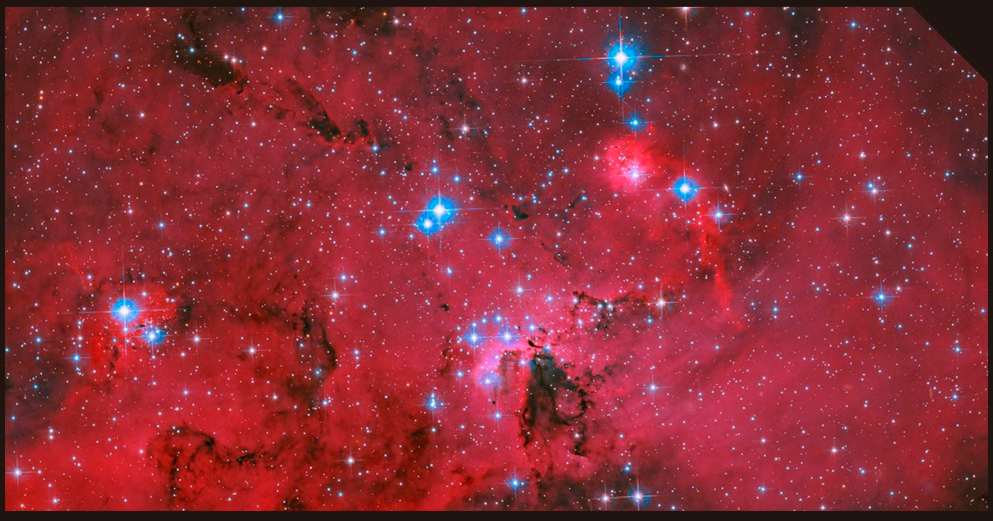
Blucher



# Construindo sistemas de IA agêntica

Crie agentes de IA inteligentes e autônomos que podem raciocinar, planejar e se adaptar

ANJANAVA BISWAS
WRICK TALUKDAR

# Construindo sistemas de IA agêntica

Crie agentes de IA inteligentes e autônomos que podem raciocinar, planejar e se adaptar

**Anjanava Biswas**

**Wrick Talukdar**

*Construindo sistemas de IA agêntica: crie agentes de IA inteligentes e autônomos que podem raciocinar, planejar e se adaptar*

Copyright © 2025 Packt Publishing

© 2026 Anjanava Biswas, Wrick Talukdar

Editora Edgard Blücher Ltda.

*Publisher* Edgard Blücher

*Editor* Eduardo Blücher

*Coordenação editorial* Rafael Fulanetti

*Coordenação de produção* Ana Cristina Garcia

*Produção editorial* Juliana Midori Horie

*Diagramação* Guilherme Salvador

*Revisão de texto* Equipe editorial

*Capa* Equipe editorial

*Imagem da capa* NASA, ESA, and N. Da Rio (The University of Virginia), G. De Marchi (European Space Agency – ESTEC), and D. Gouliermis (Universitat Heidelberg); Processing: Gladys Kober (NASA/Catholic University of America)

**Blucher**

Rua Pedroso Alvarenga, 1245, 4° andar
04531-934 – São Paulo – SP – Brasil
Tel.: 55 11 3078-5366
**contato@blucher.com.br**
www.blucher.com.br

Segundo o Novo Acordo Ortográfico, conforme 6. ed. do *Vocabulário Ortográfico da Língua Portuguesa*, Academia Brasileira de Letras, julho de 2021.

É proibida a reprodução total ou parcial por quaisquer meios sem autorização escrita da editora.

Todos os direitos reservados pela Editora Edgard Blücher Ltda.

Dados Internacionais de Catalogação na Publicação (CIP)
Heytor Diniz Teixeira, CRB-8/10570

Biswas, Anjanava
Construindo sistemas de IA agêntica : crie agentes de IA inteligentes e autônomos que podem raciocinar, planejar e se adaptar / Anjanava Biswas, Wrick Talukdar. – São Paulo : Blucher, 2026.

294 p. : il.

Bibliografia
ISBN 978-85-212-3307-7 (Impresso)
ISBN 978-85-212-3308-4 (Eletrônico – Epub)
ISBN 978-85-212-3305-3 (Eletrônico – PDF)

Título original: *Building Agentic AI Systems: Create intelligent, autonomous AI agents that can reason, plan, and adapt*

1. Inteligência artificial (IA). 2. IA generativa. 3. Sistemas agênticos generativos. 4. Autoencoders variacionais. 5. Redes generativas adversariais. 6. Arquiteturas de agentes inteligentes. 7. Desenvolvimento de agentes inteligentes. I. Talukdar, Wrick. II. Título.

CDU 004.8

Índice para catálogo sistemático:
1. Inteligência artificial CDU 004.8

# Apresentação

A inteligência artificial sempre foi movida por uma aspiração profunda: criar sistemas que não apenas executem tarefas predefinidas, mas também possam raciocinar, adaptar-se e agir com autonomia. Com a ascensão da IA generativa e de sistemas agênticos, estamos testemunhando um momento transformador, no qual agentes artificiais podem refletir sobre suas experiências, traçar estratégias de forma eficaz e colaborar de maneira fluida. Este livro chega precisamente neste momento crucial, oferecendo uma exploração abrangente deste campo emergente e de grande impacto.

Como alguém com décadas de experiência em IA, abrangendo pesquisa fundamental e desenvolvimento de produto como líder técnico na Microsoft Research e em startups de IA apoiadas por capital de risco, estive profundamente envolvido em pesquisa de IA de alto impacto e na implantação de produtórios de IA em larga escala que atendem a milhões de usuários, tanto no segmento de consumo quanto no corporativo. Meu envolvimento no IEEE em diversas funções de liderança em IA também me proporcionou uma perspectiva única sobre a importância crítica do avanço ético e responsável da IA.

Conheci os autores, Wrick Talukdar e Anjanava Biswas, por meio do nosso trabalho colaborativo em iniciativas globais de IA do IEEE. Sua ampla expertise, demonstrada por suas funções técnicas de alto impacto na AWS, trabalho em IA e ML, bem como sua pesquisa acadêmica em IA, é muito impressionante. Tanto Anjanava quanto Wrick são líderes reconhecidos em IA generativa, trazendo expertise de classe mundial, dedicação e insights profundos oriundos de arquitetar, desenvolver e implantar plataformas de IA de ponta em escala global. Sua rica compreensão das complexidades técnicas e das considerações éticas os qualifica de forma única para escrever este importante livro.

Organizado em três partes, o livro inicialmente estabelece uma base teórica sólida ao definir e desmistificar claramente a IA generativa e as características essenciais de sistemas agênticos.

