

# Tudo o que se precisa saber sobre LLM

Adaptado por: Dr. Arnaldo de Carvalho Junior

Neste tutorial, será apresentado o que é um **LLM (Large Language Model)**, como ele é aplicado, quais ferramentas são utilizadas e suas aplicações práticas [1].

## 1. O que é um LLM?

Um modelo de linguagem natural ou *large language model* (LLM), é um tipo de modelo de linguagem de alta capacidade que utiliza algoritmos avançados de processamento de linguagem natural (*natural language processing* - NLP) para entender e gerar texto. Esses modelos são treinados em grandes quantidades de dados textuais, como livros, artigos, sites e muito mais, a fim de aprender a estrutura da linguagem e as relações semânticas entre palavras e frases.

Um dos LLMs mais conhecidos é o transformador generativo pré-treinado (*Generative Pre-trained Transformer* - GPT), que utiliza a arquitetura de transformadores (*transformers*) para processar o texto de forma eficiente e capturar dependências de longo alcance. Esses modelos são treinados em tarefas de previsão de palavras ou frases seguintes, o que lhes permite gerar texto coerente e de qualidade com base no contexto fornecido.

*LLM — “Language Model Learning” é alimentado com grandes quantidades de texto e aprende a prever a próxima palavra em uma frase, permitindo que gere respostas coerentes e relevantes às entradas do usuário.*

## 2. Onde um LLM é aplicado?

Os LLMs têm uma ampla gama de aplicações em diferentes áreas. Aqui estão alguns exemplos:

1. **Geração de texto:** Os LLMs podem ser usados para gerar automaticamente resumos de texto, artigos, redações e até mesmos diálogos fictícios.
2. **Tradução automática:** Esses modelos são frequentemente utilizados para traduzir texto entre diferentes idiomas, fornecendo resultados mais precisos e fluentes.

3. **Assistentes virtuais:** Muitos assistentes virtuais, chatbots e sistemas de perguntas e respostas são alimentados por LLMs para entender e gerar respostas coerentes em tempo real.
4. **Análise de sentimentos:** LLMs podem ser aplicados para analisar o sentimento em textos, identificando se eles são positivos, negativos ou neutros. Isso é útil em áreas como análise de mídias sociais e pesquisa de opinião.
5. **Recomendação de conteúdo:** Plataformas de streaming, comércio eletrônico e mídias sociais utilizam LLMs para recomendar conteúdo personalizado com base nos interesses e comportamentos dos usuários.
6. **Correção automática:** Os LLMs também são usados em teclados e editores de texto para fornecer sugestões de palavras e corrigir erros ortográficos e gramaticais.

Esses são apenas alguns exemplos das diversas aplicações dos LLMs. Sua versatilidade torna-os ferramentas poderosas para lidar com tarefas relacionadas ao processamento de linguagem natural, e não apenas texto.

A IA generativa está em andamento há algum tempo, mas nos últimos anos, todo o ecossistema de IA generativa passou por um desenvolvimento significativo. No entanto, para entender completamente o estado atual das coisas e apreciar todo o potencial da IA generativa, é importante mergulhar nos avanços feitos no campo do processamento de linguagem natural. O advento dos modelos de transformadores desempenhou um papel crucial a esse respeito. Por meio do uso de transformadores, a IA agora pode processar e gerar linguagem, imagens e vídeos e trabalhar em várias modalidades combinadas. A Figura 1 apresenta um cronograma da evolução recente da IA.

### 3. O Avanço das IA Generativas

Nos últimos anos, a humanidade tem testemunhado um avanço significativo no campo das Inteligências Artificiais (IAs) generativas, impulsionado em grande parte por modelos baseados em LLMs. Esses modelos, como o ChatGPT, Midjourney e Leonardo AI, estão ganhando destaque por sua capacidade de gerar texto coerente em diversas aplicações.

