce more equitable outcomes for diverse groups.

Imputation method
: Techniques for filling in missing data using estimates based on other available information.

Linear regression
: A supervised learning method that models the relationship between one or more independent variables and a continuous outcome.

Logistic regression
: A supervised algorithm used to predict binary outcomes, such as yes/no or true/false decisions.

Machine learning
: A subset of AI where algorithms learn from data to make predictions or decisions without being explicitly programmed.

Principal component analysis (PCA)
: A dimensionality reduction method that transforms correlated variables into smaller uncorrelated components.

Random forests
: An ensemble learning method that combines multiple decision trees to improve accuracy and reduce overfitting.

Support vector machines (SVMs)
: A supervised learning algorithm that finds the best boundary to separate different categories of data.

Synthetic data
: Artificially generated data used to supplement or balance real datasets, especially when data is limited or imbalanced.

Time series analysis
: A technique for analyzing data points collected over time to identify trends and make future predictions.

Closed-domain chatbot
: An AI system trained on specialized, domain-specific data to handle questions or tasks within a narrowly defined area (e.g., legal, financial, or medical topics).

FOCUS framework
: A strategy for writing effective prompts that are Focused, Objective, include Context, are Unambiguous, and pose a Single query.

Hallucination (AI)
: A response generated by an AI model that sounds plausible but is factually incorrect, misleading, or entirely fabricated.

Hard prompt
: A complex or ambiguous prompt that challenges an AI model’s ability to reason, combine concepts, or generate reliable responses.

Iterative refinement
: The process of improving AI output by evaluating the initial response and then adjusting the prompt through multiple rounds.

Open-domain (open-ended) chatbot
: An AI system trained on broad, general-purpose data capable of answering a wide range of questions across many topics.

Persona (in prompting)
: A fictional role or identity assigned to the AI (e.g., lawyer, coach, customer service rep) to guide tone, content, or point of view in responses.

Prompt engineering
: The process of crafting and refining inputs to AI systems to get the most accurate, relevant, or creative output.

Retrieval-augmented generation (RAG)
: A method that enhances AI responses by combining generative models with real-time retrieval from external, trusted data sources, improving factual accuracy and reducing hallucinations.

Scaffolding (in prompting)
: Techniques used to guide AI performance by providing examples, constraints, or step-by-step structure to improve output quality and reliability.

Task decomposition
: Breaking down a complex task into smaller, simpler, and clearly defined subtasks, allowing each subtask to be separately addressed by an AI model. This approach leads to clearer, more precise, and effective responses.

Verification (AI responses)
: The practice of checking AI-generated content for accuracy, completeness, and credibility before using or sharing it.

Application Programming Interface (API)
: A software interface that enables applications to send requests to a foundation model (such as for summarization or translation) and receive generated outputs in response.

Autoregressive generation
: A sequential process whereby a model forecasts the most probable subsequent token and appends it to the output sequence, continuing this procedure until the response is fully generated.

Diffusion models
: A type of generative AI model that creates images by starting with random noise and gradually refining it to match a text prompt. Used in tools like DALL-E 2 and Stable Diffusion, these models “reverse” the process of adding noise to images during training.

Fine-tuning
: The process of adapting a pretrained model to a specific domain or task by continuing training on a smaller, specialized dataset. Fine-tuning helps the model reflect particular vocabulary, tone, or rules.

Foundation model
: Also known as a pretrained model; a large, general-purpose model such as ChatGPT 5, Claude, or Gemini, trained on extensive public data. These models can be utilized directly or adjusted for specific tasks.

Generative adversarial network (GAN)
: An image-generation method where two networks, called “generator” and “discriminator,” interact. The generator creates images, while the discriminator evaluates them, until realistic outputs are achieved.

Generative AI
: Creates original content—text, images, audio, video, or designs—based on patterns from large datasets; unlike predictive AI, which forecasts or classifies, generative AI produces new creations.

Hallucination
: A phenomenon where a generative model produces information that appears credible but is inaccurate or unsubstantiated, thereby presenting significant reliability concerns.

Image-based generative AI
: A type of generative AI that creates original images from text prompts or other inputs. These models are used in marketing, design, and art creation.

Mathematical vector
: A list of numbers that represents information in a way that AI models can understand and process. In generative AI, vectors are used to encode the meaning of words, images, or other inputs.

Multimodal generative AI
: Generative AI systems that can handle and combine different types of inputs or outputs—such as text, images, and audio—in a single model. These tools enable experiences like describing an image in words or generating a video from a script.

Multimodal model
: A system capable of processing multiple data types (e.g., text and image), which allows tasks such as describing a photo or generating an image from a text prompt.

Predictive AI
: Involves traditional models that classify, score, or forecast outcomes using structured data (e.g., fraud detection, risk scoring), rather than creating new outputs.

Retrieval-augmented generation (RAG)
: An AI technique that combines a language model with an external search component. Instead of relying only on what the model has memorized, RAG retrieves relevant documents or facts in real time to help generate more accurate and up-to-date responses.

Text-based generative AI

A type of generative AI that produces humanlike text based on input prompts. It can be used for writing emails, reports, product descriptions, or chatbot responses.

Tokenization
: Involves dividing input text into smaller units, such as words or characters, which can then be processed by the model effectively.

Traditional AI
: AI systems designed to perform narrowly defined tasks such as classification, prediction, or optimization using structured data and rule-based or statistical methods. Often referred to as predictive or task-specific AI.

Adaptive systems
: AI-driven systems that adjust their behavior or responses based on real-time user interactions or environmental changes.

Energy-efficient AI (green AI)
: Efforts to design and deploy AI models that require less computational power and energy, reducing environmental impact while maintaining performance.

Few-shot learning
: A machine learning method where models are trained to perform tasks based on only a small number of labeled examples.

Input layer
: The first layer of a neural network that receives raw data—such as numbers, images, or text—and passes it to the next layer for processing.

Multimodal learning
: A method where AI systems can simultaneously process and relate information from multiple data types, such as text, images, and audio.

Neural network
: Machine learning model inspired by the structure of the human brain, designed to recognize patterns and relationships in data.

Output layer
: The final layer of a neural network that produces the model's result—such as a classification, prediction, or decision—based on the data processed by earlier layers.

Self-supervised learning
: A machine learning approach where models generate their own labeled data from raw inputs, reducing the need for manual labeling.

Unsupervised learning
: A machine learning technique where models identify patterns or groupings in data without using labeled examples.

Weights and biases
: Internal values in a neural network that determine how input signals are combined and adjusted as they move through the network. Weights control the strength of each connection, while biases allow the network to shift or fine-tune its outputs.

Zero-shot learning
: A deep learning technique where a model makes accurate predictions about categories it was never explicitly trained on by using patterns and relationships learned from related data.

AI agents
: Tools that can carry out multistep tasks with minimal human input.

AI governance
: The framework of policies, procedures, and oversight mechanisms that guide the ethical, transparent, and accountable use of AI technologies within an organization.

AI readiness index
: A composite metric that assesses how prepared an organization is to adopt and scale AI, including technical infrastructure, skills, data, and leadership support.

Balanced scorecard
: A strategic performance management tool that measures an organization's success across multiple perspectives, including financial, customer, internal processes, and learning/growth metrics.

Closed-domain AI
: An AI system trained on a specific, limited set of data relevant to a narrow topic or task, which allows for higher accuracy and reliability within that scope (e.g., customer service for a single product line).

Cultural readiness
: An organization's openness to change, innovation, and cross-functional collaboration—especially its willingness to experiment, learn from failure, and adopt new technologies like AI.

Data silo
: A repository of data that is isolated from other parts of the organization, typically because it is stored in a separate system or controlled by a specific department.

Dynamic pricing
: Strategy where prices are adjusted in real time based on market demand, competition, customer behavior, and other factors.

Edge-case testing
: The process of testing AI systems on rare, unexpected, or extreme input scenarios to ensure robust and reliable

performance.

Hallucinations
: False or misleading outputs generated by AI systems, particularly large language models, that may sound plausible but are factually incorrect or fabricated.

Human-in-the-loop (HITL)
: A system design where human review or intervention is included at critical stages of the AI decision-making process, especially in high-risk applications.

Innovation strategy
: An organizational plan integrating new technologies like AI into broader goals like product development, customer experience, or business model transformation. In AI implementation, an innovation strategy helps align technical initiatives with long-term value creation and fosters a culture of experimentation and continuous improvement.

Intelligent process automation (IPA)
: The combination of robotic process automation (RPA) and AI to automate workflows and processes, including cognitive tasks like decision making.

MLOps
: Short for “machine learning operations,” it refers to the practices and tools used to deploy, manage, and monitor machine learning models in production.

Pilot trap
: A common challenge in AI implementation, where projects remain stuck in pilot phases and fail to scale into full production.

Retrieval-augmented generation (RAG)
: A method that enhances AI responses by combining generative models with real-time retrieval from external, trusted data sources, improving factual accuracy and reducing hallucinations.

Robotic process automation (RPA)
: Software technology that automates repetitive, rule-based tasks by mimicking human actions within digital systems, such as data entry or invoice processing.

Sentiment analysis
: An NLP technique used to determine the emotional tone or attitude expressed in text, such as customer reviews, survey responses, or social media posts. Businesses use sentiment analysis to gauge public perception, improve marketing strategies, and respond proactively to customer feedback.

Skills gap analysis
: A systematic process of identifying the difference between the skills employees currently possess and the skills required to meet an organization’s goals. This analysis helps determine areas where additional training, reskilling, or hiring is needed.

Misinformation
: False or misleading information that spreads online, often amplified by AI-powered recommendation algorithms.

Privacy
: The right of individuals to control how their data is collected, used, and shared.

Deepfake
: AI-generated or manipulated media, such as videos or images, designed to impersonate real individuals or events convincingly.

Chatbot
: An AI-powered program that interacts with users via text or speech, commonly used in customer service and virtual assistants.

Intellectual property (IP)
: Legal rights that protect creations of the mind, such as patents, trademarks, and copyrights, which AI-generated content raises questions about.

Speech recognition
: AI’s ability to convert spoken language into text, enabling voice assistants and automated transcription services.