Os leitores são guiados por conceitos centrais, de autoencoders variacionais e redes generativas adversariais a arquiteturas sofisticadas de agentes, permitindo que tanto iniciantes quanto profissionais experientes naveguem com confiança por esse cenário complexo.

Em sua segunda parte, o livro desloca o foco para técnicas práticas necessárias ao desenvolvimento de agentes inteligentes capazes de autoconsciência, adaptabilidade e tomada de decisão colaborativa. Os capítulos sobre reflexão, introspecção, planejamento e colaboração multiagente fornecem metodologias e insights inestimáveis, capacitando profissionais a projetar sistemas que aprendem dinamicamente e melhoram continuamente.

A seção final aborda de forma ponderada as implicações mais amplas da implantação de sistemas agênticos generativos em contextos do mundo real. Ao enfatizar transparência, explicabilidade, segurança e governança ética, este livro ressalta as responsabilidades essenciais associadas a tecnologias avançadas de IA. Ele fornece uma estrutura crítica para fomentar a confiança e garantir que essas poderosas ferramentas estejam alinhadas com valores sociais e padrões éticos.

Mais do que uma referência técnica, este livro serve como um guia essencial para moldar o futuro da IA generativa e de agentes inteligentes. Ele convida pesquisadores, líderes do setor de atuação (indústria) e inovadores a participar ativamente da elaboração do próximo Capítulo da IA, que avance de forma responsável as fronteiras tecnológicas enquanto enriquece a sociedade.

Endosso plenamente este trabalho oportuno e perspicaz. Encorajo você a se aprofundar nos capítulos que se seguem, assumindo seu papel de colaborador ativo na fronteira empolgante e em constante evolução da inteligência artificial.

*Matthew R. Scott,*

*Diretor de Tecnologia, Minset.ai*

*Primavera de 2025*

O mundo foi pego de surpresa quando o ChatGPT foi lançado em novembro de 2022. Foi um momento "aha" — de repente, parecia que você estava conversando com uma pessoa de verdade. As respostas não eram apenas fluentes; eram informativas e úteis. Desde então, a tecnologia evoluiu em um ritmo impressionante, com centenas de milhões de pessoas usando o ChatGPT e sistemas semelhantes como assistentes de escrita ou para encontrar respostas que não são facilmente encontradas por meio de buscas tradicionais na web.

No entanto, quando você pede a um assistente humano para reservar uma viagem, ele não apenas fala sobre isso — ele realmente reserva os voos e o hotel. Essa é a essência dos sistemas agênticos. **Modelos de linguagem de grande porte** (**LLMs**), como os usados no ChatGPT, podem ser reaproveitados não apenas para gerar texto, mas também para agir — executando tarefas por meio da geração de código, chamadas a APIs e interação com serviços web. Assim

como a World Wide Web passou de servir conteúdo estático para permitir aplicativos interativos, os sistemas agentes estão prestes a revolucionar a forma como interagimos com a IA.

Ao contrário de outros livros que se concentram principalmente na teoria por trás dos sistemas agênticos, este livro enfatiza a implementação no mundo real. Os autores fornecem exemplos de código práticos, casos de uso práticos e orientações passo a passo sobre como integrar fluxos de trabalho agênticos nas organizações para resolver desafios reais de negócios. Este livro serve como um recurso acessível e, ao mesmo tempo, abrangente para desenvolvedores de software, engenheiros de ML, líderes técnicos, diretores e outros tomadores de decisão que buscam se atualizar sobre sistemas agênticos.

Aqueles que já usaram assistentes de IA como o ChatGPT entendem seu valor como assistentes de escrita e programação, em que os usuários validam a saída antes de agir. No entanto, a IA às vezes pode gerar respostas com tanta confiança que os usuários podem, por engano, confiar nelas. Ao contrário dos humanos, que frequentemente atenuam suas afirmações com frases como "Não tenho certeza, mas acho que…", os sistemas de IA carecem de indicadores de incerteza nativos. Queremos confiar que nossos agentes de IA tomarão as ações corretas e nos perguntarão quando estiverem em dúvida. Este livro aprofunda questões cruciais de confiança, transparência, explicabilidade e confiabilidade, bem como considerações éticas como viés e segurança. Ele descreve como sistemas agênticos podem ser personalizados e tornados sensíveis ao contexto, e como podem planejar e reagir aos resultados de suas ações.

Como alguém que trabalha com inteligência artificial há décadas, acredito que este livro será um ótimo recurso para estudantes, pesquisadores e profissionais, traçando um caminho claro em uma era profundamente influenciada por sistemas inteligentes. Conheci Wrick Talukdar pela primeira vez em um painel de conferência do IEEE e fiquei imediatamente impressionado com sua capacidade de fazer a ponte entre teoria e prática. Em *Construindo sistemas de IA agêntica*, ele e Anjanava Biswas fazem exatamente isso: oferecem um guia prático para o futuro da automação impulsionada por IA.

*Dr. Alex Acero,*

*Membro da National Academy of Engineering,*

*IEEE Fellow*

# 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**

# Prefácio

*Construindo sistemas de IA agêntica* é projetado para fornecer tanto uma base teórica quanto orientações práticas sobre IA generativa e inteligência baseada em agentes. IA generativa e sistemas agênticos estão na vanguarda da próxima onda de IA, impulsionando automação, criatividade e tomada de decisão de maneiras antes inimagináveis. Ao permitir que máquinas gerem texto, imagens e até planos estratégicos enquanto raciocinam e se adaptam de forma autônoma, essas tecnologias estão transformando setores como saúde, finanças e robótica.

O livro começa apresentando a IA generativa, abrangendo modelos fundamentais como **Autoencoders Variacionais** (**VAEs**), **redes generativas adversariais** (**GANs**), e modelos autorregressivos. Exploramos suas aplicações em criação de conteúdo, design e pesquisa científica, ao mesmo tempo que abordamos as limitações e os desafios desses modelos.

Em seguida, mergulhamos no mundo dos sistemas agênticos, definindo conceitos como agência, autonomia e colaboração multiagente. Analisamos diferentes arquiteturas de agentes — deliberativas, reativas e híbridas — e exploramos como múltiplos agentes podem interagir, cooperar e se coordenar em direção a objetivos comuns.

Com as bases estabelecidas, avançamos para a implementação prática. Exploramos como agentes podem refletir sobre seus próprios processos de raciocínio, planejar e usar ferramentas externas de forma eficaz. Isso inclui técnicas práticas de metarraciocínio, autoexplicação, planejamento estratégico e coordenação multiagente. O livro também apresenta práticas recomendadas para projetar agentes de IA inteligentes e confiáveis, equilibrando autonomia com controle e garantindo um desenvolvimento de IA responsável e ético.