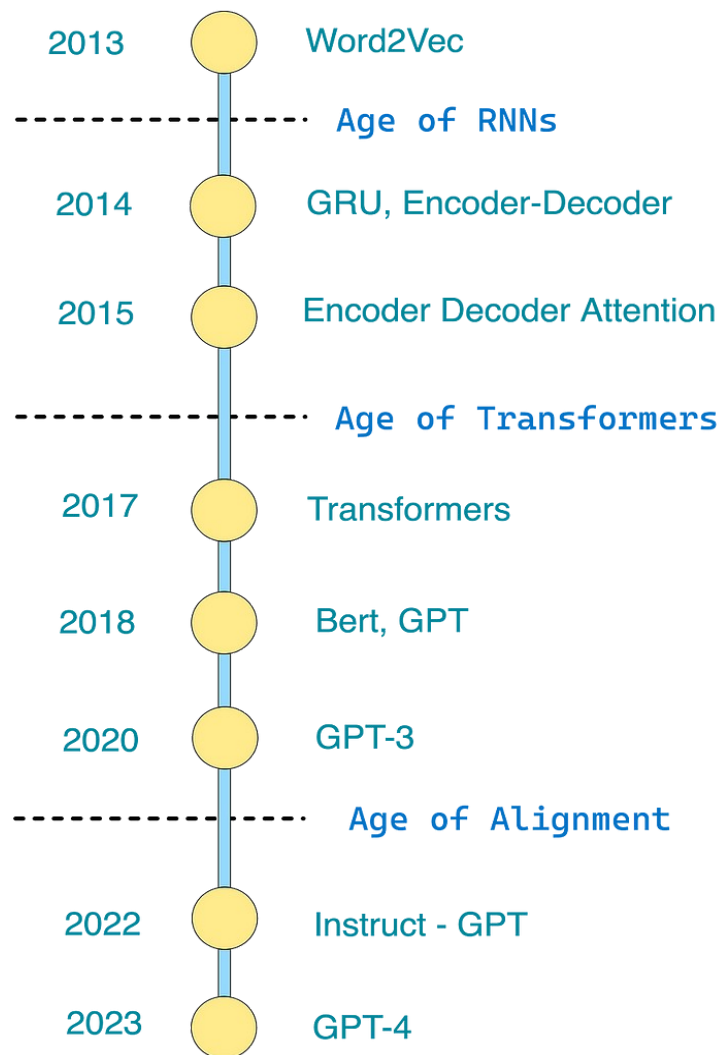


Figura 1 – Evolução da IA de 2013-2023.

A Figura 2 apresenta o cenário de aplicação da IA generativa.

### 3.1. ChatGPT

O ChatGPT é um exemplo proeminente de IA generativa baseada em LLMs. Criado pela Open AI, empresa que tem entre fundadores e investidores nomes como Elon Musk e Sam Altman, o ChatGPT é uma ferramenta de IA utilizada para gerar diálogos virtuais [2]. Ele utiliza a arquitetura GPT para entender e responder a perguntas, fornecendo uma experiência conversacional mais natural e sofisticada. Ele é treinado em grandes quantidades de dados textuais, permitindo que ele gere respostas relevantes e contextualmente coerentes.