Visualization
: The graphical representation of data or information to make patterns, trends, and insights easier to understand. Common forms include charts, graphs, maps, and dashboards.

Hidden layers
: In a neural network, a layer of neurons located between the input layer and the output layer. Hidden layers transform the incoming data by applying weights, biases, and activation functions, enabling the network to detect patterns and relationships that are not directly observable in the raw input.

Deep learning
: A subset of machine learning that uses multi-layered neural networks to process complex data and make predictions.

Robotic process automation (RPA)
: The use of robots or AI-driven automation tools to perform repetitive business processes, such as data entry or invoice processing, without human intervention.

Edge computing
: A method of processing data closer to where it is generated (such as a sensor or device) rather than sending it to a central server, which helps reduce delays and enables faster, more efficient real-time decision making.

Imputation
: A method for filling in missing data using statistical estimates or predictive modeling to maintain dataset integrity.

Role-based access control
: A method for managing access to resources by assigning permissions based on job roles ensuring that individuals only have access to the data and tools they need for their specific responsibilities.

Process optimization and automation
: Using machine learning to streamline operations, reduce costs, and improve efficiency through automated decision-making.

Reinforcement learning
: A learning method where an algorithm improves performance by interacting with an environment and receiving feedback in the form of rewards or penalties.

Foundation models
: Large-scale, pretrained AI models that can be adapted for a wide range of downstream tasks with minimal additional training.

Pretraining
: The initial phase of training a generative AI model using large datasets from diverse sources to learn general language or image patterns. This step creates a foundation model that can be adapted for many tasks.

Custom generative AI model
: A version of a generative AI model that has been adapted to meet the specific needs of a business or domain. Most custom models are created by fine-tuning a large, pretrained model on proprietary data to reflect a company's voice, terminology, or tasks.

Retrieval-augmented generation (RAG)
: An AI technique that combines a language model with an external search component. Instead of relying only on what the model has memorized, RAG retrieves relevant documents or facts in real time to help generate more accurate and up-to-date responses.

Remarks

Sample annotation #1.

inserting an annotation

inseting annotakjslfkjLKD

annotation sample

[1] https://news.airbnb.com/anti-party-crackdown-over-new-years-eve/ Back to reference: 1

[2] https://news.airbnb.com/anti-party-crackdown-over-new-years-eve/ Back to reference: 2

[3]*Mohmad, Parvin. "Top 10 Business Applications of Neural Networks in 2023." Analytics Insight, 24 May 2023*. https://www.analyticsinsight.net/business-intelligence-and-analytics/top-10-business-applications-of-neural-networks-in-2023; *Gupta, Jatinder N. D., and Kate A. Smith. Neural Networks in Business: Techniques and Applications. Idea Group Inc (IGI), 2003.* Back to reference: 3

[4] *"Microsoft Lens gets a hot new feature — the Office app can 'read' your handwriting." Laptop Mag, 2024. https://www.laptopmag.com/news/microsoft-lens-gets-a-hot-new-feature-the-office-app-can-read-your-handwriting*. Back to reference: 4

[5] https://news.samsung.com/global/samsung-announces-global-launch-of-samsung-food-an-ai-powered-personalized-food-and-recipe-service Back to reference: 5

[7] https://bluerivertechnology.com/technology/ Back to reference: 7

[8] https://hbr.org/2018/02/7-years-later-what-really-happened-to-siri-and-apple; https://www.fastcompany.com/3065339/siri-rising-inside-apples-voice-controlled-brain-trust Back to reference: 8

[9] https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent.html Back to reference: 9

[11] https://www.bcs.org/media/10764/data-ownership-and-privacy.pdf Back to reference: 11

[12] International Energy Agency (IEA). *Data Centres and Data Transmission Networks – Tracking Report*. 2021. https://www.iea.org/reports/data-centres-and-data-transmission-networks Back to reference: 12

[13] Google. *Environmental Report 2024*. Google Sustainability. 2024. https://www.gstatic.com/gumdrop/sustainability/google-2024-environmental-report.pdf Back to reference: 13

[14] Forti, V., Baldé, C. P., Kuehr, R., and Bel, G. *The Global E-Waste Monitor 2020: Quantities, Flows, and the Circular Economy Potential*. United Nations University, International Telecommunication Union, and International Solid Waste Association. 2020. https://collections.unu.edu/eserv/UNU:7737/GEM_2020_def_july1.pdf Back to reference: 14

[15] Stanford AI Index Report (2022): https://aiindex.stanford.edu/report/ Back to reference: 15

[16] International Telecommunication Union, 2024: https://www.itu.int/itu-d/reports/statistics/facts-figures-2024/index/ Back to reference: 16

[19] Pew Research Center, 2021: https://www.pewresearch.org/ Back to reference: 19

[20] https://about.google/intl/ALL_us/values-in-action/internet-saathi/ Back to reference: 20

[27] Panettieri, Joe. "Generative AI Lawsuits Timeline: Legal Cases vs. OpenAI, Microsoft, Anthropic, Nvidia, Perplexity, Intel and More." *Sustainable Tech Partner*, January 9, 2025. https://sustainabletechpartner.com/topics/ai/generative-ai-lawsuit-timeline/

[28] Smiley, Lauren. “The Legal Saga of Uber’s Fatal Self-Driving Car Crash Is Over.” *WIRED*, July 28, 2023. https://www.wired.com/story/ubers-fatal-self-driving-car-crash-saga-over-operator-avoids-prison/

[1] Jones, Milo, and Philippe Silberzahn. "Without An Opinion, You're Just Another Person With Data." *Forbes*, March 15, 2016. https://www.forbes.com/sites/milojones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/

[2] Alder, Steve. “UHG Increases Change Healthcare Data Breach Victim Count to 190 Million.” *HIPAA Journal*, January 25, 2025. https://www.hipaajournal.com/change-healthcare-responding-to-cyberattack/ Back to reference: 2

[3] Amazon Web Services. "IAM Roles." *AWS Identity and Access Management User Guide*. Last modified January 15, 2025. https://docs.aws.amazon.com/IAM/latest/UserGuide/id_roles.html Back to reference: 3

[4] Fink, Camille. “Hurricane Ida Response and Recovery in Louisiana.” *Health Facilities Management*, September 17, 2021. https://www.hfmmagazine.com/articles/4284-hurricane-ida-response-and-recovery-in-louisiana Back to reference: 4

[5] Office of Cybersecurity, Energy Security, and Emergency Response. “Colonial Pipeline Cyber Incident.” *U.S. Department of Energy*. https://www.energy.gov/ceser/colonial-pipeline-cyber-incident Back to reference: 5

[6] National Institute of Standards and Technology. “The NIST Cybersecurity Framework (CSF) 2.0.” Last modified February 26, 2024. https://doi.org/10.6028/NIST.CSWP.29 Back to reference: 6

[7] Lohr, Steve. “For Big-Data Scientists, ‘Janitor Work’ Is Key Hurdle to Insights.” *The New York Times*, August 17, 2014. https://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-janitor-work-is-key-hurdle-to-insights.html Back to reference: 7

[8] Editverse. *Data Cleaning Techniques: Ensuring uality in your 2024–2025 Research*, August 2024. https://editverse.com/data-cleaning-techniques-ensuring-quality-in-your-2024-2025-research/ Back to reference: 8

[9] Thota, Prahalad. “Inside Wells Fargo’s Data Management System.” *ET CIO*, October 29, 2019. https://cio.economictimes.indiatimes.com/news/strategy-and-management/inside-wells-fargos-data-management-system/71799797 Back to reference: 9

[10] https://gdpr.eu/fines/_; Wolford, Ben. “What are the GDPR Fines?” *GDPR.eu*; Data Privacy Manager. “20 Biggest GDPR Fines So Far [2025].” *Data Privacy Manager*, February 12, 2024. https://dataprivacymanager.net/5-biggest-gdpr-fines-so-far-2020/ Back to reference: 10

[11] California Privacy Protection Agency. (2024, January 2). *California Consumer Privacy Act Regulations*, January 2, 2024. https://cppa.ca.gov/regulations/pdf/cppa_regs.pdf Back to reference: 11

[12] California Privacy Protection Agency. *California Consumer Privacy Act Regulations*, January 2, 2024. https://cppa.ca.gov/regulations/pdf/cppa_regs.pdf Back to reference: 12

[13] California Privacy Protection Agency. *California Consumer Privacy Act Regulations*, January 2, 2024. https://cppa.ca.gov/regulations/pdf/cppa_regs.pdf Back to reference: 13

[14] California Department of Justice. “Attorney General Bonta Announces Settlement with Sephora as Part of Ongoing Enforcement of California Consumer Privacy Act.” *State of California Department of Justice*, August 24, 2022. https://oag.ca.gov/news/press-releases/attorney-general-bonta-announces-settlement-sephora-part-ongoing-enforcement Back to reference: 14

[15] European Parliament. “EU AI Act: First Regulation on Artificial Intelligence.” *European Parliament*, June 8, 2023. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence Back to reference: 15

[16] Californians for Consumer Privacy. “Text of the California Privacy Rights Act (CPRA) of 2020.” https://www.caprivacy.org/cpra-text/ Back to reference: 16

[17] NASA. *Advanced Computing: Powering NASA's Science and Engineering Discoveries*, March 2021. https://www.nas.nasa.gov/SC21/assets/pdf/NASA_AdvancedComputing_2021-03.pdf Back to reference: 17

[18] Google. “Positive Energy: Belgian Site Becomes First Google Data Center to Add On-Site Solar.” *Google Sustainability*,

September 2018. https://sustainability.google/operating-sustainably/stories/belgium-solar/; St. John, Jeff. “Google Plans to Build Gigawatts of Clean Power and Data Centers Together.” *Canary Media*, December 10, 2024. https://www.canarymedia.com/articles/clean-energy/google-has-a-20b-plan-to-build-data-centers-and-clean-power-together Back to reference: 18

[19] Green Mountain. “Sustainability.” https://greenmountain.no/why-green-mountain/renewable-power/ Back to reference: 19

[20] Ungleich Glarus AG. “Data Center Light: 99.9% Hydropower.” https://datacenterlight.ch/en-us/cms/hydropower/ Back to reference: 20