Para concluir, examinamos casos de uso e aplicações do mundo real em vários domínios, de NLP e robótica até suporte à decisão e otimização. Também exploramos confiança, transparência, mitigação de viés e segurança em IA — elementos-chave para garantir a confiabilidade de sistemas impulsionados por IA.

Ao longo deste livro, você encontrará exemplos de código, exercícios práticos e estratégias de implementação para ajudar a preencher a lacuna entre a teoria e a aplicação no mundo real. Seja você um profissional de IA, pesquisador, engenheiro ou líder de tecnologia, este livro o equipará com as habilidades e o conhecimento para construir agentes de IA autônomos, adaptativos e inteligentes que possam raciocinar, colaborar e evoluir.

Vamos embarcar nesta jornada juntos, moldando o futuro dos sistemas inteligentes — um agente de cada vez.

## Para quem é este livro

Este livro é destinado a profissionais de IA, desenvolvedores, pesquisadores, engenheiros e líderes de tecnologia que desejam compreender e construir agentes impulsionados por IA que apresentem autonomia, adaptabilidade e inteligência. Seja você um desenvolvedor que busca integrar modelos generativos a sistemas inteligentes ou um arquiteto de IA explorando capacidades agênticas avançadas, este livro fornecerá tanto fundamentos teóricos quanto estratégias práticas de implementação.

## O que este livro aborda

*Capítulo 1*, *Fundamentos de IA generativa*, apresenta a IA generativa, explicando seus conceitos centrais, vários tipos de modelos — incluindo VAEs, GANs e modelos autorregressivos —, aplicações no mundo real e desafios como viés, limitações e questões éticas.

*Capítulo 2*, *Princípios de Sistemas Agênticos*, define sistemas agênticos, abrangendo agência, autonomia e as características essenciais de agentes inteligentes, incluindo reatividade, proatividade e habilidade social. Também explora diferentes arquiteturas de agentes e colaboração multiagente.

*Capítulo 3*, *Componentes Essenciais de Agentes Inteligentes*, detalha elementos-chave de agentes inteligentes, incluindo representação do conhecimento, raciocínio, mecanismos de aprendizado, tomada de decisão e o papel da IA generativa no aprimoramento das capacidades dos agentes.

*Capítulo 4*, *Reflexão e Introspecção em Agentes*, explora como agentes inteligentes analisam seu raciocínio, aprendem com a experiência e melhoram a tomada de decisão usando técnicas como metarraciocínio, autoexplicação e automodelagem.

*Capítulo 5*, *Habilitando o Uso de Ferramentas e o Planejamento em Agentes*, discute como os agentes alavancam ferramentas externas, implementam algoritmos de planejamento e integram o uso de ferramentas com a tomada de decisão estratégica para melhorar a eficiência e o alcance de objetivos.

*Capítulo 6*, *Explorando a Abordagem de Coordenador, Executor e Delegador*, apresenta o modelo CWD para colaboração multiagente, explicando como os agentes assumem papéis especializados—coordenador, executor ou delegador—para otimizar a execução de tarefas e a alocação de recursos.

*Capítulo 7*, *Técnicas Eficazes de Projeto de Sistemas Agênticos*, abrange práticas recomendadas para projetar agentes inteligentes, incluindo instruções focadas, definição de guardrails e restrições, balanceamento entre autonomia e controle e garantia de transparência e responsabilização.

*Capítulo 8*, *Construindo Confiança em Sistemas de IA generativa*, examina técnicas para fomentar a confiança na IA, incluindo transparência, explicabilidade, tratamento de incerteza e viés e o projeto de sistemas de IA que sejam confiáveis e interpretáveis.

*Capítulo 9*, *Gerenciando Segurança e Considerações Éticas*, aborda os riscos e desafios da IA generativa, estratégias para garantir o desenvolvimento de IA responsável, diretrizes éticas e considerações de privacidade e segurança para implantações de IA.

*Capítulo 10*, *Casos de Uso Comuns e Aplicações*, apresenta aplicações do mundo real de IA generativa, abrangendo áreas como geração de conteúdo criativo, IA conversacional, robótica e sistemas de apoio à decisão.

*Capítulo 11*, *Conclusão e Perspectivas Futuras*, resume os principais conceitos abordados no livro, explora tendências emergentes em IA generativa e inteligência agêntica, discute **inteligência artificial geral** (**AGI**) e destaca desafios e oportunidades futuros no campo.

## Para aproveitar ao máximo este livro

Acompanhar será um pouco mais fácil se você tiver o seguinte:

- **Familiaridade com conceitos de IA e machine learning**: Embora o livro aborde princípios fundamentais, conhecimento prévio de IA/ML, deep learning e programação em Python ajudará você a compreender os tópicos mais avançados.
- **Prática aplicada**: Experimente os exemplos de código e frameworks fornecidos para construir IA generativa e sistemas de agentes. Configurar um ambiente de desenvolvimento local ou baseado em nuvem aprimorará sua experiência de aprendizado.
- **Pense criticamente sobre ética e segurança em IA**: Ao explorar IA generativa e agentes autônomos, considere as implicações de confiança, viés e design de IA responsável para construir sistemas inteligentes que estejam alinhados com diretrizes éticas.

| Software/hardware abordados no livro | Requisitos do sistema operacional |
|---|---|
| Python, Jupyter Notebooks e CrewAI | Windows, macOS, Linux |

## Aviso legal sobre imagens

Algumas imagens neste título são apresentadas para fins contextuais, e a legibilidade do gráfico não é crucial para a discussão.

## Convenções utilizadas

Há diversas convenções de texto utilizadas ao longo deste livro.

**Negrito**: Indica um novo termo, uma palavra importante ou palavras que você vê na tela. Por exemplo, palavras em menus ou caixas de diálogo aparecem em negrito. Aqui está um exemplo: "**Plano de onboarding personalizado**: Com base nas metas e necessidades identificadas, crie um plano de onboarding personalizado que descreva as etapas, os marcos e os cronogramas para alcançar os objetivos definidos."

> **Dicas ou observações importantes**
>
> Aparecem assim.

# Parte 1: Fundamentos de IA generativa e sistemas agênticos

Esta parte estabelece os conceitos fundamentais de IA generativa e de sistemas agênticos, fornecendo uma base teórica sólida para a compreensão e o desenvolvimento de agentes inteligentes e autônomos.