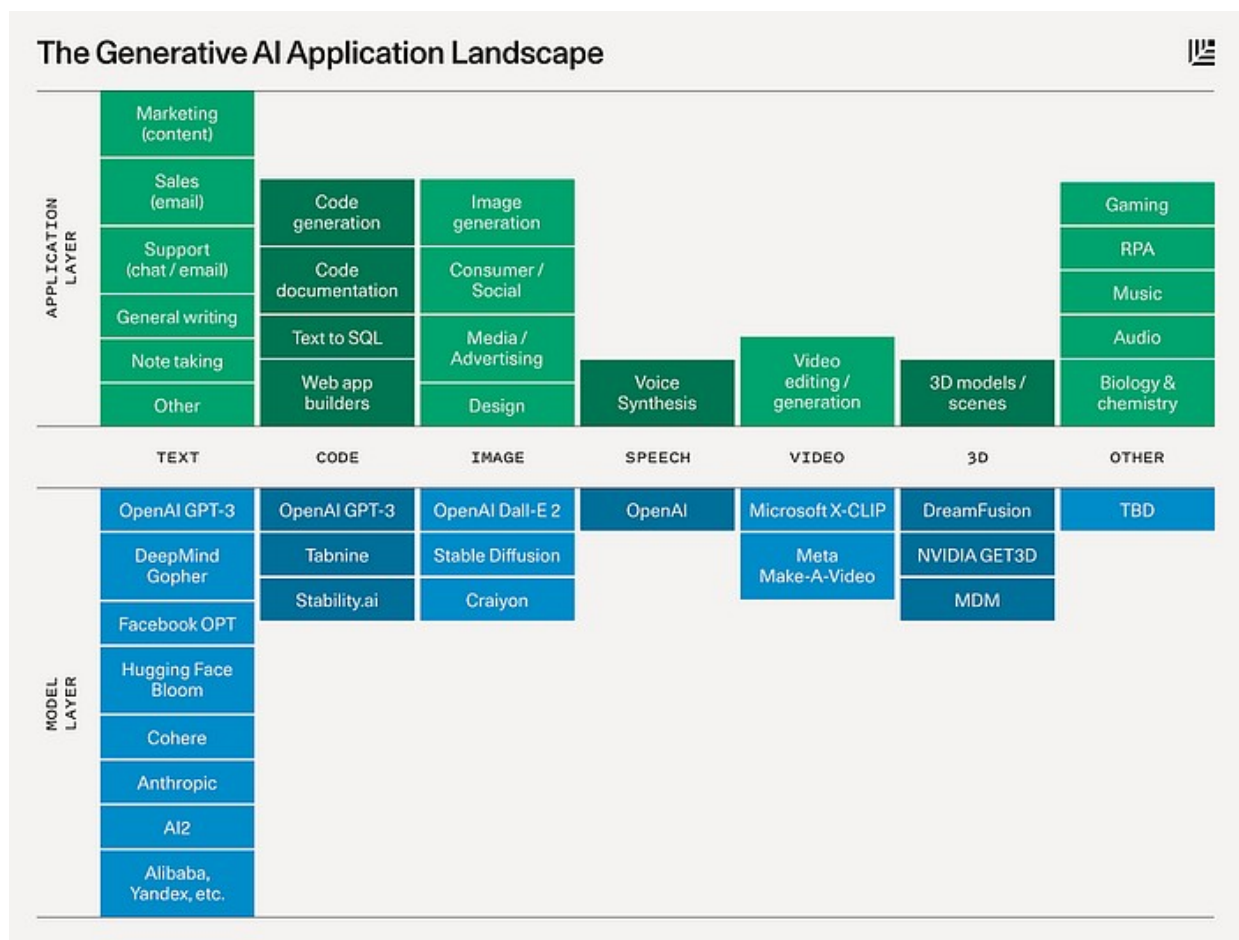


Figura 2 – O Cenário de Aplicação da IA generativa.  
 Fonte: Sequoia Capital's Report (Sep 2022)

Em dezembro de 2022, o ChatGPT foi lançado. A introdução deste chatbot AI marcou um ponto de viragem na história da tecnologia. Seu rápido crescimento superou o de qualquer outra plataforma na história e provocou uma revolução no campo das aplicações generativas de IA. Essa nova onda impactou praticamente todos os domínios e campos, desde saúde até finanças e entretenimento. Como resultado, as tecnologias generativas de IA têm muitos usos potenciais e seu impacto na sociedade ainda está sendo explorado. A Figura 3 apresenta uma comparação de usuários ativos por mês do ChatGPT para outras aplicações populares.

O ChatGPT tem sido usado em várias aplicações, desde assistentes virtuais e *chatbots* até sistemas de perguntas e respostas em tempo real. Ele é capaz de interagir com os usuários, compreender suas consultas e fornecer respostas relevantes e úteis, conforme Figura 4.

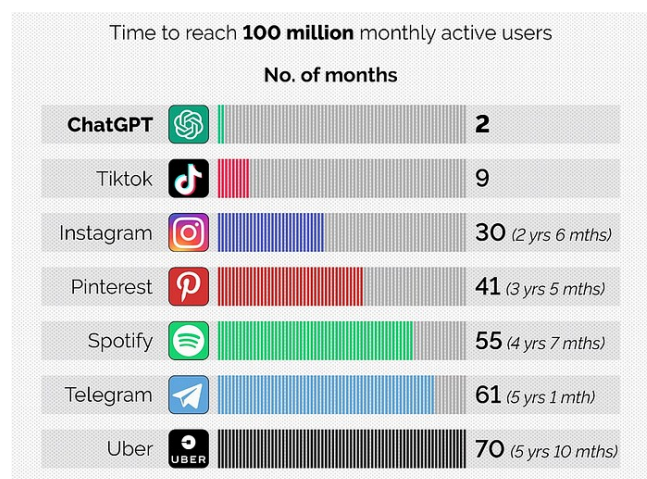


Figura 3 – Comparação de acessos mensais do ChatGPT e outros aplicativos [1].