[21] Terrell, Michael. “A First-of-Its-Kind Geothermal Project Is Now Operational.” *The Keyword*, November 28, 2023. https://blog.google/outreach-initiatives/sustainability/google-fervo-geothermal-energy-partnership/ Back to reference: 21

[22] Meta. “New Geothermal Energy Project to Support Our Data Centers.” *Meta Newsroom*, August 26, 2024. https://about.fb.com/news/2024/08/new-geothermal-energy-project-to-support-our-data-centers/ Back to reference: 22

[23] “Big Tech Pours $1.5bn into Nuclear Reactors to Power AI Boom.” *Financial Times*, February 19, 2025. https://www.ft.com/content/2d84198e-7eeb-4154-bbf2-9a469b0cc700; “Nuclear-Powered Data Centers: When Will SMRs Finally Take Off?” *Data Center Knowledge*, January 22, 2025. https://www.datacenterknowledge.com/energy-power-supply/nuclear-powered-data-centers-when-will-smrs-finally-take-off- Back to reference: 23

[24] European Parliament and Council. “Directive (EU) 2022/2464 of the European Parliament and of the Council of 14 December 2022 Amending Regulation (EU) No 537/2014, Directive 2004/109/EC, Directive 2006/43/EC and Directive 2013/34/EU, as Regards Corporate Sustainability Reporting.” *Official Journal of the European Union*, December 16, 2022. https://eur-lex.europa.eu/eli/dir/2022/2464/oj/eng Back to reference: 24

[25] Climate Neutral Data Centre Pact. *Climate Neutral Data Centre Pact*. (n.d.). https://www.climateneutraldatacentre.net/ Back to reference: 25

[26] Allison, Ian. “Guardtime Secures Over a Million Estonian Healthcare Records on the Blockchain.” *International Business Times*, March 3, 2016. https://www.ibtimes.co.uk/guardtime-secures-over-million-estonian-healthcare-records-blockchain-1547367 Back to reference: 26

[27] de Best, Raynor. “Transaction Speed Ranking of 45 Crypto—Including DeFi and Metaverse—in 2025.” *Statista*, February 5, 2025. https://www.statista.com/statistics/944355/cryptocurrency-transaction-speed/ Back to reference: 27

[28] Arrow Electronics. “Sensors and Edge AI: How They Work Together to Provide Autonomous Driving.” Last modified May 17, 2022. https://www.arrow.com/en/research-and-events/articles/sensors-and-edge-ai-how-they-work-together-to-provide-autonomous-driving Back to reference: 28

[29] Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H. R., Albarqouni, S., Bakas, S., et al. (2020). “The Future of Digital Health with Federated Learning.” *npj Digital Medicine* 3(1), 19. Back to reference: 29

[30] Daniel, R. M., & Shevina, V. (2024). “Relationship between Ethical Business and Customer Retention at Tech Industry.” *International Research Journal of Modernization in Engineering Technology and Science*, 6(10), 1501–1505. Back to reference: 30

[1] Turing, A. M. (1950). “Computing Machinery and Intelligence.” *Mind*, 59(236), 433–460. Back to reference: 1

[2] Cheng, X., Kosinski, M., & Stillwell, D. (2024). A Turing test of whether AI chatbots are behaviorally similar to humans. *Proceedings of the National Academy of Sciences*, 121(9), e2313925121. https://doi.org/10.1073/pnas.2313925121 Back to reference: 2

[3] McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.” Back to reference: 3

[4] Newell, A., & Simon, H. A. (1956). “The Logic Theory Machine—A Complex Information Processing System.” *IRE Transactions on Information Theory*, 2(3), 61–79. Back to reference: 4

[5] Weizenbaum, J. (1966). “ELIZA—A Computer Program for the Study of Natural Language Communication between Man and Machine.” *Communications of the ACM*, 9(1), 36–45. Back to reference: 5

[6] Shortliffe, E. H. (1976). *Computer-Based Medical Consultations: MYCIN*. Elsevier. Back to reference: 6

[7] Khandelwal, K., Patel, S., Patel, J., & Pnachal, M. (2023). "A Study to Know —Use of AI for Personalized Recommendation, Streaming Optimization, and Original Content Production at Netflix." *International Journal of Scientific Research in Engineering and Technology* 9(6): 435–445. https://ijsret.com/wp-content/uploads/2023/11/IJSRET_V9_issue6_435.pdf Back to reference: 7

[8] H2O.ai. (2020). "Driving Away Fraudsters at PayPal." https://h2o.ai/content/dam/h2o/en/marketing/documents/2020/01/PayPal-Customer-Case-Study-rnd2-1.pdf Back to reference: 8

[9] Jiang, Fei, Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., & Wang, Y. (2017). "Artificial Intelligence in Healthcare: Past, Present and Future." *Stroke and Vascular Neurology* 2(4): 230–243; "Artificial Intelligence and Healthcare: A Journey through History." *Journal of Medical Internet Research* 25(1) (2023): e11122160. Back to reference: 9

[10] Strickland, Eliza. "How IBM Watson Overpromised and Underdelivered on AI Health Care." *IEEE Spectrum*, April 2, 2019; "Confronting the Criticisms Facing Watson for Oncology." *The ASCO Post*, September 10, 2019. Back to reference: 10

[11] DeepSeek-AI. "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning." *GitHub*. https://github.com/deepseek-ai/DeepSeek-R1; Tau, Byron. "Researchers Link DeepSeek's Blockbuster Chatbot to Chinese Telecom Banned from Doing Business in US." *AP News*, February 5, 2025. https://apnews.com/article/deepseek-china-generative-ai-internet-security-concerns-c52562f8c4760a81c4f76bc5fbdebad0 Back to reference: 11

[12] McKinsey & Company. "The State of AI in Early 2024." June 2024. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai. Back to reference: 12

[13] International Telecommunication Union. *ITU Internet Reports 2005: The Internet of Things*. 7th ed. Geneva: International Telecommunication Union, 2005. Back to reference: 13

[14] Greenberg, Sarah. "Birth and Evolution of Internet of Things (IoT)." *BCC Research*, July 30, 2021. https://blog.bccresearch.com/birth-and-evolution-of-internet-of-things-iot Back to reference: 14

[15] Matellio Inc. "IoT for Smart Cities: Enabling Smarter & Connected Solutions." January 7, 2025. https://www.matellio.com/blog/iot-for-smart-cities/ Back to reference: 15

[16] Quy, T. E., Roy, A., Iosifidis, V., Zhang, W., & Ntoutsi, E. "A Survey on Datasets for Fairness-Aware Machine Learning." January 21, 2022. https://arxiv.org/abs/2110.00530 Back to reference: 16

[17] SciDev.Net. "AI Risks in Healthcare: Misdiagnosis, Inequality, and Ethical Concerns." *News-Medical.Net*, January 23, 2024. https://www.news-medical.net/news/20240123/AI-risks-in-healthcare-Misdiagnosis-inequality-and-ethical-concerns.aspx Back to reference: 17

[18] Federal Aviation Administration. *Summary of the FAA's Review of the Boeing 737 MAX*. November 18, 2020. https://www.faa.gov/sites/faa.gov/files/2022-08/737_RTS_Summary.pdf Back to reference: 18

[19] Columbus, Louis. "Defending SOCs Under Siege: Battling Adversarial AI Attacks." *VentureBeat*, December 9, 2024. https://venturebeat.com/security/defending-socs-under-siege-battling-adversarial-ai-attacks/. Back to reference: 19

[20] Brumfiel, Geoff. "How AI is Revolutionizing How Governments Conduct Surveillance." *NPR*, June 13, 2023. https://www.npr.org/2023/06/13/1181868277/how-ai-is-revolutionizing-how-governments-conduct-surveillance Back to reference: 20

[21] TIME. "How OpenAI's Sam Altman Is Thinking About AGI and Superintelligence in 2025." January 8, 2025. https://time.com/7205596/sam-altman-superintelligence-agi/ Back to reference: 21

[4] Raman, Aditya. "Explaining Spotify Song Classification and Principal Components Analysis using Python." *Medium*, August 26, 2022. https://medium.com/@adityaraman/explaining-spotify-song-classification-and-principal-components-analysis-using-python

[5] Torlay, Lucas. "Creating a Zillow Zestimate Using Machine Learning." *Medium*, May 26, 2022. https://medium.com/@ltorlay/creating-a-zillow-zestimate-using-machine-learning-a4785b9e541

[6] Xue, D., Frisch, A., He, D. (2019). Differential Diagnosis of Heart Disease in Emergency Departments Using Decision Tree and Medical Knowledge. In: Gadepally, V., *et al*. Heterogeneous Data Management, Polystores, and Analytics for Healthcare. DMAH Poly 2019 2019. Lecture Notes in Computer Science(), vol 11721. Springer, Cham. https://doi.org/10.1007/978-3-030-33752-0_16; S. Palaniappan and R. Awang, "Intelligent heart disease prediction System using data mining techniques", *IEEE/ACS International Conference on Computer Systems and Applications AICCSA 2008*, pp. 108-115, March 31 2008-Apri14 2008.

[7] ActiveWizards Group LLC. "5 Real-world Examples of Logistic Regression Application." *ActiveWizards Blog*. https://activewizards.com/blog/5-real-world-examples-of-logistic-regression-application; Lynn, Jacob. "Predicting Cancellations with Survival Modeling." *Booking.com Data Science Blog*. November 11, 2024. https://booking.ai/predicting-cancellations-with-survival-modeling-a299af54249b

[8] Pinem, Fransiska Jesinta, Rachmadita Andreswari, and Muhammad Azani Hasibuan. "Sentiment Analysis to Measure Celebrity Endorsement's Effect using Support Vector Machine Algorithm." In *Proceedings of the 2018 International Conference on Electrical Engineering and Computer Science and Informatics* (EECSI), 690-695. Malang, Indonesia: IEEE, 2018. https://core.ac.uk/download/pdf/296976312.pdf

[9] Qi, Y. (2019). Random Forest for Bioinformatics. In *Machine Learning Techniques for Bioinformatics* (pp. 1-18). Springer. https://www.cs.cmu.edu/~qyj/papersA08/11-rfbook.pdf

[10] Mallouhy, Roxane Elias, Christophe Guyeux, Chady Abou Jaoude, and Abdallah Makhoul. "K-mean Clustering: A Case Study in Yvelines, Île-de-France." In *2022 14th International Conference on Computational Intelligence and Communication Networks (CICN)*, 1-6. Al-Khobar, Saudi Arabia: IEEE, 2022. https://doi.org/10.1109/CICN56167.2022.10008344.

