

IA Generativa

Glossário do Curso

Bem-vindo! Este glossário em ordem alfabética contém muitos dos termos deste curso. Este glossário abrangente também inclui termos adicionais reconhecidos pela indústria que não são usados nos vídeos do curso. Esses termos são essenciais para você reconhecer ao trabalhar na indústria, participar de grupos de usuários e em outros programas de certificação.

Tempo estimado de leitura: 4 minutos

Termo	Definição
Bidirectional and Auto-Regressive Transformers (BART)	Modelo de linguagem grande (LLM) de sequência para sequência, que segue uma arquitetura de codificador-decodificador. Ele aproveita a codificação para compreensão contextual e a decodificação para gerar texto.
Bidirectional Encoder Representations from Transformers (BERT)	Modelo de linguagem grande (LLM), que utiliza uma arquitetura de transformador somente com codificador. É excepcional na compreensão do contexto de uma palavra dentro de uma frase, o que é crucial para tarefas sutis como análise de sentimento.
Data analysis learning with language model for generation and exploration (DALL-E)	Modelo de IA desenvolvido pela OpenAI, conhecido por gerar imagens complexas e criativas a partir de descrições textuais usando técnicas de aprendizado profundo.
Data loader	Componente de aplicação que permite o agrupamento e a mistura eficientes de dados, essenciais para o treinamento de redes neurais. Ele permite o pré-processamento em tempo real, o que otimiza o uso da memória.
Data set	Coleção de amostras de dados e seus rótulos
Diffusion model	Modelo probabilístico de IA generativa comumente usado para geração de imagens. Um modelo de difusão é treinado para gerar imagens aprendendo a remover ruído ou reconstruir exemplos de seus dados de treinamento que foram distorcidos além do reconhecimento.
Fine-tuning	Ajuste de um modelo pré-treinado para melhorar o desempenho em uma tarefa ou conjunto de dados específico. Isso faz com que o modelo gere conteúdo mais preciso e contextualmente relevante.
Generative adversarial network (GAN)	Modelo de IA generativa que pode gerar imagens a partir de vetores de entrada aleatórios ou imagens sementes. Consiste em um gerador e um discriminador, que trabalham em modo competitivo.
Generative AI	Modelos de aprendizado profundo que podem gerar texto, imagens e outros conteúdos de alta qualidade com base nos dados nos quais foram treinados. Esses modelos são desenvolvidos e treinados para entender padrões e estruturas dentro dos dados existentes e aplicar essa compreensão para produzir novos dados relevantes.
Generative pre-trained transformers (GPT)	Modelo de IA generativa baseado em arquitetura de transformador. Ele foi pré-treinado em grandes quantidades de dados textuais e pode prever e gerar sequências de texto com base nos padrões aprendidos de seus dados de treinamento.
Hugging Face	Plataforma que oferece uma biblioteca de código aberto com modelos pré-treinados e ferramentas para agilizar o processo de treinamento e ajuste fino de modelos de IA generativa.
Iterator	Um objeto que pode ser percorrido em um loop. Ele contém elementos que podem ser iterados e normalmente inclui dois métodos: iter e next.
LangChain	Framework de código aberto que ajuda a agilizar o desenvolvimento de aplicativos de IA usando modelos de linguagem grandes (LLMs).
Large language models (LLMs)	Modelos fundamentais que usam IA e aprendizado profundo com vastos conjuntos de dados para gerar texto, traduzir idiomas e criar vários tipos de conteúdo. Eles são chamados de modelos de linguagem grandes devido ao tamanho do conjunto de dados de treinamento e ao número de parâmetros.
Natural language processing (NLP)	Subcampo da inteligência artificial (IA) que lida com a interação de computadores e humanos em linguagem humana. Envolve a criação de algoritmos e modelos que ajudarão os computadores a entender e compreender a linguagem humana e gerar texto contextualmente relevante em linguagem humana.
NLTK	Biblioteca Python usada em processamento de linguagem natural (NLP) para tarefas como tokenização e processamento de texto.
Pydantic	Biblioteca Python que ajuda a simplificar o manuseio de dados. Pode ser usada para analisar e validar seus dados.
PyTorch	Framework dinâmico de aprendizado profundo desenvolvido pelo laboratório de Pesquisa em IA do Facebook. É uma biblioteca baseada em Python bem conhecida por sua facilidade de uso, flexibilidade e gráficos de computação dinâmicos.

Termo	Definição
Recurrent neural networks (or RNNs)	Redes neurais artificiais que usam dados sequenciais ou de séries temporais. RNNs são usadas para resolver problemas relacionados a dados com uma ordem natural ou dependências baseadas no tempo.
SentencePiece	Algoritmo de tokenização baseado em subpalavras que segmenta texto em partes gerenciáveis e atribui IDs únicos.
spaCy	Biblioteca de código aberto usada em processamento de linguagem natural. Fornece ferramentas para tarefas como tokenização e embeddings de palavras.
TensorFlow	Framework de aprendizado de máquina de código aberto. Fornece um conjunto de ferramentas e bibliotecas para facilitar o desenvolvimento e a implantação de modelos de aprendizado de máquina.
Text-to-Text Transfer Transformer (T5)	Modelo de linguagem grande baseado em transformador, que usa uma estrutura de texto para texto. Ele aproveita a codificação para compreensão contextual e a decodificação para gerar texto.
Tokenization	Divisão de texto em partes menores ou tokens. Os tokens ajudam um modelo de IA generativa a entender melhor o texto.
Tokenizer	Programa que divide o texto em tokens individuais.
Transformers	Modelos de aprendizado profundo que podem traduzir texto e fala em tempo quase real. Eles pegam dados, como palavras ou números, e os passam por diferentes camadas, com informações fluindo em uma direção.
Unigram	Algoritmo de tokenização baseado em subpalavras que divide o texto em partes menores. Começa com uma lista grande de possibilidades e gradualmente reduz com base em quão frequentemente aparecem no texto.
variational autoencoders (VAEs)	Modelo de IA generativa que opera em uma estrutura de codificador-decodificador. A rede do codificador primeiro comprime os dados de entrada em um espaço simplificado e abstrato que captura características essenciais. A rede do decodificador então usa essas informações condensadas para recriar os dados originais.
WaveNet	Modelo de IA generativa projetado para gerar conteúdo de áudio. Pode ser usado para tarefas como síntese de fala.
WordPiece	Algoritmo de tokenização baseado em subpalavras que avalia os benefícios e desvantagens de dividir e mesclar dois símbolos para garantir que suas decisões sejam valiosas.



Skills Network