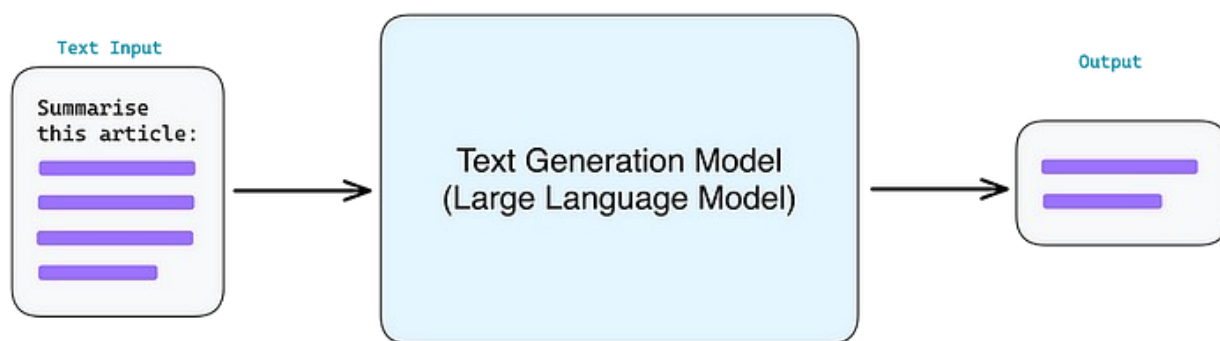


Figura 4 –Interação do ChatGPT com usuário [1].

### 3.2. Bard

O Bard do Google é um chatbot de IA que compete com o ChatGPT da OpenAI. Lançado no Brasil em julho de 2023, a ferramenta é capaz de criar textos naturais, como se fossem escritos por humanos, e também permite criar listas, estruturar planilhas, agendar reuniões e mais. Disponível apenas no navegador, mas integrada a outros serviços do Google, a IA promete aprimorar a experiência de assistência virtual para os usuários. Porém, de acordo com a empresa, o bot ainda está em fase experimental, e sua capacidade é limitada [3].

### 3.3. Claude IA:

O Claude 2 traz melhorias e oferece uma experiência mais aprimorada, inclusive com algumas vantagens sobre o ChatGPT [4].

- a) Melhor leitura, análise e resumo de arquivos carregados;

- b) Maior capacidade de processamento de palavras;
- c) Possui dados até 2022;
- d) Resumo de conteúdos por meio de link.

#### 4. Comparando capacidades de LLMs (Large Language Models)

A seguir, um comparando alguns dos principais LLMs do mercado:

##### 4.1. Generative AI — Multimodal

A maioria dos LLMs atuais são somente de texto, ou seja, eles se destacam apenas em aplicações baseadas em texto e têm capacidade limitada de compreender outros tipos de dados. Exemplo de LLM somente texto: GPT-3.

Já os LLMs multimodais combinam outros tipos de dados, como imagens, vídeos, áudio, junto ao texto. A integração da multimodalidade nos LLMs aborda algumas das limitações dos atuais modelos somente de texto e abre possibilidades para novas aplicações que antes eram impossíveis.

O recentemente lançado GPT-4 pela Open AI é um exemplo de LLM Multimodal. Ele pode aceitar entradas de imagem e texto e mostrou desempenho de nível humano em vários *benchmarks*. Além instruções de entrada via voz e saída também via voz. Fora o poder de gerar imagens incríveis através de prompts de texto (DALL-E 3). A seguir, exemplo de outras ferramentas que geram imagem a partir de texto.

##### 4.1. Midjourney

O Midjourney é um novo gerador de texto para imagem gerido por inteligência artificial. A ferramenta opera dentro do *Discord* e funciona de forma colaborativa. Deste modo, as criações são disponibilizadas para todos que estiverem dentro do grupo onde as artes foram criadas.