[11] AI Tech Daily. "AI in Automotive Manufacturing: Driving the Future of Vehicle Production." *AI Tech Daily*, August 5, 2024. Accessed May 2, 2025. https://medium.com/ai-tech-daily/ai-in-automotive-manufacturing-driving-the-future-of-vehicle-production.

[12] Chopra, Rohit, Kristen Clarke, Charlotte A. Burrows, and Lina M. Khan. "Joint Statement on Enforcement Efforts Against Discrimination and Bias in Automated Systems." *Consumer Financial Protection Bureau*, May 2, 2025. Accessed May 2, 2025. https://www.consumerfinance.gov/about-us/newsroom/joint-statement-on-enforcement-efforts-against-discrimination-and-bias-in-automated-systems/

[13] Groves, Lara, Jacob Metcalf, Alayna Kennedy, Briana Vecchione, and Andrew Strait. "Auditing Work: Exploring the New York City Algorithmic Bias Audit Regime." arXiv, February 2024. https://arxiv.org/html/2402.08101v1

[14] Krug, Steve. Don't Make Me Think: A Common Sense Approach to Web Usability. 2nd ed. Berkeley: New Riders, 2005.

[1] Schmidt, Douglas C., et al. *Evaluating Personified Expert Effectiveness in AI-Powered Systems.* Vanderbilt University, 2023. https://www.dre.vanderbilt.edu/~schmidt/PDF/Evaluating_Personified_Expert_Effectiveness_Conference.pdf

[2] Incubity. "10 Examples of Persona Patterns in Prompt Engineering." *Incubity*, 2023. https://incubity.ambilio.com/10-examples-of-persona-patterns-in-prompt-engineering/; Skim AI. "Personas in Prompt Engineering." *Skim AI*, 2023. https://skimai.com/personas-in-prompt-engineering/; Wan, Yixin, Jieyu Zhao, Aman Chadha, Nanyun Peng, and Kai-Wei Chang. "Are Personalized Stochastic Parrots More Dangerous? Evaluating Persona Biases in Dialogue Systems." *arXiv*. Last modified November 2, 2023. https://arxiv.org/abs/2310.05280

[3] Neff, Gina, and Peter Nagy. "Talking to Bots: Symbiotic Agency and the Case of Tay." *International Journal of Communication*, vol. 10, 2016, pp. 4915–4931 https://ijoc.org/index.php/ijoc/article/view/6277/1804

[1] Schmidt, Douglas C., et al. *Evaluating Personified Expert Effectiveness in AI-Powered Systems*. Vanderbilt University, 2023. https://www.dre.vanderbilt.edu/~schmidt/PDF/Evaluating_Personified_Expert_Effectiveness_Conference.pdf Back to reference: 1

[2] Incubity. "10 Examples of Persona Patterns in Prompt Engineering." *Incubity*, 2023. https://incubity.ambilio.com/10-examples-of-persona-patterns-in-prompt-engineering/; Skim AI. "Personas in Prompt Engineering." 2023. https://skimai.com/personas-in-prompt-engineering/; Wan, Y., Zhao, J., Chadha, A., Peng, N., & Chang, K.-W. "Are Personalized Stochastic Parrots More Dangerous? Evaluating Persona Biases in Dialogue Systems." *arXiv*. Last modified November 2, 2023. https://arxiv.org/abs/2310.05280 Back to reference: 2

[3] Neff, Gina, & Nagy, P. (2016). "Talking to Bots: Symbiotic Agency and the Case of Tay." *International Journal of*

Communication, 10, 4915–4931. https://ijoc.org/index.php/ijoc/article/view/6277/1804 Back to reference: 3

[1] Duhigg, Charles. “How Companies Learn Your Secrets.” *New York Times Magazine*, February 16, 2012. https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html. Back to reference: 1

[2] Hays, Constance L. “What Wal-Mart Knows About Customers’ Habits.” *The New York Times*, November 14, 2004. https://www.nytimes.com/2004/11/14/business/yourmoney/what-walmart-knows-about-customers-habits.html Back to reference: 2

[3] Mnih, V., Kavukcuoglu, K., Silver D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., et al. (2015). “Human-Level Control through Deep Reinforcement Learning.” *Nature* 518(7540), 529–533. https://doi.org/10.1038/nature14236 Back to reference: 3

[4] Raman, Aditya. “Explaining Spotify Song Classification and Principal Components Analysis using Python.” *Medium*, August 26, 2022. https://medium.com/@adityaraman/explaining-spotify-song-classification-and-principal-components-analysis-using-python
Back to reference: 4

[5] Torlay, Lucas. “Creating a Zillow Zestimate Using Machine Learning.” *Medium*, May 26, 2022. https://medium.com/@ltorlay/creating-a-zillow-zestimate-using-machine-learning-a4785b9e541 Back to reference: 5

[6] Xue, D., Frisch, A., and He, D. (2019). Differential Diagnosis of Heart Disease in Emergency Departments Using Decision Tree and Medical Knowledge. In: Gadepally, V., et al. *Heterogeneous Data Management, Polystores, and Analytics for Healthcare*. DMAH Poly 2019. Lecture Notes in Computer Science, vol 11721. Springer, Cham. https://doi.org/10.1007/978-3-030-33752-0_16; Palaniappan, S., & Awang, R., “Intelligent Heart Disease Prediction System Using Data Mining Techniques.” *IEEE/ACS International Conference on Computer Systems and Applications AICCSA 2008*, pp. 108– 115, March 31, 2008–April 14, 2008. Back to reference: 6

[7] ActiveWizards Group LLC. “5 Real-World Examples of Logistic Regression Application.” *ActiveWizards Blog*. https://activewizards.com/blog/5-real-world-examples-of-logistic-regression-application; Lynn, Jacob. “Predicting Cancellations with Survival Modeling.” *Booking.com Data Science Blog*. November 11, 2024. https://booking.ai/predicting-cancellations-with-survival-modeling-a299af54249b Back to reference: 7

[8] Pinem, F. J., Andreswari, R., & Hasibuan, M. A.. “Sentiment Analysis to Measure Celebrity Endorsement’s Effect using Support Vector Machine Algorithm.” In *Proceedings of the 2018 International Conference on Electrical Engineering and Computer Science and Informatics (EECSI)*, pp. 690–695. Malang, Indonesia: IEEE, 2018. https://core.ac.uk/download/pdf/296976312.pdf Back to reference: 8

[9] Qi, Y. (2019). “Random Forest for Bioinformatics.” In *Machine Learning Techniques for Bioinformatics* (pp. 1–18). Springer. https://www.cs.cmu.edu/~qyj/papersA08/11-rfbook.pdf Back to reference: 9

[10] Mallouhy, R. E., Guyeux, C., Jaoude, C. A., & Makhoul, A. “K-Mean Clustering: A Case Study in Yvelines, Île-de-France.” In *2022 14th International Conference on Computational Intelligence and Communication Networks (CICN)*, pp. 1–6. Al-Khobar, Saudi Arabia: IEEE, 2022. https://doi.org/10.1109/CICN56167.2022.10008344. Back to reference: 10

[11] *AI Tech Daily*. “AI in Automotive Manufacturing: Driving the Future of Vehicle Production.” August 5, 2024. https://medium.com/ai-tech-daily/ai-in-automotive-manufacturing-driving-the-future-of-vehicle-production Back to reference: 11

[12] Chopra, R., Clarke, K., Burrows, C. A., & Khan, L. M. “Joint Statement on Enforcement Efforts against Discrimination and Bias in Automated Systems.” *Consumer Financial Protection Bureau*, May 2, 2025. https://www.consumerfinance.gov/about-us/newsroom/joint-statement-on-enforcement-efforts-against-discrimination-and-bias-in-automated-systems/ Back to reference: 12

[13] Groves, L., Metcalf, J., Kennedy, A., Vecchione, B., & Strait, A. “Auditing Work: Exploring the New York City Algorithmic Bias Audit Regime.” arXiv, February 2024. https://arxiv.org/html/2402.08101v1 Back to reference: 13

[14] Krug, Steve. *Don’t Make Me Think: A Common Sense Approach to Web Usability.* 2nd ed. Berkeley: New Riders, 2005. Back to reference: 14

[1] Singh, Poornima, Sanjay Singh, and Gayatri S. Pandi-Jain. “Effective Heart Disease Prediction System Using Data Mining Techniques.” *International Journal of Nanomedicine* 13, no. T-NANO 2014 Abstracts (March 15, 2018): 121–124.

https://doi.org/10.2147/IJN.S124998.

[2] Sharma, Atul, Ayush Singh, Anup Kumar, and Elaiyaraja P. “Heart Disease Predictor.” *Journal of Emerging Technologies and Innovative Research* 10, no. 5 (May 2023): 64-70. https://www.jetir.org/papers/JETIR2305838.pdf.

[3] Pandita, Rahul. “How Netflix & Amazon Use AI to Do Better Customer Segmentation.” *Teqfocus Consulting LLC*, July 14, 2024. https://www.teqfocus.com/blog/how-netflix-amazon-use-ai-to-do-better-customer-segmentation.

[4] Rafalski, Kacper. “AI in Logistics - Top Benefits and Practical Applications.” *Netguru*, April 24, 2025. https://www.netguru.com/blog/ai-in-logistics.

[5] Robin, Meredith. 2025. “A Global Dress Distributor Gave AI a Chance. Now It Makes Creative Content Faster, and Overstock Is Down 40%.” *Business Insider*, March 7, 2025. https://www.businessinsider.com/wholesale-formal-gown-distributor-using-ai-for-ecommerce-operations.

[6] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. “Deep Learning.” *Nature* 521 (7553): 436–444. https://doi.org/10.1038/nature14539.

[7] Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, et al. 2021. “On the Opportunities and Risks of Foundation Models.” *arXiv preprint* . https://arxiv.org/abs/2108.07258.

