

Modelos de IA generativa: unimodais x multimodais

Modelos Unimodais

São projetados para trabalhar com um único tipo de dado, **permitindo maior especialização e eficiência** nesse domínio específico. Confira alguns exemplos:

- Modelos de texto: Geram texto baseado em grandes volumes de dados textuais. Esses modelos são treinados em uma vasta base de conhecimento de texto, o que permite a eles gerarem respostas coerentes e contextualmente relevantes a partir de prompts textuais.
- Modelos de imagem: São usadas para reconhecimento e geração de imagens baseadas em padrões aprendidos, na qual possuem a possibilidade de criar imagens novas e até artísticas.
- Modelos de áudio: Modelos que são capazes de processar e gerar áudio, como síntese de voz ou reconhecimento de fala. Também são capazes de transcrever um áudio.

Vantagens dos modelos unimodais:

- Especialização: Focam um único tipo de dado, permitindo maior especialização e otimização. O que resulta em um desempenho superior para tarefas específicas.
- Simplicidade: São mais leves e podem ser usados em tecnologias menos robustas, que não demanda de muito processamento de dados.

Limitações dos modelos unimodais:

- Falta de contexto multimodal: Esses modelos não conseguem integrar múltiplos tipos de dados para fornecer uma visão mais holística e contextual. Por exemplo, um modelo de texto não conseguirá analisar um áudio e realizar uma transcrição textual.

Modelos Multimodais

É justamente o contrário dos modelos unimodais, dado que integram e processam múltiplos tipos de dados simultaneamente, como combinações de texto e imagem ou áudio e vídeo.

Vantagens dos modelos multimodais:

- **Compreensão contextual enriquecida:** A capacidade de integrar diferentes tipos de dados permite uma compreensão mais profunda e contextual de problemas, resultando em respostas mais abrangentes.
- **Ampliação das possibilidades de questionamento e resultados:** A multimodalidade aumenta a precisão e a relevância das respostas, especialmente quando prompts misturam texto, imagem e vídeo.

Limitações dos modelos multimodais:

- **Recursos computacionais:** Ao contrário dos modelos unimodais, há a necessidade de mais recursos computacionais e uma integração complexa de dados variados, ocasionalmente gerando respostas lentas, dependendo do processamento da tecnologia que está utilizando.
- **Custo e complexidade de treinamento:** Normalmente, são modelos pagos que conseguem obter melhores resultados, dado que demanda um investimento maior para se treinar o modelo de dados. Ou seja, em modelos *open source*, talvez não consigamos os melhores resultados.

Referência

MAGALHÃES, Marcelo Almeida. **Modelos para IA Generativa**. IBMEC, 2023.