Em comparação direta ao DALLE (OpenAI), que renderiza imagens de uma forma mais realista, o Midjourney é voltado para a reprodução baseada em diferentes estilos de artes, conforme Figura 5, ou seja, tem uma pegada mais criativa e abstrata, enquanto seu rival busca fidelidade ao texto interpretado.



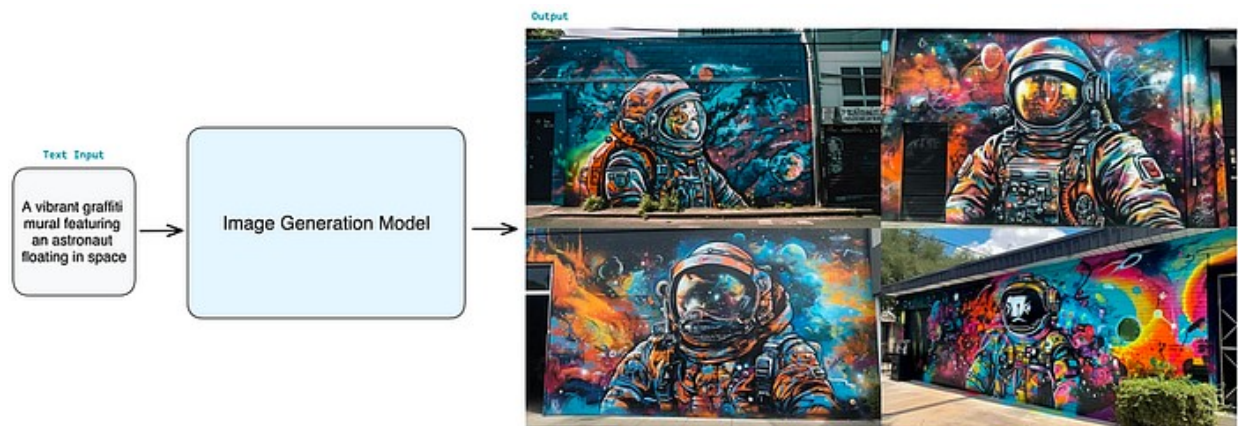


Figura 5 – Exemplos de Artes geradas pelo Midjourney [1].

#### 4.2. Leonardo AI

Para criar belos recursos para jogos, como itens, ambientes, capacetes, edifícios e arte conceitual, a Leonardo AI usa modelos de IA. É uma interface intuitiva projetada com

artistas em mente, permitindo que os usuários gerem ideias rapidamente, treinem seus próprios modelos de IA e produzam materiais originais adequados para produção, conforme Figura 6.

O Leonardo AI usa modelos de aprendizado profundo de última geração para gerar ativos de jogos realistas e diversificados que correspondam às suas especificações. Você pode escolher entre uma variedade de gêneros, estilos, temas e humores para se adequar à sua visão de jogo. Você também pode misturar e combinar diferentes elementos para criar ativos únicos e originais.

