

Contents

Preface — VII

Part I: Innovation

- 1 Innovation Dynamics: AI's Transformative Potential — 3**
 - 1.1 What Is a General Purpose Technology? — 3
 - 1.2 AI: The Next General Purpose Technology — 6
 - 1.3 The Productivity Paradox: Why GPT Benefits Take Time — 8
 - 1.4 Capturing AI's GPT Potential: A Playbook for Leaders — 10
 - 1.5 Conclusion: Preparing for the AI-Driven Future — 12
- 2 Innovation Culture: Igniting AI Creativity in Organizations — 14**
 - 2.1 Prototypes as Innovation Catalysts — 14
 - 2.2 The Experimental Organization: Test and Learn — 16
 - 2.3 Small Teams, Big Learning — 20
 - 2.4 Engaging Lead Users: Innovation from the Outside In — 21
 - 2.5 Conclusion — 24

Part II: Computation

- 3 Algorithms: The Logic of Computing — 27**
 - 3.1 What Is an Algorithm? — 27
 - 3.2 An Algorithm in Action — 30
 - 3.3 Making Algorithms Better — 31
 - 3.4 Algorithmic Speed — 33
- 4 Architecture: How Computers Work — 35**
 - 4.1 Inventing the Modern Computer — 35
 - 4.2 The Core Components of a von Neumann Machine — 37
 - 4.3 A Simple Programming Example — 39
 - 4.4 Multiple Cores — 40
 - 4.5 Graphics Processing Units — 41
 - 4.6 Conclusion — 43

Part III: Machine Learning

- 5 Machine Learning: Learning from Data — 47**
 - 5.1 Traditional Programming vs. Machine Learning — 47
 - 5.2 The ImageNet Breakthrough — 50
 - 5.3 Types of Machine Learning — 54

- 6 Classification: Predicting Categories (Case Study) — 58**
 - 6.1 Problem Statement — 58
 - 6.2 Supervised Machine Learning Model Development Workflow — 59
 - 6.3 Step 1: Prepare the Data — 60
 - 6.4 Step 2: Choose the Model — 62

- 7 Regression: Predicting Continuous Values (Case Study) — 68**
 - 7.1 Problem Statement — 68
 - 7.2 Step 1: Prepare the Data — 69
 - 7.3 Step 2: Choose the Model — 70
 - 7.4 A Theoretical Interlude — 70
 - 7.5 Step 3: Fit the Model to the Data — 74
 - 7.6 Step 4: Evaluate the Model — 76
 - 7.7 Step 5: Predict New Data — 77
 - 7.8 Step 6: Human-in-the-Loop—Applications in Education — 78
 - 7.9 Conclusion — 79

Part IV: Deep Learning

- 8 Deep Learning: The Big Picture — 83**
 - 8.1 Neural Networks in Biology and Early AI — 83
 - 8.2 The Five Pillars of Deep Learning — 89
 - 8.3 Conclusion — 91

- 9 Deep Learning: Under the Hood — 92**
 - 9.1 Layers — 95
 - 9.2 Forward Propagation — 98
 - 9.3 The Loss Function — 100
 - 9.4 Backpropagation — 102
 - 9.5 Conclusion — 103

Part V: Generative AI

- 10 Foundation Models: Large Language Model Basics — 107**
- 11 Foundation Models: How Large Language Models Become Smart — 121**
 - 11.1 The Context Problem — 121
 - 11.2 Context Windows: The Scope of Understanding — 122
 - 11.3 Transformer Architecture — 124
 - 11.4 The Attention Mechanism — 126
 - 11.4.1 Positional Encoding — 126
 - 11.4.2 Visualizing the Attention Mechanism — 127
 - 11.5 Vector Dimensions: Beyond Position in Space — 127
 - 11.6 Feed-Forward Layer — 129
 - 11.7 Attention Beyond Ambiguity — 130
 - 11.8 Putting It All Together: A Complete Example — 131
 - 11.9 Conclusion — 132
- 12 Expert Models: Grounding AI in Trusted Knowledge — 133**
 - 12.1 Comparing Approaches—Foundation vs. Expert Models — 133
 - 12.2 The Limitations of Foundation Models — 135
 - 12.3 Applied Example: Medical Guidance in Clinical Settings — 138
 - 12.4 How Expert Models Work — 141
 - 12.5 Standalone Expert Models: An Alternative Approach — 144
 - 12.6 Design Considerations and Trade-offs — 145
 - 12.7 Conclusion — 148
- 13 Agentic AI: From Knowledge to Action — 149**
 - 13.1 Overview — 149
 - 13.2 The Need for AI Agents — 150
 - 13.3 The Reflection Dimension — 151
 - 13.4 The Action Dimension — 152
 - 13.5 The Adaptation Dimension — 153
 - 13.6 How AI Agents Work: The ReACT Framework — 155
 - 13.7 Conclusion — 157

Part VI: Risk

- 14 System Failures: Why and How AI Makes Mistakes — 161**
 - 14.1 The Multi-Agent Paradigm and Its Promise — 162

- 14.2 The Berkeley Study: A Taxonomy of Failure — **162**
- 14.3 Failure Category 1: Poor Specification and System Design — **163**
- 14.4 Failure Category 2: Inter-Agent Misalignment — **164**
- 14.5 Failure Category 3: Verification and Quality Control — **165**
- 14.6 AI as a Complex System: Moving Beyond Technical Fixes — **165**
- 14.7 A Framework for Evaluating AI System Risk — **166**
- 14.8 Conclusion: Toward Reliable AI Systems — **167**

- 15 Catastrophic Accidents: How to Identify and Address Major Failures — 169**
- 15.1 Normal Accident Theory: Understanding Inevitable Failures — **169**
- 15.2 AI Systems Through Perrow’s Lens — **171**
- 15.3 A Normal Accident Theory Checklist for AI Systems — **172**
- 15.4 Case Studies: Normal Accidents in AI — **175**
- 15.5 Implications for AI System Design — **177**
- 15.6 Conclusion: Beyond Prevention to Resilience — **179**

- 16 Risk Across the Action Boundary: From Passive to Agentic Systems — 180**
- 16.1 The Action Boundary Defined — **180**
- 16.2 Risk Amplification at the Action Boundary — **182**
- 16.3 The Agentic Risk Assessment Matrix — **183**
- 16.4 Control Frameworks by Risk Level — **184**
- 16.5 Case Example: Financial Service Automation — **185**
- 16.6 The Organizational Readiness Component — **185**
- 16.7 Bridging Technical and Organizational Controls — **186**
- 16.8 The Path Forward: Progressive Agency — **187**
- 16.9 Conclusion: Agency as Responsibility — **188**

- Bibliography — 189**

- List of Figures — 191**

- List of Tables — 197**

- Index — 199**