[8] WealthStack Roundup. “AI @ Morgan Stanley Assistant Now Live.” https://www.wealthmanagement.com/artificial-intelligence/wealthstack-roundup-ai-morgan-stanley-assistant-now-live

[9] InterVision Systems. “What is Multimodal Gen AI and How Can it Evolve Your Business Strategy?” March 13, 2025. https://intervision.com/blog-what-is-multimodal-gen-ai-and-how-can-it-evolve-your-business-strategy/

[10] Rubio-Licht, Nat. “Adobe’s Multimodal Model Could Make AI Summaries Clearer.” *The Daily Upside*, January 25, 2024. https://www.thedailyupside.com/adobe-multimodal-model-ai-summaries-clearer.

[11] Brown, Tom B., Benjamin Mann, Nick Ryder, et al. 2020. “Language Models Are Few-Shot Learners.” *Advances in Neural Information Processing Systems* 33: 1877–1901. https://arxiv.org/abs/2005.14165.

[12] Yoo TK, Choi JY, Kim HK. Feasibility study to improve deep learning in OCT diagnosis of rare retinal diseases with few-shot classification. Med Biol Eng Comput. 2021 Feb;59(2):401-415. doi: 10.1007/s11517-021-02321-1. Epub 2021 Jan 25. PMID: 33492598; PMCID: PMC7829497.

[13] Schwartz, Roy, Jesse Dodge, Noah A. Smith, and Oren Etzioni. 2020. “Green AI.” *Communications of the ACM* 63 (12): 54–63. https://doi.org/10.1145/3381831.

[14] Schwartz, Roy, Jesse Dodge, Noah A. Smith, and Oren Etzioni. 2020. “Green AI.” *Communications of the ACM* 63 (12): 54–63. https://doi.org/10.1145/3381831.

[15] Henry, Parker. “How Duolingo Uses AI to Create Lessons Faster.” Last modified June 22, 2023. Accessed May 12, 2025. https://www.duolingo.com/blog/how-duolingo-uses-ai.

[16] Epoch AI, *AI Model Timeline*, accessed June 7, 2025, https://epoch.ai/data

[17] https://www.bloomberg.com/news/articles/2013-05-07/tesla-ceo-talking-with-google-about-autopilot-systems

[18] Synopsys Editorial Staff. “The 6 Levels of Vehicle Autonomy Explained.” Synopsys, February 15, 2025. https://www.synopsys.com/automotive/autonomous-vehicles/6-levels-of-vehicle-autonomy.html.

[19] Synopsys Editorial Staff. “The 6 Levels of Vehicle Autonomy Explained.” Synopsys, February 15, 2025. https://www.synopsys.com/automotive/autonomous-vehicles/6-levels-of-vehicle-autonomy.html.

[1] Singh, P., Singh, S., & Pandi-Jain, G. S. (March 15, 2018). “Effective Heart Disease Prediction System Using Data Mining Techniques.” *International Journal of Nanomedicine* 13(T-NANO 2014 Abstracts), 121–124. https://doi.org/10.2147/IJN.S124998 Back to reference: 1

[2] Sharma, A., Singh, A., Kumar, A., & Elaiyaraja, P. (May 2023). “Heart Disease Predictor.” *Journal of Emerging Technologies*

and Innovative Research 10(5), 64-70. https://www.jetir.org/papers/JETIR2305838.pdf Back to reference: 2

[3] Pandita, Rahul. "How Netflix & Amazon Use AI to Do Better Customer Segmentation." *Teqfocus Consulting LLC*, July 14, 2024. https://www.teqfocus.com/blog/how-netflix-amazon-use-ai-to-do-better-customer-segmentation Back to reference: 3

[4] Rafalski, Kacper. "AI in Logistics—Top Benefits and Practical Applications." *Netguru*, April 24, 2025. https://www.netguru.com/blog/ai-in-logistics Back to reference: 4

[5] Robin, Meredith. "A Global Dress Distributor Gave AI a Chance. Now It Makes Creative Content Faster, and Overstock Is Down 40%." *Business Insider*, March 7, 2025. https://www.businessinsider.com/wholesale-formal-gown-distributor-using-ai-for-ecommerce-operations Back to reference: 5

[6] LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep Learning." *Nature* 521(7553), 436–444. https://doi.org/10.1038/nature14539 Back to reference: 6

[7] Bommasani, R., Hudson, D. A., Adeli, E., et al. (2021). "On the Opportunities and Risks of Foundation Models." *arXiv preprint*. https://arxiv.org/abs/2108.07258 Back to reference: 7

[8] WealthStack Roundup. "AI @ Morgan Stanley Assistant Now Live." https://www.wealthmanagement.com/artificial-intelligence/wealthstack-roundup-ai-morgan-stanley-assistant-now-live Back to reference: 8

[9] InterVision Systems. "What is Multimodal Gen AI and How Can it Evolve Your Business Strategy?" March 13, 2025. https://intervision.com/blog-what-is-multimodal-gen-ai-and-how-can-it-evolve-your-business-strategy/. Back to reference: 9

[10] Rubio-Licht, Nat. "Adobe's Multimodal Model Could Make AI Summaries Clearer." *The Daily Upside*, January 25, 2024. https://www.thedailyupside.com/adobe-multimodal-model-ai-summaries-clearer Back to reference: 10

[11] Brown, T. B., Mann, B., Ryder, N., et al. (2020). "Language Models Are Few-Shot Learners." *Advances in Neural Information Processing Systems*, 33, 1877–1901. https://arxiv.org/abs/2005.14165 Back to reference: 11

[12] Yoo, T. K., Choi, J. Y., & Kim, H. K. "Feasibility Study to Improve Deep Learning in OCT Diagnosis of Rare Retinal Diseases with Few-Shot Classification." *Medical & Biological Engineering Computing*, 59(2), 401–415. doi: 10.1007/s11517-021-02321-1. Epub 2021 Jan 25. PMID: 33492598; PMCID: PMC7829497. Back to reference: 12

[13] Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). "Green AI." *Communications of the ACM* 63(12), 54–63. https://doi.org/10.1145/3381831 Back to reference: 13

[14] Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). "Green AI." *Communications of the ACM* 63(12), 54–63. https://doi.org/10.1145/3381831 Back to reference: 14

[15] Henry, Parker. "How Duolingo Uses AI to Create Lessons Faster." Last modified June 22, 2023. https://www.duolingo.com/blog/how-duolingo-uses-ai Back to reference: 15

[16] Epoch AI, *AI Model Timeline*, accessed June 7, 2025, https://epoch.ai/data

[17] https://www.bloomberg.com/news/articles/2013-05-07/tesla-ceo-talking-with-google-about-autopilot-systems

[16] Synopsys Editorial Staff. "The 6 Levels of Vehicle Autonomy Explained." February 15, 2025. https://www.synopsys.com/automotive/autonomous-vehicles/6-levels-of-vehicle-autonomy.html Back to reference: 16

[17] Synopsys Editorial Staff. "The 6 Levels of Vehicle Autonomy Explained." February 15, 2025. https://www.synopsys.com/automotive/autonomous-vehicles/6-levels-of-vehicle-autonomy.html. Back to reference: 17

[1] LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep Learning." *Nature* 521(7553), 436–444. https://doi.org/10.1038/nature14539 Back to reference: 1

[2] McCulloch, Warren S., & Pitts, Walter. (1943). "A Logical Calculus of the Ideas Immanent in Nervous Activity." *The Bulletin of Mathematical Biophysics* 5(4), 115–133. https://doi.org/10.1007/BF02478259 Back to reference: 2

[3] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). "Learning Representations by Back-Propagating Errors." *Nature* 323(6088), 533–536. https://doi.org/10.1038/323533a0 Back to reference: 3

[4] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. Available online: https://www.deeplearningbook.org/ Back to reference: 4

[5] LeCun, Y., Lon Bottou, L., Bengio, Y., & Haffner, P. (1998). “Gradient-Based Learning Applied to Document Recognition.” *Proceedings of the IEEE* 86(11), 2278–2324. https://doi.org/10.1109/5.726791 Back to reference: 5

[6] Hochreiter, Sepp, & Schmidhuber, Jürgen. (1997). “Long Short-Term Memory.” *Neural Computation* 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735 Back to reference: 6

[7] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Aidan N. Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). “Attention Is All You Need.” *Advances in Neural Information Processing Systems* 30(NeurIPS 2017). https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf Back to reference: 7

[8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Aidan N. Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). “Attention Is All You Need.” *Advances in Neural Information Processing Systems* 30(NeurIPS 2017). https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf Back to reference: 8

[9] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Aidan N. Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). “Attention Is All You Need.” *Advances in Neural Information Processing Systems* 30(NeurIPS 2017). https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf Back to reference: 9

[10] Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). “Recent Trends in Deep Learning Based Natural Language Processing.” *IEEE Computational Intelligence Magazine* 13(3), 55–75. https://doi.org/10.1109/MCI.2018.2840738 Back to reference: 10

[11] Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). “Recent Trends in Deep Learning Based Natural Language Processing.” *IEEE Computational Intelligence Magazine* 13(3), 55–75. https://doi.org/10.1109/MCI.2018.2840738 Back to reference: 11

[1] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. “Deep Learning.” *Nature* 521 (7553): 436–444. https://doi.org/10.1038/nature14539.

[2] McCulloch, Warren S., and Walter Pitts. 1943. “A Logical Calculus of the Ideas Immanent in Nervous Activity.” *The Bulletin of Mathematical Biophysics* 5 (4): 115–133. https://doi.org/10.1007/BF02478259.

[3] Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. 1986. “Learning Representations by Back-Propagating Errors.” *Nature* 323 (6088): 533–536. https://doi.org/10.1038/323533a0.

[4] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. Cambridge, MA: MIT Press._Available online: https://www.deeplearningbook.org/

[5] LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. “Gradient-Based Learning Applied to Document Recognition.” *Proceedings of the IEEE* 86 (11): 2278–2324. https://doi.org/10.1109/5.726791.