Esta parte contém os seguintes capítulos:

- *Capítulo 1*, *Fundamentos da IA generativa*
- *Capítulo 2*, *Princípios de sistemas agênticos*
- *Capítulo 3*, *Componentes essenciais de agentes inteligentes*

# 1
# Fundamentos de IA generativa

A IA generativa rapidamente se tornou uma tecnologia transformadora no campo da **inteligência artificial** (**IA**) e aprendizado de máquina, revolucionando processos criativos e a resolução de problemas em diversos setores e casos de uso. Ela está expandindo os limites da autonomia em sistemas inteligentes baseados em agentes. Neste capítulo, vamos abordar os fundamentos da IA generativa. Exploraremos o que é IA generativa juntamente com um breve histórico dos modelos de IA generativa. Em seguida, vamos discutir os diferentes tipos de modelos generativos, que incluem **autoencoders variacionais** (**VAEs**), **redes generativas adversariais** (**GANs**), autoregressivos e modelos transformadores. Em seguida, vamos nos aprofundar nas aplicações de IA generativa e encerrar o capítulo com uma breve discussão sobre algumas das limitações e desafios relacionados à IA generativa.

Abordaremos os seguintes tópicos neste capítulo introdutório, que ajudarão a estabelecer as bases da IA generativa à medida que exploramos mais a fundo as capacidades de agentes inteligentes autônomos impulsionados por IA generativa:

- Introdução à IA generativa
- Tipos de modelos de IA generativa
- Aplicações da IA generativa
- Desafios e limitações da IA generativa

Ao final deste capítulo, você terá adquirido uma compreensão abrangente da IA generativa, incluindo seus conceitos fundamentais, aplicações diversas e desafios atuais. Você também conhecerá o potencial e as limitações da tecnologia, com foco especial em seu papel crítico no avanço de agentes inteligentes e sistemas agênticos.

## Introdução à IA generativa

IA generativa refere-se a uma classe de tecnologias de IA que é capaz de gerar diversas formas de conteúdo, incluindo, mas não se limitando a, texto, imagens, áudio e vídeo. Esses sistemas de IA podem gerar novo conteúdo, com base em seus dados de treinamento e parâmetros de entrada, que geralmente incluem prompts de texto, mas também podem envolver outras formas de entrada, como imagens. A recente repercussão em torno da IA generativa decorre da simplicidade com que novas interfaces de usuário, impulsionadas por essa tecnologia de IA, podem criar texto, gráficos e vídeos de alta qualidade em segundos.

Em termos muito simples, a IA generativa trata de criar novos dados que se parecem com os dados a partir dos quais ela foi treinada. Em outras palavras, aprender os padrões, as estruturas e as distribuições subjacentes dos dados de entrada habilita um procedimento dentro do modelo que lhe permite gerar novos dados de maneira semelhante. Por exemplo, se treinado em um conjunto de dados de rostos humanos, um modelo de IA generativa será capaz de criar rostos completamente novos de pessoas que não existem de fato na vida real, mas que são muito realistas. Em essência, modelos de IA generativa funcionam aprendendo a distribuição de probabilidade de um conjunto de dados e, então, realizando amostragem a partir dessa distribuição para criar novas instâncias. Essa abordagem é diferente dos modelos discriminativos, cujo objetivo é aprender a distinguir classes de dados. Por exemplo, um modelo discriminativo poderia ser treinado para separar imagens de gatos e cães, mas um modelo generativo seria treinado para criar imagens completamente novas de gatos ou cães.

O conceito de IA generativa vem de muito tempo; na verdade, remonta à época em que o machine learning estava em sua infância. No entanto, foi apenas recentemente que esse campo veio à tona devido às redes neurais e às melhorias drásticas no poder computacional nos últimos 10 anos. Nas primeiras tentativas de modelagem generativa—modelos de mistura Gaussiana e modelos ocultos de Markov—muitas técnicas estatísticas mais simples estavam envolvidas. Métodos de modelagem de distribuições de dados arbitrariamente complexas têm uma compatibilidade potencial muito maior com os tipos de métodos que anteriormente eram dominantes, mas que recentemente foram superados pelo deep learning. O deep learning foi, portanto, um divisor de águas para o campo da IA generativa. Modelos, como VAEs, que começaram a surgir por volta do início da década de 2010, foram capazes de fazer alavancagem de redes neurais profundas para modelar distribuições de dados mais complexas. Por volta desse mesmo período, GANs foram propostas; pela primeira vez, ofereceram uma nova forma de duas redes neurais trabalharem juntas para criar dados, usando princípios da teoria dos jogos. De modo geral, esses avanços deram saltos em direção às possibilidades de tornar real a criação de dados de alta qualidade e realista.

Isso pode levar à inovação na maioria dos campos em diversos setores, como saúde, finanças, educação, viagens e hotelaria, manufatura e mais. A IA generativa nas indústrias criativas ajuda artistas e designers a pensar fora da caixa, auxiliando na concepção de conteúdo novo e inovador. Ela abre caminhos para novas rotas de descoberta de fármacos e medicina personalizada. No mundo dos negócios, a IA generativa atua fornecendo engajamento de clientes eficiente por meio de conteúdo e experiências personalizadas. O outro ponto envolve

questionar a própria postura em relação à criatividade e à originalidade: quando as máquinas conseguem produzir em massa material muito semelhante, a pessoa é automaticamente levada a pensar sobre o que é criatividade, quem é realmente o artista e quais deveriam ser os parâmetros éticos para conteúdo criado por IA.

Tendo entendido o que é IA generativa e seu breve histórico, vamos explorar os diferentes tipos de modelos de IA generativa.

# Tipos de modelos de IA generativa

IA generativa é um interessante domínio da IA que trata da geração de novos dados sintéticos por meio do aprendizado de padrões a partir de conjuntos de dados já existentes, visando gerar saídas que compartilhem propriedades e características estatísticas semelhantes aos dados de treinamento. Aqui está uma visão geral de alguns dos tipos mais proeminentes de modelos generativos: VAEs, GANs e modelos autorregressivos.

## VAEs

Um dos modelos generativos mais populares é o VAE. A ideia central por trás do VAE consiste em aprender um mapeamento probabilístico entre dados e um espaço latente, e vice-versa. Isso significa aprender como converter dados reais em uma representação simplificada (como uma forma comprimida) e depois em dados que pareçam reais, novamente. VAEs são projetados para garantir uma alta verossimilhança dos dados, ao mesmo tempo em que preservam um espaço latente bem estruturado para viabilizar a geração de novas amostras de dados semelhantes aos dados de treinamento. Algumas das variantes de VAE mais comuns são as seguintes:

- **VAE**: O modelo básico que comprime e reconstrói dados. Ele fornece um framework para gerar novas amostras a partir do espaço latente aprendido. O VAE é um modelo generativo que aprende a comprimir e reconstruir dados enquanto também aprende uma distribuição de probabilidade do espaço latente. Em termos mais simples, um VAE é como um algoritmo inteligente que aprende a entender e recriar dados. Imagine-o como um artista habilidoso que consegue comprimir uma pintura detalhada em um esboço simples (codificação) e então recriar a pintura completa a partir desse esboço (decodificação). O que torna os VAEs especiais é que eles não apenas copiam dados, mas também aprendem a essência deles, permitindo criar novos dados semelhantes. Isso é como um artista aprendendo não apenas a copiar pinturas específicas, mas também a entender tão bem o estilo que consegue pintar novas obras originais nesse estilo. Na prática, VAEs têm sido usados de maneiras muito criativas. Por exemplo, na descoberta de fármacos, VAEs têm sido usados para gerar novas estruturas moleculares. Pesquisadores da AstraZeneca usaram VAEs para explorar o espaço químico e propor novas moléculas semelhantes a fármacos com propriedades desejadas, potencialmente acelerando o processo de desenvolvimento de medicamentos [1].
- **Beta-VAE**: Uma extensão do VAE que introduz um hiperparâmetro para controlar explicitamente o compromisso entre a qualidade de reconstrução e o desemaranhamento

do espaço latente. O modelo promove a criação de representações mais representações interpretáveis e desemaranhadas. Em termos mais simples, o Beta-VAE é como uma versão mais flexível do VAE original. Ele permite que pesquisadores ajustem o quanto o modelo foca em recriar detalhes exatos versus entender as características subjacentes dos dados. Isso é semelhante a ensinar um estudante de arte não apenas a copiar uma pintura, mas também a entender e separar os elementos-chave, como cor, forma e estilo. Essa capacidade de *desemaranhar* características torna o Beta-VAE particularmente útil em áreas como visão computacional e robótica. Por exemplo, pesquisadores usaram o Beta-VAE para ensinar robôs a entender melhor objetos. Ao aprender a separar características como tamanho, cor e posição, robôs podem reconhecer e manipular objetos mais facilmente em diferentes situações, tornando-os mais adaptáveis e eficientes em várias tarefas [2].

- **Autoencoder Variacional Condicional** (**CVAE**): Uma variante que condiciona o processo de geração a alguma informação extra, tipicamente rótulos de classe, produzindo não apenas dados semelhantes ao conjunto de treinamento, mas também amostras que podem então ser extraídas de classes específicas . Pense no CVAE como uma versão mais controlada do VAE, como um artista que pode pintar em diferentes estilos sob demanda. Em vez de apenas criar novas pinturas aleatórias, esse artista pode ser instruído a "*Pinte uma paisagem*" ou "*Pinte um retrato*" e eles criarão uma nova arte nessa categoria específica. Esse controle adicional torna o CVAE muito útil em aplicações práticas. Por exemplo, na área de jogos eletrônicos, o CVAE tem sido usado para desenvolvimento de jogos e geração procedural de conteúdo, com a criação de elementos de jogo como design de personagens, layouts de níveis, música e efeitos sonoros, e assim por diante. Ao fornecer diferentes condições como "*Crie um nível de floresta*" ou "*Crie um nível de deserto*", o CVAE pode produzir uma ampla variedade de ambientes de jogo, economizando tempo para os designers e aprimorando a experiência do jogador com mundos de jogo mais diversos e interessantes [3].

## GANs

Uma GAN é basicamente composta por duas redes neurais: um **gerador** e um **discriminador**. O gerador gera amostras de dados sintéticos; a outra treinada rede neural deveria então ser capaz de distinguir entre amostras reais e criadas. Durante o treinamento dessas redes, elas são treinadas juntas de forma antagônica: o gerador tenta enganar o discriminador, enquanto o discriminador tenta classificar corretamente dados reais versus falsos. Nessa competição, o gerador fica cada vez melhor em gerar dados falsos. A seguir estão alguns dos diferentes tipos de GANs:

- **GAN**: O modelo básico no qual o gerador e o discriminador são treinados adversarialmente; é o modelo base para a maioria das inovações em modelagem generativa. Como explicado anteriormente, você pode imaginar a GAN como um jogo entre dois jogadores — um falsificador (gerador) tentando criar dinheiro falso, e um detetive (discriminador) tentando identificar as falsificações. À medida que competem, ambos ficam melhores em seus trabalhos, o que significa que o falsificador fica cada vez melhor

em criar dinheiro falso. Essa configuração inteligente permite que GANs criem dados falsos incrivelmente realistas, como imagens ou sons [4].

- **GAN convolucional profunda** (**DCGAN**): Este é um refinamento do modelo base de GAN com redes neurais convolucionais profundas; no momento, é uma das melhores arquiteturas para gerar imagens de alta qualidade. Pense na DCGAN como um artista mais sofisticado em comparação com a GAN básica. É como fazer um upgrade de ferramentas simples de esboço para um estúdio completo de arte digital, permitindo a criação de imagens muito mais detalhadas e realistas. As DCGANs são particularmente boas em entender e recriar padrões complexos em imagens [5].
- **GAN de Wasserstein** (**WGAN**): Propõe uma função de perda diferente com a distância de Wasserstein. Isso remove problemas durante o treinamento com uma GAN, tornando-o mais estável e melhorando a qualidade das amostras. Isso ajuda a medir melhor a distância entre duas distribuições de probabilidade. A WGAN é como dar ao artista (gerador) e ao crítico (discriminador) em uma GAN uma maneira melhor de se comunicar e avaliar o trabalho um do outro. Em vez de apenas dizer "*bom*" ou "*ruim*", eles agora podem dar um feedback mais detalhado, como "*Você está chegando mais perto*" ou "*Você está muito longe*". Isso leva a uma melhoria mais consistente e a resultados de alta qualidade. Em aplicações práticas, WGANs têm sido usadas na geração de imagens médicas sintéticas para fins de treinamento. Isso ajuda a criar conjuntos de dados maiores e mais diversos para o treinamento de sistemas de IA diagnósticos, potencialmente melhorando sua acurácia na detecção de doenças a partir de tomografias e raios X [6].
- **StyleGAN**: Gera imagens realistas de alta qualidade. O modelo é especialmente bom no tratamento separado de estilo e conteúdo. As evoluções trazidas pelas introduções do StyleGAN2 e do StyleGAN3 continuam voltadas a melhorar a fidelidade e o realismo das imagens. Pense no StyleGAN como um artista digital avançado que não apenas pode criar imagens realistas, mas também misturar e combinar diferentes estilos e conteúdos. É como ter um pintor que pode pegar o estilo de Van Gogh e aplicá-lo a uma paisagem urbana moderna. Essa flexibilidade torna o StyleGAN incrivelmente versátil [7].