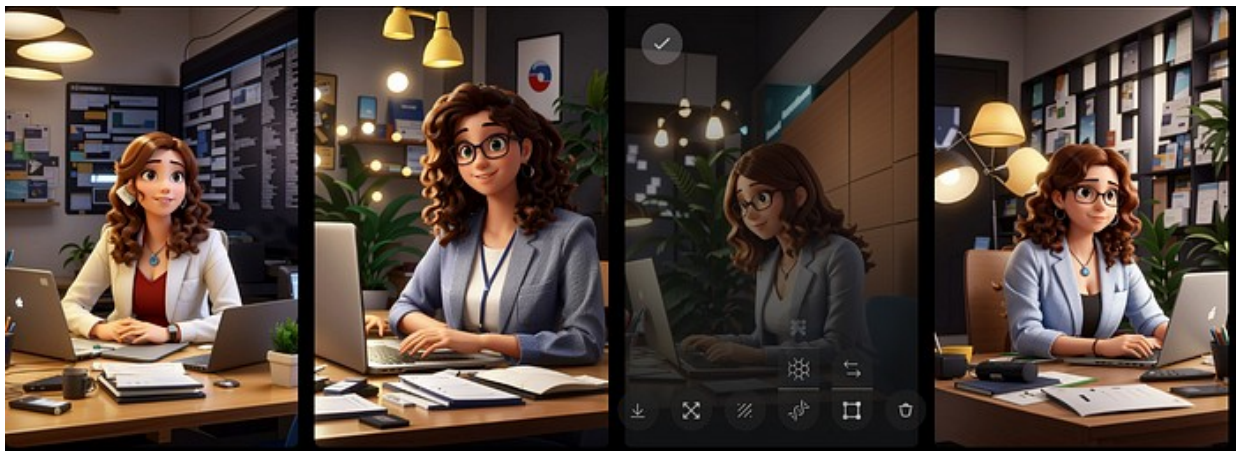


Figura 6 – Exemplos de Artes geradas pelo Leonardo AI [1].

## 5. Aplicações em Engenharia de Controle e Automação

Um estudo comparativo de performance das LLMs na criação e simulação de um foguete está na referência [5]. Já a referência [6] apresenta um estudo da LLM na área de engenharia de controle e automação.

## 6. O Futuro das IAs Generativas

As IAs generativas baseadas em LLMs estão em constante evolução, impulsionando avanços significativos no campo do processamento de linguagem natural. À medida que a pesquisa e o desenvolvimento nessa área continuam a avançar, podemos esperar melhorias contínuas na qualidade das respostas geradas e na capacidade de entender o contexto e a intenção do usuário [1].



No entanto, é importante mencionar que esses modelos também enfrentam desafios, como a necessidade de treinamento em grandes conjuntos de dados, a garantia da segurança e a mitigação de vieses. A comunidade de pesquisa e as empresas estão trabalhando ativamente para enfrentar essas questões, visando criar IAs generativas mais confiáveis, éticas e eficazes [1].

Em resumo, as IAs generativas baseadas em LLMs, como o ChatGPT, Midjourney e Leonardo AI, estão transformando a forma como interagimos com a tecnologia. Esses modelos têm um enorme potencial em diversas áreas, desde assistência virtual até suporte ao cliente, e continuarão a evoluir para oferecer experiências mais aprimoradas e personalizadas no futuro [1].

## **Referências:**

[1] SOUZA, A. Tudo o que você precisa saber sobre LLM (Large Language Model), Blog do Souza, Medium, 2023. Disponível em: <<https://medium.com/blog-do-zouza/tudo-o-que-voce-precisa-saber-sobre-llm-large-language-model-a36be85bbf8f>>. Acesso: Maio 08, 2025.

[2] SOUZA, A. ChatGPT — Visão Geral, Blog do Souza, Medium, 2023. Disponível em: <<https://medium.com/blog-do-zouza/chatgpt-visao-geral-f68ed1d1cf54>>. Acesso: Maio 08, 2025.

[3] VILLARINHO, J. O que é Bard? Guia ensina como funciona e como usar a IA do Google, Techtudo, 2023. Disponível em: <<https://www.techtudo.com.br/guia/2023/07/o-que-e-bard-guia-ensina-como-funciona-e-como-usar-a-ia-do-google-edsoftwares.ghtml?>>. Acesso: Maio 08, 2025.

[4] CHAVES, M. Claude IA: 4 coisas que o chatbot pode fazer que o ChatGPT não consegue, Olhar Digital, 2023. Disponível em: <<https://olhardigital.com.br/2023/08/24/dicas-e-tutoriais/claude-ia-4-coisas-que-o-chatbot-pode-fazer-que-o-chatgpt-nao-consegue/?>>. Acesso: Maio 08, 2025.

[5] SIMONDS, T.. LLMs for Engineering: Teaching Models to Design High Powered Rockets. arXiv preprint arXiv:2504.19394, 2025. Disponível em: <<https://arxiv.org/abs/2504.19394>>. Acesso: Maio 08, 2025.

[6] GUO, Xingang et al. ControlAgent: Automating Control System Design via Novel Integration of LLM Agents and Domain Expertise. arXiv preprint arXiv:2410.19811, 2024. Disponível em: <<https://arxiv.org/abs/2410.19811>>. Acesso: Maio 08, 2025.