Sumário

Prefácio xxv

Parte 1: Fundamentos de IA generativa e sistemas agênticos

1

Fundamentos de IA generativa 3

Introdução à IA generativa 4
Tipos de modelos de IA generativa 5
- VAEs 5
- GANs 6
- Modelos autorregressivos e arquitetura Transformer 7
- Agentes de IA baseados em LLMs 11

Aplicações de IA generativa 16
Desafios e limitações da IA generativa 20
- Qualidade dos dados e viés 21
- Privacidade de dados 21
- Recursos computacionais 22

Implicações éticas e sociais 23
Generalização e criatividade 23
Resumo 24
Perguntas 24
Respostas 24
Leitura adicional 25
Referências 25

2

Princípios de sistemas agênticos 27

Compreendendo autogovernança, capacidade de agir e autonomia 27
Autogovernança 28
Capacidade de agir 29
Autonomia 30
Exemplo de capacidade de agir e autonomia em agentes 31
Revisão de agentes inteligentes e suas características 35
Explorando a arquitetura de sistemas agênticos 36
Arquiteturas deliberativas 36
Arquiteturas reativas 38
Arquiteturas híbridas 39
Compreendendo sistemas multiagentes 41
Definição e características de SMAs 42
Mecanismos de interação em SMAs 43
Resumo 49
Perguntas 49
Respostas 50

3

Componentes essenciais de agentes inteligentes 51

Representação do conhecimento em agentes inteligentes 52
Redes semânticas 52
Quadros 54
Representações baseadas em lógica 55

Raciocínio em agentes inteligentes 57
Raciocínio dedutivo 57
Raciocínio indutivo 59
Raciocínio abdutivo 60
Mecanismos de aprendizado para agentes adaptativos 62
Tomada de decisão e planejamento em sistemas agênticos 63
Função utilitária 64
Algoritmos de planejamento 66
Planejamento baseado em grafos 66
Busca heurística 67
Monte Carlo Tree Search 68
Planejamento hierárquico 68
Satisfação de restrições 69
Aprimorando as capacidades de agentes com IA generativa 69
Comece a construir IA agêntica 71
Resumo 74
Perguntas 74
Respostas 74

Parte 2: Projetando e implementando agentes baseados em IA generativa

4

Reflexão e introspecção em agentes 79

A importância da reflexão em agentes 80
Tomada de decisão aprimorada 80
Adaptação 81
Considerações éticas 82
Interação humano-computador 83
Introspecção em agentes inteligentes 84
Implementando capacidades reflexivas 85
Raciocínio tradicional 85
Metarraciocínio 86
Monitoramento de performance 92
Métricas específicas para acompanhar 93

Como as métricas ajustam o comportamento 93
Alocação de recursos 94
Autoexplicação 95
Transparência 96
Aprendizado e refinamento 98
Engajamento do usuário e colaboração 99
Automodelagem 100
Gerenciamento de metas 101
Atualização de conhecimento 101
Casos de uso e exemplos 102
Chatbots de atendimento ao cliente 102
Agentes de marketing pessoal 103
Sistemas de negociação financeira 104
Agentes de previsão 105
Estratégias de preços no e-commerce 107
Resumo 108
Perguntas 109
Respostas 109

5

Habilitando o uso de ferramentas e planejamento em agentes 111

Compreendendo o conceito de uso de ferramentas por agentes 112
Chamada de ferramentas e funções 113
Definição de ferramentas para agentes 115
Abordagem baseada em framework – usando docstrings 115
Integração direta com LLM 116
Tipos de ferramentas 117
A importância das ferramentas em sistemas agênticos 120
Algoritmos de planejamento para agentes 121
Algoritmos de planejamento menos práticos 122
STRIPS 122
*Planejamento com A** 122
GraphPlan 122
Monte Carlo Tree Search (MCTS) 123

Algoritmo de planejamento moderadamente prático – FF 123
Algoritmos de planejamento mais práticos 124
Planejamento baseado em LLM 125
HTN 127
Integrando o uso de ferramentas e o planejamento 131
Raciocínio sobre ferramentas 131
Planejamento do uso de ferramentas 133
Explorando implementações práticas 134
Exemplo de CrewAI 134
Exemplo de AutoGen 136
Exemplo de LangGraph 138
Resumo 140
Perguntas 141
Respostas 141