## Modelos autorregressivos e arquitetura Transformer

Modelos autoregressivos geram seus pontos de dados um de cada vez, condicionando cada um de seus pontos de dados ao anterior. Surpreendentemente, isso acaba se mostrando muito útil em tarefas nas quais a sequência ou a estrutura dos dados é relevante, como geração de texto e geração de imagens. A arquitetura Transformer, introduzida no artigo *Attention Is All You Need* [8], é uma arquitetura de modelo que revolucionou muitas tarefas sequenciais de dados, especialmente em **processamento de linguagem natural** (**NLP**). Seus componentes-chave são demonstrados na *Figura 1.1*:

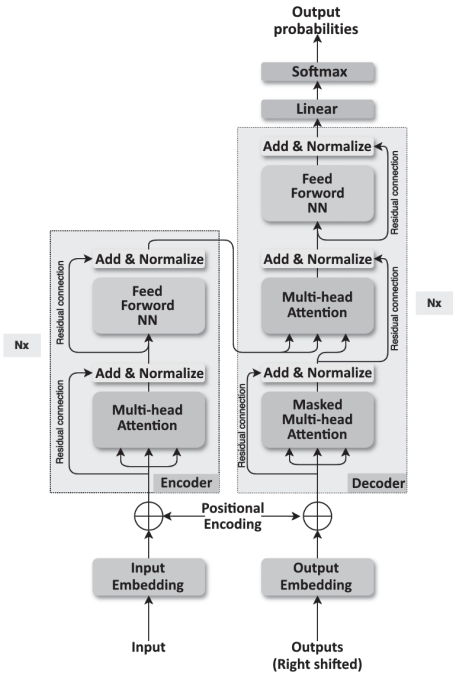



Figura 1.1 – A arquitetura Transformer

Estes principais componentes incluem o seguinte:

- **Mecanismo de autoatenção**: Uma técnica computacional que permite ao modelo concentrar-se dinamicamente em diferentes partes da entrada ao processar cada elemento
- **Atenção multi-cabeça**: Múltiplos mecanismos de atenção executando em paralelo, permitindo que o modelo se concentre em diferentes aspectos da entrada simultaneamente
- **Codificação posicional**: Adiciona informação sobre a posição de cada elemento na sequência
- **Redes neurais feed-forward**: Processam a saída das camadas de atenção
- **Normalização de camada e conexões residuais**: Melhoram a estabilidade do treinamento e o fluxo de informação

Transformers podem ser utilizados tanto em configurações autorregressivas quanto não autorregressivas, o que os torna versáteis para diversas tarefas. A seguir estão alguns exemplos:

- **PixelCNN**: Gera imagens de forma autorregressiva, pixel a pixel, aplicando condicionamento a cada pixel com base nos anteriormente gerados. Este modelo é muito eficaz na geração de imagens de alta qualidade, com detalhes finos [9].

- **PixelSNAIL**: Um aprimoramento em relação ao PixelCNN que adiciona mecanismos de atenção ao modelo para capturar dependências fortes entre pixels e, portanto, uma melhor qualidade de imagem [10].
- **GPT**: Sigla para **Transformer Generativo Pré-treinado**, é uma linha de modelos especializados em geração de texto—isto é, prevendo quais palavras devem vir a seguir. Entre eles estão GPT-2, GPT-3 e GPT-4, representando um grande salto na geração de texto coerente e contextualmente relevante [11] [12].
- **BERT**: Ao contrário do GPT, o BERT, que é a sigla para **Bidirectional Encoder Representations from Transformers**, é projetado para entender o contexto em ambas as direções no texto. Ele usa a parte do codificador do Transformer e é tipicamente usado para tarefas que exigem a compreensão de toda a entrada de uma só vez, em vez de gerar texto de forma autorregressiva [13].
- **Text-To-Text Transfer Transformer** (**T5**): Este modelo enquadra todas as tarefas de NLP em um formato texto-para-texto. Ele usa a arquitetura Transformer completa (tanto o codificador quanto o decodificador) e pode lidar com diversas tarefas de geração de texto [14].

Embora dependa da tarefa, esses modelos demonstram como a arquitetura Transformer pode ser adaptada tanto para tarefas autorregressivas (como GPT) quanto para não autorregressivas (como BERT), evidenciando sua versatilidade no tratamento de dados sequenciais. Os pontos fortes variam entre modelos generativos. Essa classe de modelos vai desde a geração de imagens fotorrealistas até texto coerente e amostras de dados totalmente novas.

Com base no sucesso de modelos como GPT, pesquisadores escalaram essas arquiteturas para criar **modelos de linguagem de grande porte** (**LLMs**). Esses modelos, frequentemente treinados em grandes volumes de texto aberto da web e outros dados de texto livremente disponíveis, demonstraram proficiência excepcional em compreender e gerar texto semelhante ao humano em diversas aplicações, evidenciando sua versatilidade e habilidades linguísticas avançadas. Exemplos de LLMs incluem GPT-3, GPT-4, PaLM e BERT-large. Esses modelos ampliaram os limites do que é possível em NLP, mostrando proficiência em tarefas que vão de resposta a perguntas e resumização até escrita criativa e geração de código. A seguir está uma visão geral dos tipos comuns de LLMs:

- **LLMs autorregressivos**: Esses modelos de linguagem são tipicamente capazes de gerar texto sequencialmente, isto é, um token por vez. Algumas das tarefas comuns para as quais esses tipos de modelos são usados incluem geração de texto, tarefas de completação de texto e escrita criativa. Alguns exemplos populares de LLMs autorregressivos são a série GPT (GPT-3, GPT-4) e PaLM.
- **LLMs do tipo apenas codificador**: Esses modelos se especializam em **compreensão de linguagem natural** (**NLU**) tarefas, que envolvem analisar e compreender entrada de texto sem gerar um novo texto. LLMs do tipo apenas codificador, como BERT e suas variantes incluindo RoBERTa, concentram-se em compreender o contexto e o significado do texto de entrada. Esses modelos processam a sequência de entrada

inteira simultaneamente, permitindo capturar o contexto e ajudando em tarefas como classificação de texto, **reconhecimento de entidades nomeadas** (**NER**), e análise de sentimento.