[6] Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. “Long Short-Term Memory.” *Neural Computation* 9 (8): 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

[7] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. “Attention Is All You Need.” *Advances in Neural Information Processing Systems* 30 (NeurIPS 2017). https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[8] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. “Attention Is All You Need.” *Advances in Neural Information Processing Systems* 30 (NeurIPS 2017). https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[9] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. “Attention Is All You Need.” *Advances in Neural Information Processing Systems* 30 (NeurIPS 2017). https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[10] Young, Tom, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. 2018. “Recent Trends in Deep Learning Based Natural Language Processing.” *IEEE Computational Intelligence Magazine* 13 (3): 55–75. https://doi.org/10.1109/MCI.2018.2840738.

[11] Young, Tom, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. 2018. “Recent Trends in Deep Learning Based Natural Language Processing.” *IEEE Computational Intelligence Magazine* 13 (3): 55–75. https://doi.org/10.1109/MCI.2018.2840738.

[1] Eloundou, T., Manning, S., Mishkin, P., & Rock, D. “GPTs Are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models.” *OpenAI*, March 2023. https://arxiv.org/abs/2303.10130 Back to reference: 1

[2] 618Media. “Services 618Media: #1 Digital Marketing Agency.” https://www.618media.com/services. Back to reference: 2

[3] Radford, Alec, et al. “Improving Language Understanding by Generative Pre-Training.” *OpenAI*, 2018. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf Back to reference: 3

[4] SendX. “A/B Testing: How to Optimize Your Email Campaigns.” SendX Blog. https://www.sendx.io/blog/email-subject-line-ab-testing#:~:text=You%20send%20each%20version%20to%20a%20small%2C,for%20the%20rest%20of%20your%20email%20lis Back to reference: 4

[5] Tom’s Guide. “OpenAI Launches Custom Chatbots—This Is How They Work and What They Can Do.” Last modified November 7, 2023. https://www.tomsguide.com/news/openai-launches-custom-chatbots-this-is-how-they-work-and-what-they-can-d Back to reference: 5

[6] Boyer, David. “Case Studies: How SMEs Are Using AI to Compete with Big Players.” *Inside Small Business*, November 4, 2024. https://insidesmallbusiness.com.au/technology/case-studies-how-smes-are-using-ai-to-compete-with-big-players Back to reference: 6

[7] Lufthansa Industry Solutions. “SmartAssistantAI: AI Chatbot Implementation in Line with Your Needs.” *Lufthansa Industry Solutions*. https://www.lufthansa-industry-solutions.com/de-en/solutions-products/artificial-intelligence/smartassistantai-ai-chatbot-implementation-in-line-with-your-needs Back to reference: 7

[8] Elomia. “Elomia: Your Mental Health AI Companion.” *Elomia*. https://elomia.com/ Back to reference: 8

[9] Humanly. “Conversational AI for Hiring.” *Humanly.io*. https://www.humanly.io/ Back to reference: 9

[10] Duolingo. “Introducing Duolingo Max: AI-Powered Learning with GPT-4.” *Duolingo Blog*, March 14, 2023. https://blog.duolingo.com/duolingo-max/ Back to reference: 10

[11] Tidio. “80+ Chatbot Statistics & Trends in 2025 [Usage, Adoption Rates].” Last modified October 18, 2024. https://www.tidio.com/blog/chatbot-statistics/ Back to reference: 11

[12] Tidio. “80+ Chatbot Statistics & Trends in 2025 [Usage, Adoption Rates].” Last modified October 18, 2024. https://www.tidio.com/blog/chatbot-statistics/ Back to reference: 12

[13] Parsani, Puran. “Beauty and the Bot: How Sephora Reimagined Customer Experience with AI.” *Cut the SaaS*, February 23, 2024. https://www.cut-the-saas.com/ai/beauty-and-the-bot-how-sephora-reimagined-customer-experience-with-ai Back to reference: 13

[14] Autodesk. (2019, November 19). “Autodesk and Airbus Demonstrate the Impact of Generative Design on Making and Building.” *Autodesk News.* https://adsknews.autodesk.com/en/news/autodesk-airbus-generative-design-aerospace-factory/ Back to reference: 14

[15] Autodesk. (2019, November 19). “Autodesk and Airbus Demonstrate the Impact of Generative Design on Making and Building.” *Autodesk News.* https://adsknews.autodesk.com/en/news/autodesk-airbus-generative-design-aerospace-factory/ Back to reference: 15

[16] BMW.com. “AI Design: Artificial Intelligence Gets Creative.” https://www.bmw.com/en/design/ai-design-and-digital-art.html Back to reference: 16

[17] Superior Data Science. “J.P Morgan—COiN—A Case Study of AI in Finance.” Last modified 2024. https://superiordatascience.tech/jpmorgan-coin-case-study Back to reference: 17

[18] UNESCO. “Generative AI: UNESCO Study Reveals Alarming Evidence of Regressive Gender Stereotypes.” *UNESCO*, March 7, 2024. https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regressive-gender-

stereotypes; Esposito, P., Atighehchian, P., Anastasis Germanidis, A., & Ghadiyaram, D. “Mitigating Stereotypical Biases in Text-to-Image Generative Systems.” *arXiv*, October 10, 2023. https://arxiv.org/abs/2310.06904 Back to reference: 18

[19] Superior Data Science. “J.P Morgan—COiN—A Case Study of AI in Finance.” Last modified 2024. https://superiordatascience.tech/jpmorgan-coin-case-study Back to reference: 19

[20] Allyn, Bobby. “Lucrative Tech Jobs Vanish as Klarna Replaces Workers with AI Chatbot.” *SFGATE*, May 17, 2024. https://www.sfgate.com/tech/article/lucrative-tech-jobs-vanish-klarna-gig-work-20339137.php Back to reference: 20

[21] OpenAI. “GPT-4.” https://www.openai.com/gpt-4 Back to reference: 21

[22] Anthropic. “Claude.” https://www.anthropic.com/claude Back to reference: 22

[23] Google. “Gemini.” https://www.google.com/gemini Back to reference: 23

[24] Cohere. “Enterprise Language Models.” https://www.cohere.com Back to reference: 24

[25] Jasper. “Generative AI for Marketing.” https://www.jasper.ai Back to reference: 25

[26] Grammarly. “GrammarlyGO.” https://www.grammarly.com/grammarlygo Back to reference: 26

[27] Copy.ai. “AI Writing Tools.” https://www.copy.ai Back to reference: 27

[28] Harvey. “AI for Legal Tasks.” https://www.harvey.ai Back to reference: 28

[29] Nabla. “AI in Healthcare.” https://www.nabla.com Back to reference: 29

[30] Replit. “AI-Powered Coding Environment.” https://www.replit.com Back to reference: 30

[31] Amazon Web Services. “AWS Bedrock.” https://aws.amazon.com/bedrock Back to reference: 31

[32] Microsoft Azure. “Azure AI Studio.” https://azure.microsoft.com/en-us/services/machine-learning/ Back to reference: 32

[33] NVIDIA. “DGX Systems.” https://www.nvidia.com/en-us/data-center/dgx-systems/ Back to reference: 33

[34] Hugging Face. “Open Models and Enterprise Support.” https://www.huggingface.co Back to reference: 34

[35] Mistral. “Open-Source Models.” https://www.mistral.ai Back to reference: 35

[36] Stability AI. “Open Models and Services.” https://www.stability.ai Back to reference: 36

[1] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). “Efficient Estimation of Word Representations in Vector Space.” arXiv preprint *arXiv*:1301.3781. https://arxiv.org/abs/1301.3781 Back to reference: 1

[2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). “Attention Is All You Need.” *Advances in Neural Information Processing Systems* 30. https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf Back to reference: 2

[3] Roberts, Jesse. “How Powerful Are Decoder-Only Transformer Neural Models?” *International Joint Conference on Neural Networks (IJCNN)*, 2024. https://arxiv.org/abs/2305.17026 Back to reference: 3

[4] Ho, J., Jain, A., & Abbeel, P. (2020). “Denoising Diffusion Probabilistic Models.” *arXiv* preprint arXiv:2006.11239. https://arxiv.org/abs/2006.11239 Back to reference: 4

[5] Roberts, Jesse. “How Powerful are Decoder-Only Transformer Neural Models?” *International Joint Conference on Neural Networks (IJCNN)*, 2024. https://arxiv.org/abs/2305.17026 Back to reference: 5

[6] Roberts, Jesse. “How Powerful are Decoder-Only Transformer Neural Models?” *International Joint Conference on Neural Networks (IJCNN)*, 2024. https://arxiv.org/abs/2305.17026 Back to reference: 6

[7] Ho, J., Jain, A., & Abbeel, P. (2020). “Denoising Diffusion Probabilistic Models.” *arXiv* preprint arXiv:2006.11239. https://arxiv.org/abs/2006.11239 Back to reference: 7

[8] Maxime Lebonne. “Decoding Strategies in Large Language Models.” *Hugging Face Blog*, 2025. https://huggingface.co/blog/mlabonne/decoding-strategies Back to reference: 8

[16] *EpochAI*. AI Model Timeline. https://epoch.ai/data

[4] O16 Labs. (n.d.). *Cost and technologies used in AI chatbot development*. O16 Labs. https://www.o16labs.com/blog/cost-and-technologies-used-in-ai-chatbot-development

[1] Harroch, Dominique A., and Richard D. Harroch. “15 Quotes on the Future of AI.” *TIME* May 2025. https://time.com/partner-article/7279245/15-quotes-on-the-future-of-ai/.