6

Explorando a abordagem coordenador, executor e delegador 143

Compreendendo o Modelo CWD 144
Princípios-chave do Modelo CWD 145
O Modelo CWD para o agente de viagens inteligente 147
Projetando agentes com atribuição de papéis 149
Papéis e responsabilidades de cada agente 152
Comunicação e colaboração entre agentes 158
Comunicação 158
Mecanismo de coordenação 159
Negociação e resolução de conflitos 159
Compartilhamento de conhecimento 160
Implementando a abordagem CWD em sistemas de IA generativa 161
Prompts de sistema e comportamento do agente 161
Formatação de instruções 162
Padrões de interação 163
Resumo 164
Perguntas 164
Respostas 165

7

Técnicas eficazes de projeto de sistemas agênticos 167

Prompts de sistema focados e instruções para agentes 168
- Definindo objetivos 168
- Especificações de tarefas 169
- Consciência de contexto 171

Espaços de estados e modelagem do ambiente 173
- Representação do espaço de estados 173
- Modelagem de ambiente 175
- Padrões de integração e interação 176
- Monitoramento e adaptação 179

Arquitetura de memória do agente e gerenciamento de contexto 179
- Memória de curto prazo (memória de trabalho) 180
- Memória de longo prazo (base de conhecimento) 181
- Memória episódica (histórico de interação) 182
- Gerenciamento de contexto 183
- Integração com a tomada de decisão 185

Processamento sequencial e paralelo em fluxos de trabalho agênticos 186
- Processamento sequencial 186
- Processamento paralelo 187
- Otimização de fluxo de trabalho 188

Resumo 190
Perguntas 190
Respostas 191

Parte 3: Confiança, segurança, ética e aplicações

8

Construindo confiança em sistemas de IA generativa 195

Importância da confiança na IA 196
Técnicas para estabelecer confiança 197

Transparência e explicabilidade 197
Lidando com incerteza e vieses 201
Comunicação de saída eficaz 202
Controle e consentimento do usuário 203
Desenvolvimento ético e responsabilidade 204

Implementando transparência e explicabilidade 206
Tratamento de incerteza e vieses 207
Resumo 208
Perguntas 209
Respostas 209

9

Gerenciando segurança e considerações éticas 211

Compreendendo riscos e potenciais desafios 212
Ataques adversariais 212
Viés e discriminação 213
Desinformação e alucinações 215
Violações de privacidade de dados 216
Riscos de propriedade intelectual 217

Garantindo IA segura e responsável 219
Explorando diretrizes e frameworks éticos 223
Design centrado no ser humano 223
Prestação de contas e responsabilidade 223
Privacidade e proteção de dados 224
Envolvimento de stakeholders diversos 224

Abordando preocupações de privacidade e segurança 225
Resumo 226
Perguntas 227
Respostas 227

10

Casos de uso e aplicações comuns 229

Aplicações criativas e artísticas 230
Evolução de agentes criativos e artísticos 230
Aplicações do mundo real 231

Descrição do problema 231
Ambiente e sistemas externos 232
Por que isso é melhor 232
Processamento de linguagem natural e agentes conversacionais **233**
Evolução de agentes de linguagem 233
Aplicações do mundo real 234
Descrição do problema 234
Ambiente e sistemas externos 235
Por que isso é melhor 236
Robótica e sistemas autônomos **236**
Evolução de agentes robóticos 237
Aplicações no mundo real 237
Descrição do problema 237
Ambiente e sistemas externos 238
Por que isso é melhor 239
Suporte à decisão e otimização **240**
Evolução de agentes de apoio à decisão 240
Aplicações do mundo real 240
Declaração do problema 241
Ambiente e sistemas externos 242
Por que isso é melhor 243
Resumo **244**
Perguntas **245**
Respostas **245**

11

Conclusão e perspectivas futuras 247

Recapitulação dos principais conceitos **248**
Tendências emergentes e direções de pesquisa **249**
Inteligência multimodal – integrando entradas diversas 249
Compreensão avançada de linguagem 249
Aprendizagem baseada na experiência – inovações em Aprendizado por Reforço 250
Implicações práticas em diferentes setores 251
Inteligência artificial geral **251**
O que torna a AGI diferente 251
O grande desafio 251

Aprendendo a aprender 252
Compreensão do mundo real 252
Desafios e oportunidades **253**
Resumo **255**
Índice remissivo **257**