- **LLMs do tipo codificador-decodificador**: Esta é uma combinação de arquitetura de codificador e decodificador em que os modelos não apenas conseguem entender o contexto do texto de entrada (NLU), mas também são capazes de gerar saída textual. O modelo T5 discutido anteriormente é um exemplo muito popular de um LLM do tipo codificador-decodificador, juntamente com modelos como BERT. Esses modelos são capazes de tradução de linguagem, resumização e resposta a perguntas.
- **LLMs multimodais**: Multimodalidade é um conceito em que um modelo de IA pode não apenas lidar com texto, mas também outras modalidades, como imagem, vídeo e áudio. LLMs multimodais podem processar e gerar conteúdo em várias modalidades, como texto, imagem, áudio, vídeo e qualquer combinação delas. Esses modelos são bastante recentes por natureza, e avanços recentes em capacidade de computação tornaram possível treinar LLMs multimodais. Alguns dos LLMs multimodais mais conhecidos são DALL-E, Stable Diffusion, Flamingo, GPT-4 e LlaVa.
- **LLMs ajustados por instruções**: LLMs são frequentemente treinados em um grande corpus de texto da web ou outros textos abertamente disponíveis. Assim, em certos cenários, esses modelos podem não ter um desempenho eficaz em tarefas que usam dados que o modelo nunca viu antes e, portanto, podem falhar em seguir a intenção do usuário. Por exemplo, um manual de fabricação de um instrumento proprietário pode conter linguagem proprietária específica de domínio, à qual o modelo pode não ter sido exposto durante o treinamento. Nesses casos, LLMs podem passar por ajuste fino usando os dados específicos de domínio para seguir instruções e prompts específicos (isto é, texto de entrada para o modelo). Esse ajuste fino especial permite que o modelo execute tarefas específicas de domínio nas quais o modelo foi ajustado, além de suas outras capacidades. Pense nisso como personalizar um modelo para ser bom em realizar uma tarefa que seja significativa para o seu caso de uso. Um dos exemplos comuns de tal modelo é o InstructGPT, em que pesquisadores realizaram ajuste fino do modelo GPT-3 com instruções específicas. Esses tipos de modelos são usados em uma infinidade de tarefas diferentes, como chatbots, assistentes virtuais e aplicações específicas de tarefa.
- **LLMs específicos de domínio**: Como mencionado anteriormente, embora LLMs sejam frequentemente treinados em um grande conjunto de dados abertos da web, geralmente disponíveis, eles podem não ter um desempenho muito bom em ambientes específicos de domínio. Embora LLMs ajustados por instruções sejam um caminho para fazer com que esses modelos de uso geral funcionem para tarefas específicas de domínio, pesquisadores inovaram no treinamento de LLMs específicos de domínio desde o zero. Isso significa que esses modelos foram pré-treinados em conhecimento de domínio altamente especializado. Por exemplo, o modelo BioBERT é treinado em dados médicos hiperespecializados, ao passo que o LegalBERT é treinado em documentos jurídicos. Esses modelos são úteis para tarefas especializadas em áreas específicas, como medicina, direito e finanças.

## Agentes de IA baseados em LLMs

Cada um dos tipos de LLM discutidos anteriormente tem seus próprios pontos fortes e fracos e são adequados para diferentes aplicações. No entanto, nosso foco neste livro será aprofundar-se nas capacidades de agentes de LLM e explorar como sistemas avançados podem ser construídos utilizando-as. É importante observar que agentes de LLM não se encaixam perfeitamente em uma única categoria, pois representam uma aplicação avançada da tecnologia de LLM, em vez de um tipo específico de arquitetura de modelo. Agentes são tipicamente construídos usando LLMs ajustados por instruções como sua base, frequentemente combinados com componentes e capacidades adicionais, como você aprenderá adiante neste livro.

Vamos analisar um exemplo simples para demonstrar as capacidades de um sistema baseado em agentes impulsionado por um LLM. Nosso exemplo é um site de viagens que possui um **chatbot assistente de reservas de voos,** que é uma interface conversacional construída com um LLM. Ele pode ajudar clientes a reservar passagens de viagem por meio de conversas semelhantes às humanas. No backend, o chatbot pode executar várias tarefas à medida que a conversa avança. Veja como uma conversa de exemplo pode ser para um cliente que está procurando reservar uma passagem aérea:

| Assistente de Reservas de Voos | |
|---|---|
|  | Olá, quero reservar um voo. |
|  | Olá! Ficarei feliz em ajudar você a reservar um voo. Para começar, poderia me informar sua cidade de partida e o destino? |
|  | Quero ir de San Diego para SFO. |
|  | Ótimo, obrigado por fornecer essa informação. Você deseja voar de San Diego (SAN) para o San Francisco International Airport (SFO). Para ajudar a encontrar as melhores opções de voo, precisarei de mais alguns detalhes.<br>1. Qual é a sua data de viagem preferida?<br>2. Esta é uma viagem só de ida ou ida e volta?<br>3. Quantos passageiros irão viajar? |
|  | Quero ir na próxima sexta-feira e voltar no domingo. Apenas eu viajando. |
|  | Obrigado pela informação adicional. Deixe-me resumir o que tenho até agora:<br>– Partida: San Diego (SAN) para San Francisco (SFO)<br>– Data de ida: Próxima sexta-feira – 08/23/2024<br>– Data de volta: Domingo – 08/25/2024<br>– Passageiros: 1 adulto |

| | |
|---|---|
| | Antes de pesquisar os voos, tenho mais algumas perguntas:<br><br>1. Você tem um horário do dia preferido para seus voos (manhã, tarde ou noite)?<br><br>2. Você tem uma companhia aérea de preferência ou está aberto a todas as companhias?<br><br>3. Qual classe de serviço você prefere (Econômica, Econômica Premium, Business ou Primeira Classe)? |
|  | Prefiro voos pela manhã. Sem preferência de companhia aérea. Classe Econômica está bom. |
|  | Obrigado por esses detalhes. Vou buscar voos pela manhã na classe Econômica em todas as companhias aéreas para a sua viagem. Dê-me um momento para encontrar as melhores opções para você.<br><br>[procurando voos... por favor, aguarde....] |
|  | Encontrei algumas opções para a sua viagem. Aqui estão os melhores voos pela manhã:<br><br>Ida (SAN para SFO) - Próxima sexta-feira:<br><br>1. United Airlines UA1234: Parte às 7:15 AM, Chega às 8:45 AM<br><br>2. Alaska Airlines AS5678: Parte às 8:30 AM, Chega às 10:00 AM<br><br>Volta (SFO para SAN) - Domingo:<br><br>1. Delta Air Lines DL9876: Parte às 9:00 AM, Chega às 10:30 AM<br><br>2. Southwest Airlines WN5432: Parte às 10:15 AM, Chega às 11:45 AM<br><br>O preço total da viagem de ida e volta, incluindo taxas e impostos, é:<br><br>– Opção 1 (United/Alaska): $325<br><br>– Opção 2 (Delta/Southwest): $310<br><br>Qual opção você prefere? Ou gostaria que eu procurasse voos diferentes? |
|  | A segunda opção parece boa. Pode reservar essa para mim? |
|  | Certamente! Ficarei feliz em reservar a segunda opção para você. Apenas para confirmar, vou reservar o seguinte itinerário:<br><br>Ida: Alaska Airlines AS5678<br><br>– Data: 08/23/2024<br><br>– Partida de San Diego às 8:30 AM, Chegada em SFO às 10:00 AM |