[1] AvePoint. *Artificial Intelligence and Information Management Report 2024*. 2024. https://cdn.avepoint.com/pdfs/en/shifthappens/AI-IM-Whitepaper-v4.pdf Back to reference: 1

[2] Ptolemay. (n.d.). *SaaS development costs: Breakdown and savings*. https://www.ptolemay.com/post/saas-development-costs-breakdown-and-savings Back to reference: 2

[3] O16 Labs. (n.d.). *Cost and technologies used in AI chatbot development*. https://www.o16labs.com/blog/cost-and-technologies-used-in-ai-chatbot-development Back to reference: 3

[4] World Economic Forum. *The Future of Jobs Report 2023.* 2023. https://www.weforum.org/publications/the-future-of-jobs-report-2023/ Back to reference: 4

[5] Bain & Company. “Widening Talent Gap Threatens Executives’ AI Ambitions.” March 4, 2025. https://www.bain.com/about/media-center/press-releases/20252/widening-talent-gap-threatens-executives-ai-ambitions--bain--company/ Back to reference: 5

[6] Cox, J. (2024) “AI Skills Gaps Threaten to Exacerbate Labor Shortages, Study Shows,” *Forbes*, November 24, 2024. https://www.forbes.com/sites/josiecox/2024/11/12/ai-skills-gaps-threaten-to-exacerbate-labor-shortages-study-shows/ Back to reference: 6

[7] Singh, Gitanjali. “Reskilling and Upskilling in the AI Age.” *LinkedIn*, October 2024. https://www.linkedin.com/pulse/reskilling-upskilling-ai-age-gitanjali-singh-vexwc Back to reference: 7

[8] Murire, Obrain Tinashe. (2024). “Artificial Intelligence and Its Role in Shaping Organizational Work Practices and Culture.” *Administrative Sciences* 14(12), 316. https://doi.org/10.3390/admsci14120316 Back to reference: 8

[9] Bughin, J., Chui, M., Manyika, J., et al. *Notes from the AI Frontier: Modeling the Impact of AI on the World Economy*. McKinsey Global Institute, September 2018. https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy Back to reference: 9

[10] Chui, M., Yee, L., Hall, B., Singla, A., & Sukharevsky, A. “The State of AI in 2023: Generative AI’s Breakout Year.” McKinsey & Company, August 1, 2023. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year Back to reference: 10

[11] Deloitte AI Institute. “The State of Generative AI in the Enterprise 2024 Year-End Report.” 2024. https://www2.deloitte.com/us/en/pages/consulting/articles/state-of-generative-ai-in-enterprise.html Back to reference: 11

[12] Talwar, Kanika. “All the Ways Zappos.com Leverages AI — And How the E-tailer Determines When to Get Humans Involved.” *Footwear News*, May 15, 2024. https://footwearnews.com/business/retail/zappos-artificial-intelligence-inclusivity-nigel-storey-1203638594/ Back to reference: 12

[13] U.S. Government Accountability Office. “Artificial Intelligence May Help IRS Close the Tax Gap.” June 2024. https://www.gao.gov/blog/artificial-intelligence-may-help-irs-close-tax-gap Back to reference: 13

[14] Sharma, H. D., & Goyal, P. (2023). “An Analysis of Sentiment: Methods, Applications, and Challenges.” *Engineering Proceedings* 59(1), 68. https://www.mdpi.com/2673-4591/59/1/68 Back to reference: 14

[15] The Coca-Cola Company. “Coca–Cola Invites Digital Artists to ‘Create Real Magic’ Using New AI Platform.” March 20, 2023. https://www.coca-colacompany.com/media-center/coca-cola-invites-digital-artists-to-create-real-magic-using-new-ai-platform Back to reference: 15

[16] Unilever Global. “Utilising AI to Redefine the Future of Customer Connectivity.” July 31, 2024. https://www.unilever.com/news/news-search/2024/utilising-ai-to-redefine-the-future-of-customer-connectivity/ Back to reference: 16

[17] Lin, Belle “Companies Had Fun Experimenting with AI. Now They Have to Show the Returns.” *The Wall Street Journal*, March 29, 2025. https://www.wsj.com/articles/companies-had-fun-experimenting-with-ai-now-they-have-to-show-the-returns-2a683592 Back to reference: 17

[18] Informatica. “New Research from Informatica Reveals More Data Leaders Plan to Capitalize on Generative AI, but Data Quality Remains the Key Obstacle to Adoption.” January 31, 2024. https://www.informatica.com/about-us/news/news-releases/2024/01/20240131-new-research-from-informatica-reveals-more-data-leaders-plan-to-capitalize-on-generative-ai-but-data-quality-remains-the-key-obstacle-to-adoption.html Back to reference: 18

[19] TechTarget. “TechTarget and Informa Tech’s Digital Businesses Combine.” *CIO Dive*, February 22, 2024. https://www.ciodive.com/news/general-mills-rolls-out-millschat-generative-ai-tool/644321/ Back to reference: 19

[20] Johnston, Lisa. “P&G’s ChatPG Generative AI Tool Scaling across Business.” Consumer Goods Technology, July 18, 2024. https://www.cgt.com/news/pg-chatpg-generative-ai-tool-scaling-across-business/ Back to reference: 20

[21] Mobilunity. “Spotify Agile Model and Its Benefits for Businesses.” January 26, 2021. https://mobilunity.com/blog/spotify-agile-model-benefits/ Back to reference: 21

[22] Mobilunity. “Spotify Agile Model and Its Benefits for Businesses.” January 26, 2021. https://mobilunity.com/blog/spotify-agile-model-benefits/ Back to reference: 22

[23] “Sappi Streamlines Operations with Intelligent Automation.” UiPath Blog. https://www.uipath.com/resources/automation-case-studies/sappi-streamlines-operations-with-intelligent-automation Back to reference: 23

[24] Wininger, Shai. “Introducing AI Jim, the Lemonade Claims Bot.” Lemonade Blog. https://www.lemonade.com/blog/secret-behind-lemonades-instant-insurance/ Back to reference: 24

[25] Hsieh, Nelson. “Lemonade’s AI Jim and Insurance Fraud Detection.” Volt Equity Blog. Published March 19, 2021. Updated November 18, 2024. Accessed April 11, 2025. https://www.voltequity.com/blog/lemonades-ai-jim-and-insurance-fraud-detection Back to reference: 25

[26] Mohammed, Irshadullah Asim. “Artificial Intelligence in Supplier Selection and Performance Monitoring: A Framework for Supply Chain Managers.” *Educational Administration: Theory and Practice*, 2023. https://www.researchgate.net/publication/386598338 Back to reference: 26

[27] Waditwar, Prajkta. (2024). “The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations.” *Open Journal of Business and Management* 12(2), 385–400. https://www.scirp.org/journal/paperinformation?paperid=137238 Back to reference: 27

[28] Waditwar, Prajkta. (2024). “The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations.” *Open Journal of Business and Management* 12(2), 385–400. https://www.scirp.org/journal/paperinformation?paperid=137238 Back to reference: 28

[29] Waditwar, Prajkta. (2024). “The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations.” *Open Journal of Business and Management* 12(2), 385–400. https://www.scirp.org/journal/paperinformation?paperid=137238 Back to reference: 29

[30] Mohammed, Irshadullah Asim. “Artificial Intelligence in Supplier Selection and Performance Monitoring: A Framework for Supply Chain Managers.” *Educational Administration: Theory and Practice*, 2023. https://www.researchgate.net/publication/386598338; Waditwar, Prajkta. (2024). “The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations.” *Open Journal of Business and Management* 12(2), 385–400. https://www.scirp.org/journal/paperinformation?paperid=137238; Morgan Lewis. *Key Considerations for Evaluating an AI Vendor*. 2024. https://www.morganlewis.com/-/media/files/publication/outside-publication/article/2024/key-considerations-for-

evaluating-an-ai-vendor.pdf Back to reference: 30

[31] Waditwar, Prajkta. (2024). “The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations.” *Open Journal of Business and Management* 12(2), 385–400. https://www.scirp.org/journal/paperinformation?paperid=137238; Morgan Lewis. *Key Considerations for Evaluating an AI Vendor*. 2024. https://www.morganlewis.com/-/media/files/publication/outside-publication/article/2024/key-considerations-for-evaluating-an-ai-vendor.pdf Back to reference: 31

[32] Mohammed, Irshadullah Asim. “Artificial Intelligence in Supplier Selection and Performance Monitoring: A Framework for Supply Chain Managers.” *Educational Administration: Theory and Practice*, 2023. https://www.researchgate.net/publication/386598338; Morgan Lewis. *Key Considerations for Evaluating an AI Vendor*. 2024. https://www.morganlewis.com/-/media/files/publication/outside-publication/article/2024/key-considerations-for-evaluating-an-ai-vendor.pdf Back to reference: 32

[33] Waditwar, Prajkta. (2024). “The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations.” *Open Journal of Business and Management* 12(2), 385–400. https://www.scirp.org/journal/paperinformation?paperid=137238; “The Role of Artificial Intelligence in the Procurement Process.” *Procedia Computer Science* 220 (2023), 127–135. https://www.sciencedirect.com/science/article/pii/S1478409223000079; Morgan Lewis. *Key Considerations for Evaluating an AI Vendor*. 2024. https://www.morganlewis.com/-/media/files/publication/outside-publication/article/2024/key-considerations-for-evaluating-an-ai-vendor.pdf Back to reference: 33

[34] Mohammed, Irshadullah Asim. “Artificial Intelligence in Supplier Selection and Performance Monitoring: A Framework for Supply Chain Managers.” *Educational Administration: Theory and Practice*, 2023. https://www.researchgate.net/publication/386598338; Waditwar, Prajkta. (2024). “The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations.” *Open Journal of Business and Management* 12(2), 385–400. https://www.scirp.org/journal/paperinformation?paperid=137238; Morgan Lewis. *Key Considerations for Evaluating an AI Vendor*. 2024. https://www.morganlewis.com/-/media/files/publication/outside-publication/article/2024/key-considerations-for-evaluating-an-ai-vendor.pdf Back to reference: 34

[35] Mohammed, Irshadullah Asim. “Artificial Intelligence in Supplier Selection and Performance Monitoring: A Framework for Supply Chain Managers.” *Educational Administration: Theory and Practice*, 2023. https://www.researchgate.net/publication/386598338; Morgan Lewis. *Key Considerations for Evaluating an AI Vendor*. 2024. https://www.morganlewis.com/-/media/files/publication/outside-publication/article/2024/key-considerations-for-evaluating-an-ai-vendor.pdf Back to reference: 35

[36] Mohammed, Irshadullah Asim. “Artificial Intelligence in Supplier Selection and Performance Monitoring: A Framework for Supply Chain Managers.” *Educational Administration: Theory and Practice*, 2023. https://www.researchgate.net/publication/386598338 Back to reference: 36

