Table of Contents

Preface xiii

1. Introduction to Agents 1
Defining AI Agents 1
The Pretraining Revolution 2
Types of Agents 3
Model Selection 5
From Synchronous to Asynchronous Operations 6
Practical Applications and Use Cases 7
Workflows and Agents 8
Principles for Building Effective Agentic Systems 11
Organizing for Success in Building Agentic Systems 12
Agentic Frameworks 13
  LangGraph 13
  AutoGen 14
  CrewAI 14
  OpenAI Agents Software Development Kit (SDK) 14
Conclusion 15

2. Designing Agent Systems 17
Our First Agent System 17
Core Components of Agent Systems 20
Model Selection 21
Tools 24
  Designing Capabilities for Specific Tasks 24
  Tool Integration and Modularity 25
Memory 25
  Short-Term Memory 26

Long-Term Memory 26
Memory Management and Retrieval 26
Orchestration 27
Design Trade-Offs 27
Performance: Speed/Accuracy Trade-Offs 27
Scalability: Engineering Scalability for Agent Systems 28
Reliability: Ensuring Robust and Consistent Agent Behavior 29
Costs: Balancing Performance and Expense 30
Architecture Design Patterns 32
Single-Agent Architectures 32
Multiagent Architectures: Collaboration, Parallelism, and Coordination 32
Best Practices 34
Iterative Design 34
Evaluation Strategy 35
Real-World Testing 37
Conclusion 39

3. User Experience Design for Agentic Systems. 41
Interaction Modalities 42
Text-Based 43
Graphical Interfaces 46
Speech and Voice Interfaces 50
Video-Based Interfaces 53
Combining Modalities for Seamless Experiences 54
The Autonomy Slider 55
Synchronous Versus Asynchronous Agent Experiences 58
Design Principles for Synchronous Experiences 58
Design Principles for Asynchronous Experiences 59
Finding the Balance Between Proactive and Intrusive Agent Behavior 59
Context Retention and Continuity 60
Maintaining State Across Interactions 61
Personalization and Adaptability 62
Communicating Agent Capabilities 63
Communicating Confidence and Uncertainty 64
Asking for Guidance and Input from Users 65
Failing Gracefully 65
Trust in Interaction Design 66
Conclusion 68

4. Tool Use. 71
LangChain Fundamentals 72
Local Tools 73

API-Based Tools 75
Plug-In Tools 78
Model Context Protocol 81
Stateful Tools 84
Automated Tool Development 85
Foundation Models as Tool Makers 85
Real-Time Code Generation 86
Tool Use Configuration 87
Conclusion 88

5. Orchestration.. 89
Agent Types 90
Reflex Agents 90
ReAct Agents 90
Planner-Executor Agents 91
Query-Decomposition Agents 91
Reflection Agents 91
Deep Research Agents 92
Tool Selection 93
Standard Tool Selection 94
Semantic Tool Selection 97
Hierarchical Tool Selection 101
Tool Execution 105
Tool Topologies 105
Single Tool Execution 106
Parallel Tool Execution 107
Chains 107
Graphs 109
Context Engineering 112
Conclusion 113

6. Knowledge and Memory.. 115
Foundational Approaches to Memory 116
Managing Context Windows 116
Traditional Full-Text Search 117
Semantic Memory and Vector Stores 119
Introduction to Semantic Search 119
Implementing Semantic Memory with Vector Stores 119
Retrieval-Augmented Generation 121
Semantic Experience Memory 122
GraphRAG 123
Using Knowledge Graphs 123

Building Knowledge Graphs 124
Promise and Peril of Dynamic Knowledge Graphs 130
Note-Taking 133
Conclusion 134

7. Learning in Agentic Systems....................................... 135
Nonparametric Learning 135
Nonparametric Exemplar Learning 135
Reflexion 137
Experiential Learning 141
Parametric Learning: Fine-Tuning 146
Fine-Tuning Large Foundation Models 146
The Promise of Small Models 151
Supervised Fine-Tuning 153
Direct Preference Optimization 158
Reinforcement Learning with Verifiable Rewards 161
Conclusion 162

8. From One Agent to Many.. 163
How Many Agents Do I Need? 163
Single-Agent Scenarios 163
Multiagent Scenarios 170
Swarms 177
Principles for Adding Agents 178
Multiagent Coordination 180
Democratic Coordination 180
Manager Coordination 181
Hierarchical Coordination 182
Actor-Critic Approaches 182
Automated Design of Agent Systems 184
Communication Techniques 189
Local Versus Distributed Communication 189
Agent-to-Agent Protocol 189
Message Brokers and Event Buses 192
Actor Frameworks: Ray, Orleans, and Akka 195
Orchestration and Workflow Engines 199
Managing State and Persistence 201
Conclusion 202

9. Validation and Measurement.. 205
Measuring Agentic Systems 205
Measurement Is the Keystone 206

Integrating Evaluation into the Development Lifecycle 207
Creating and Scaling Evaluation Sets 207
Component Evaluation 209
Evaluating Tools 209
Evaluating Planning 210
Evaluating Memory 212
Evaluating Learning 213
Holistic Evaluation 214
Performance in End-to-End Scenarios 214
Consistency 216
Coherence 217
Hallucination 218
Handling Unexpected Inputs 219
Preparing for Deployment 220
Conclusion 221

10. Monitoring in Production. 223
Monitoring Is How You Learn 224
Monitoring Stacks 226
Grafana with OpenTelemetry, Loki, and Tempo 227
ELK Stack (Elasticsearch, Logstash/Fluentd, Kibana) 227
Arize Phoenix 228
SigNoz 229
Langfuse 229
Choosing the Right Stack 230
OTel Instrumentation 230
Visualization and Alerting 232
Monitoring Patterns 235
Shadow Mode 235
Canary Deployments 235
Regression Trace Collection 236
Self-Healing Agents 236
User Feedback as an Observability Signal 236
Distribution Shifts 237
Metric Ownership and Cross-Functional Governance 239
Conclusion 241

11. Improvement Loops. 243
Feedback Pipelines 245
Automated Issue Detection and Root Cause Analysis 250
Human-in-the-Loop Review 251
Prompt and Tool Refinement 254

Aggregating and Prioritizing Improvements 259
Experimentation 260
Shadow Deployments 261
A/B Testing 262
Bayesian Bandits 263
Continuous Learning 265
In-Context Learning 265
Offline Retraining 267
Conclusion 268

12. Protecting Agentic Systems. 271
The Unique Risks of Agentic Systems 272
Emerging Threat Vectors 273
Securing Foundation Models 275
Defensive Techniques 276
Red Teaming 278
Threat Modeling with MAESTRO 281
Protecting Data in Agentic Systems 283
Data Privacy and Encryption 283
Data Provenance and Integrity 285
Handling Sensitive Data 286
Securing Agents 288
Safeguards 288
Protections from External Threats 290
Protections from Internal Failures 292
Conclusion 296

13. Human-Agent Collaboration. 297
Roles and Autonomy 297
The Changing Role of Humans in Agent Systems 298
Aligning Stakeholders and Driving Adoption 299
Scaling Collaboration 300
Agent Scope and Organizational Roles 302
Shared Memory and Context Boundaries 303
Trust, Governance, and Compliance 305
The Lifecycle of Trust 305
Accountability Frameworks 306
Escalation Design and Oversight 309
Privacy and Regulatory Compliance 310
Conclusion: The Future of Human-Agent Teams 312

Glossary 315

Index 319