| | |
|---|---|
| | Volta: Southwest Airlines WN5432<br><br>– Data: 08/25/2024<br><br>– Partida de SFO às 10:15 AM, Chegada em San Diego às 11:45 AM<br><br>Preço total: $310<br><br>Para prosseguir com a reserva, vou precisar de alguma informação adicional:<br><br>1. Seu nome completo como aparece no seu documento de identificação emitido pelo governo<br><br>2. Sua data de nascimento<br><br>3. Seu endereço de e-mail (para envio da confirmação)<br><br>4. Seu número de telefone<br><br>5. Algum pedido especial (por exemplo, preferências de refeição, preferências de assento)<br><br>Além disso, observe que você precisará fornecer uma forma de pagamento para concluir a reserva. Você está pronto para prosseguir com esta informação? |
|  | Sim, estou pronto. Meu nome é John Smith, nascido em 05/15/1985. Meu e-mail é john.smith@email.com e meu número de telefone é 555-123-4567. Sem pedidos especiais. |
|  | Obrigado por fornecer essa informação, Sr. Smith. Inseri seus dados no sistema de reservas. Aqui está um resumo da informação que tenho:<br><br>– Nome: John Smith<br><br>– Data de Nascimento: 05/15/1985<br><br>– E-mail: john.smith@email.com<br><br>– Telefone: 555-123-4567<br><br>– Pedidos Especiais: Nenhum<br><br>A etapa final é processar o pagamento de $310 pelos seus voos. Por razões de segurança, não posso aceitar informação de cartão de crédito diretamente neste chat. Em vez disso, enviarei um link de pagamento seguro para o seu endereço de e-mail (john.smith@email.com).<br><br>Assim que receber o e-mail, siga o link para concluir o processo de pagamento. Após o pagamento ser processado com sucesso, você receberá outro e-mail com a confirmação do voo e os e-tickets.<br><br>Posso prosseguir enviando o link de pagamento para o seu e-mail? |

| | |
|---|---|
|  | Sim, está bem. Por favor, envie o link de pagamento. |
|  | Excelente, Sr. Smith. Iniciei o processo para enviar o link de pagamento seguro para o seu endereço de e-mail (john.smith@email.com). Você deve recebê-lo nos próximos minutos. Por favor, verifique sua caixa de entrada e a pasta de spam se não o vir imediatamente.<br><br>Assim que você concluir o pagamento, o sistema processará automaticamente sua reserva e enviará seu e-mail de confirmação com os e-tickets. Esse e-mail incluirá informação importante como:<br><br>1. Número de confirmação da reserva<br><br>2. Itinerário detalhado do voo<br><br>3. Instruções de check-in<br><br>4. Informação sobre franquia de bagagem<br><br>5. Informação de contato da companhia aérea<br><br>Depois de receber seu e-mail de confirmação, revise todos os detalhes para garantir que tudo esteja correto. Se notar qualquer problema ou tiver alguma dúvida, não hesite em nos contatar.<br><br>Há mais alguma coisa com a qual você precise de assistência em relação à sua viagem ou ao processo de reserva? |
|  | Não, é só isso. Obrigado pela ajuda! |
|  | De nada, Sr. Smith! Fico feliz por ter podido ajudá-lo a reservar seus voos de San Diego para San Francisco. Se precisar de mais assistência com esta reserva ou tiver quaisquer dúvidas sobre sua viagem, não hesite em entrar em contato. Estamos aqui para ajudar 24/7. Tenha uma ótima viagem a San Francisco na próxima semana! |

À primeira vista, este exemplo parece bastante direto. No entanto, há várias coisas que estão acontecendo em segundo plano. Primeiro, o cliente começa com um pedido vago para reservar um voo. O assistente de IA entende que, para prosseguir com essa solicitação, precisa de mais algumas informações para consultar voos disponíveis, como cidade de partida/chegada, datas de viagem e assim por diante.

À medida que a conversa avança, o assistente de IA solicita mais informações para conseguir concluir a reserva com sucesso. As informações que ele coleta em cada etapa são usadas como valor de parâmetro para chamar APIs relacionadas à reserva de voos no backend; por exemplo, cidade/datas de viagem/número de passageiros são necessários para consultar informações de voo disponíveis usando uma API (fictícia) chamada **flightLookup**, nome/telefone/e-mail do cliente são necessários para de fato efetuar a reserva usando uma API chamada b**ookFlight**, e assim por diante. *Figura 1.2* fornece uma visão geral das decisões que o modelo toma, do fluxo da interação e das ferramentas (funções de API) que ele chama para concluir cada etapa:

# CONSTRUINDO SISTEMAS DE IA AGÊNTICA

GPT/LLMS ■ INTELIGÊNCIA ARTIFICIAL ■ MACHINE LEARNING ■ DEEP LEARNING ■ DATA SCIENCE

Os sistemas de IA deixaram de apenas gerar texto para agir: raciocinar, planejar, usar ferramentas e decidir com autonomia. Este guia prático mostra como projetar, construir e implantar agentes de IA generativa capazes de refletir, adaptar-se e colaborar. Dos fundamentos (VAEs, GANs, modelos autorregressivos) às técnicas de planejamento e colaboração multiagente, com exemplos de código e casos reais – sem descuidar de confiança, segurança e ética. Ideal para desenvolvedores, engenheiros de ML, pesquisadores e líderes técnicos que querem ir além da teoria e criar agentes autônomos e confiáveis.

O que você vai encontrar:

- **Fundamentos da IA generativa: VAEs, GANs e modelos autorregressivos**
- **Princípios dos sistemas agênticos: autonomia, reatividade e proatividade**
- **Reflexão, planejamento e uso de ferramentas pelos agentes**
- **Colaboração multiagente com o modelo Coordenador–Trabalhador–Delegador**
- **Confiança, segurança e ética, com casos de uso reais**

"Diferentemente de outros livros, focados na teoria, esta obra enfatiza a implementação no mundo real, com exemplos de código, casos de uso e orientação passo a passo."

— Dr. Alex Acero, membro da National Academy of Engineering e IEEE Fellow

**Anjanava Biswas** é arquiteto sênior de soluções em IA, com mais de 17 anos de experiência em machine learning, IA generativa, PLN e nuvem. Fellow do BCS, do IET e do IETE e senior member do IEEE, publicou pesquisas e contribui com projetos open source. Já atuou na IBM e na Oracle.

**Wrick Talukdar** é líder de tecnologia em IA generativa na Amazon, com mais de duas décadas em IA, nuvem e liderança de produto. Senior member e Chair do IEEE, é referência em IA generativa, multimodalidade, PLN e visão computacional, e palestrante em conferências internacionais.

Conheça mais livros em:
tech.Blucher.com.br

ISBN 978-85-212-3307-7