[37] Mohammed, Irshadullah Asim. “Artificial Intelligence in Supplier Selection and Performance Monitoring: A Framework for Supply Chain Managers.” *Educational Administration: Theory and Practice*, 2023. https://www.researchgate.net/publication/386598338; Waditwar, Prajkta. (2024). “The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations.” *Open Journal of Business and Management* 12(2), 385–400. https://www.scirp.org/journal/paperinformation?paperid=137238; Morgan Lewis. *Key Considerations for Evaluating an AI Vendor*. 2024. https://www.morganlewis.com/-/media/files/publication/outside-publication/article/2024/key-considerations-for-evaluating-an-ai-vendor.pdf Back to reference: 37

[38] Morgan Lewis. *Key Considerations for Evaluating an AI Vendor*. 2024. https://www.morganlewis.com/-/media/files/publication/outside-publication/article/2024/key-considerations-for-evaluating-an-ai-vendor.pdf Back to reference: 38

[39] Waditwar, Prajkta. (2024). “The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations.” *Open Journal of Business and Management* 12(2), 385–400. https://www.scirp.org/journal/paperinformation?paperid=137238 Back to reference: 39

[40] Waditwar, Prajkta. (2024). “The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations.” *Open Journal of Business and Management* 12(2), 385–400. https://www.scirp.org/journal/paperinformation?paperid=137238 Back to reference: 40

[41] “The Role of Artificial Intelligence in the Procurement Process.” *Procedia Computer Science* 220 (2023), 127–135. https://www.sciencedirect.com/science/article/pii/S1478409223000079; Morgan Lewis. *Key Considerations for Evaluating an AI Vendor*. 2024. https://www.morganlewis.com/-/media/files/publication/outside-publication/article/2024/key-considerations-for-

evaluating-an-ai-vendor.pdf Back to reference: 41

[42] "The Role of Artificial Intelligence in the Procurement Process." *Procedia Computer Science* 220 (2023), 127–135. https://www.sciencedirect.com/science/article/pii/S1478409223000079 Back to reference: 42

[43] Kumo. "How Much Does AI Chatbot Development Cost?" Last modified February 21, 2025. https://www.kumohq.co/blog/ai-chatbot-development-cost Back to reference: 43

[44] Phenom. "Electrolux Group Digitalizes Key HR Processes for Distinct Hiring Edge." *Phenom*. https://www.phenom.com/resource/electrolux-group-digitalizes-processes-hiring-edge Back to reference: 44

[45] Verizon. *Verizon 2023 TCFD Report*. March 2023. https://www.verizon.com/about/sites/default/files/Verizon-2023-TCFD-Report.pdf Back to reference: 45

[46] Verizon. "Verizon Unveils AI Strategy to Power Next-Gen AI Demands." *Verizon News Center*, February 21, 2024. https://www.verizon.com/about/news/verizon-unveils-ai-strategy-power-next-gen-ai-demands Back to reference: 46

[47] Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366, no. 6464 (2019): 447-453. https://www.science.org/doi/10.1126/science.aax2342. Back to reference: 47

[48] Forbes Finance Council. (2023, March 9). "The Future of Mortgage Lending: How AI and Humans Can Coexist." *Forbes*. https://www.forbes.com/councils/forbesfinancecouncil/2023/03/09/the-future-of-mortgage-lending-how-ai-and-humans-can-coexist/ Back to reference: 48

[49] Court, E. (2024, April 23). "AI Lawsuits against Insurers Signal Wave of Health Litigation." *Bloomberg Law*. https://news.bloomberglaw.com/health-law-and-business/ai-lawsuits-against-insurers-signal-wave-of-health-litigation Back to reference: 49

[50] Government of Singapore. "SkillsFuture: A Decade and Beyond." SkillsFuture. Last modified May 23, 2025. https://www.skillsfuture.gov.sg Back to reference: 50

[51] World Economic Forum. (2023, May 2). "Lessons from Singapore on Upskilling for the Future." *World Economic Forum*, https://www.weforum.org/stories/2023/05/lessons-from-singapore-on-upskilling-for-the-future/ Back to reference: 51

[52] Lin, Belle. "Why Amazon Is Betting on 'Automated Reasoning' to Reduce AI's Hallucinations," *The Wall Street Journal*, February 5, 2025. https://www.wsj.com/articles/why-amazon-is-betting-on-automated-reasoning-to-reduce-ais-hallucinations-b838849e. WSJ Back to reference: 52

[53] AllBusiness.com Team. "Retrieval-Augmented Generation (RAG)." *TIME*, April 3, 2025. https://time.com/collection_hub_item/definition-of-retrieval-augmented-generation/ Back to reference: 53

[54] Gupta, Abhishek. "Guardrails for Amazon Bedrock Can Now Detect Hallucinations and Safeguard Apps Built Using Custom or Third-Party FMs," *AWS News Blog*, July 10, 2024. https://aws.amazon.com/blogs/aws/guardrails-for-amazon-bedrock-can-now-detect-hallucinations-and-safeguard-apps-built-using-custom-or-third-party-fms/ Back to reference: 54

[55] Cecco, Leyland. "Air Canada Ordered to Pay Customer Who Was Misled by Airline's Chatbot." *The Guardian*, February 16, 2024. https://www.theguardian.com/world/2024/feb/16/air-canada-chatbot-compensation Back to reference: 55

National Parks Board. 2022. *Skyrise Greenery Incentive Scheme*. Accessed April 2, 2025. https://skyrisegreenery.nparks.gov.sg/incentive-scheme/about/

ST Engineering. 2023. *Smart Estate Lighting*. Accessed April 2, 2025. https://www.stengg.com/en/smart-city/urban-solutions/smart-utilities-and-infrastructure/agil-smart-lighting/agil-smart-estate-lighting/

Land Transport Authority (LTA). 2022. *Passenger Information Display Systems*. Accessed April 2, 2025. https://www.lta.gov.sg/content/ltagov/en/who_we_are/statistics_and_publications/Connect/pids.html

[1] National Parks Board. 2022. *Skyrise Greenery Incentive Scheme*. https://skyrisegreenery.nparks.gov.sg/incentive-scheme/about/ Back to reference: 1

[2] ST Engineering. 2023. *Smart Estate Lighting*. https://www.stengg.com/en/smart-city/urban-solutions/smart-utilities-and-

infrastructure/agil-smart-lighting/agil-smart-estate-lighting/ Back to reference: 2

[3] Land Transport Authority (LTA). 2022. *Passenger Information Display Systems*. https://www.lta.gov.sg/content/ltagov/en/who_we_are/statistics_and_publications/Connect/pids.html Back to reference: 3

[4] Land Transport Authority. 2022. *Smart Mobility 2030*. https://www.lta.gov.sg/ Back to reference: 4

[5] Monetary Authority of Singapore. 2021. *E-payments in Singapore*. https://www.mas.gov.sg/ Back to reference: 5

[6] Infocomm Media Development Authority. (2022). *AI and IoT in Singapore's Retail Sector*. https://www.imda.gov.sg/business Back to reference: 6

[7] Urban Redevelopment Authority. (2023). *Singapore Green Plan 2030*. Retrieved from https://www.ura.gov.sg/ Back to reference: 7

[8] Singapore Green Plan 2030. "A City of Green Possibilities." Singapore Government Agency Website. Last modified March 8, 2025. https://www.greenplan.gov.sg Back to reference: 8

[9] Building and Construction Authority. "Green Mark 2021 Certification Standards." *BuildSG*, last updated February 25, 2025. https://www.buildsg.gov.sg Back to reference: 9

[6] Gratch, J., Wang, N., Gerten, B., Strader, M., Nordquist, D., & Larson, K. (2004). *Creating Rapport in Virtual Humans*. IEEE Intelligent Systems, 19(4), 42–50; Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., ... & Alcañiz, M. (2007). *Affective Interactions Using Virtual Reality: The Link Between Presence and Emotions*. CyberPsychology & Behavior, 10(1), 45–56. Back to reference: 6

[10] European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).* Official Journal of the European Union, L 119, 1–88. Back to reference: 10

[17] Federal Communications Commission (FCC), 2021 Broadband Deployment Report: https://docs.fcc.gov/public/attachments/FCC-21-18A1.pdf Back to reference: 17

[18] Telecom Regulatory Authority of India (TRAI), 2022: https://trai.gov.in/ Back to reference: 18

[21] Horrigan, John B., and Katherine Murray. "Rural America Trails Rest of Nation in Home Broadband Use, but Gap Has Narrowed in Past Two Years." *Pew Research Center*, February 26, 2006 https://www.pewresearch.org/internet/2006/02/26/rural-america-trails-rest-of-nation-in-home-broadband-use-but-gap-has-narrowed-in-past-two-years/; Goldberg, Rafi. "New NTIA Data Show 13 Million More Internet Users in the U.S. in 2023 than 2021." *National Telecommunications and Information Administration*, June 6, 2024. https://www.ntia.gov/blog/2024/new-ntia-data-show-13-million-more-internet-users-us-2023-2021 Back to reference: 21

[22] Amazon Staff. "Upskilling 2025." *Amazon News*, October 2, 2020. https://www.aboutamazon.com/news/workplace/upskilling-2025 Back to reference: 22

[23] Google. "Grow with Google." *Grow with Google*. https://grow.google/ Back to reference: 23

[24] Microsoft. "Microsoft Launches Next Stage of Skills Initiative After Helping 30 Million People." *Microsoft Stories* https://news.microsoft.com/skills/ Back to reference: 24

[25] O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown, 2016. Back to reference: 25

[26] Orlowski, Jeff, dir. *The Social Dilemma*. Exposure Labs, 2020. Netflix. Msp page 1-23 (paragraph "the documentary") Back to reference: 26

[27] Panettieri, Joe. "Generative AI Lawsuits Timeline: Legal Cases vs. OpenAI, Microsoft, Anthropic, Nvidia, Perplexity, Intel and More." *Sustainable Tech Partner*, January 9, 2025; https://sustainabletechpartner.com/topics/ai/generative-ai-lawsuit-timeline/ Back to reference: 27

[28] Smiley, Lauren. “The Legal Saga of Uber’s Fatal Self-Driving Car Crash Is Over.” *WIRED*, July 28, 2023. https://www.wired.com/story/ubers-fatal-self-driving-car-crash-saga-over-operator-avoids-prison/ Back to reference: 28