



A Era Gemma: Desvendando a Família de Modelos Abertos do Google

Victor Pugliese
Google Developer Expert em IA



Google
Developer
Groups

Victor Pugliese

- Apaixonado por pesquisa e desenvolvimento

Graduação



MBA



Mestrado



Doutorado



- Desde 2017 utilizando Inteligência Artificial na indústria e +15 publicações acadêmicas



- Última experiência





victorpug.exe

🎵 Hélio Ziskind • Castelo Rá-Tim-Bum (Abertur...

⋮



Google
Developer
Groups

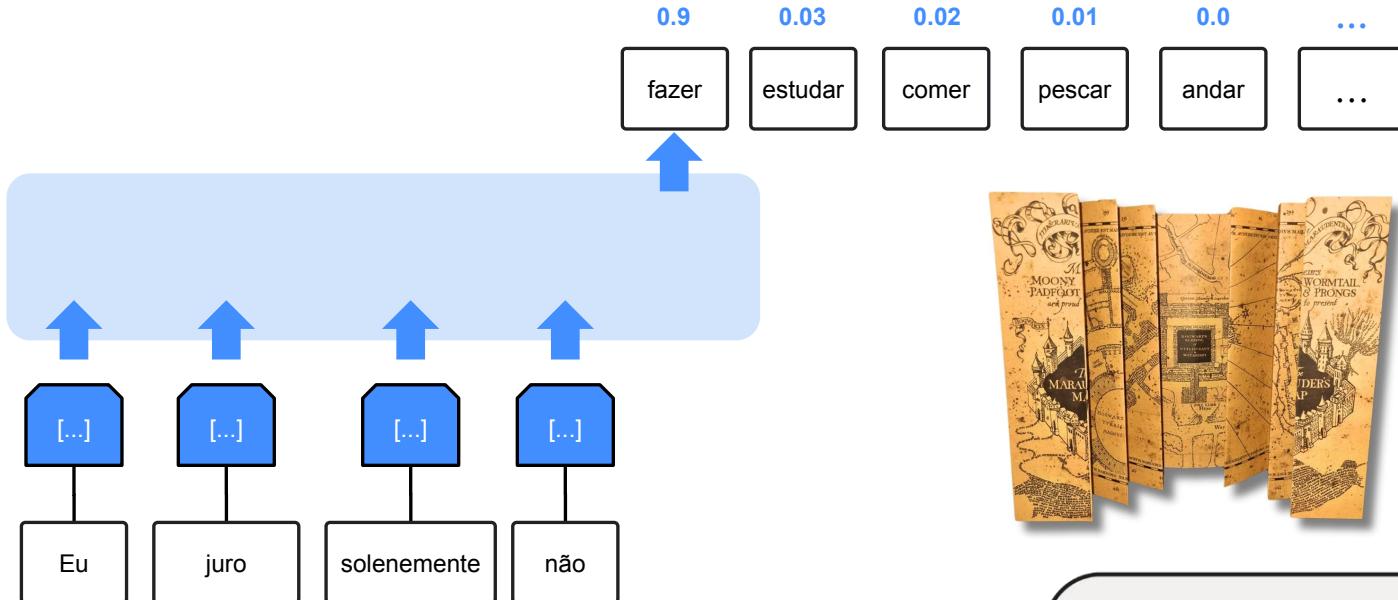
O que é um LLM?

- Um **grande modelo de linguagem** (ou LLM) é um tipo de modelo de inteligência artificial que foi treinado em um grande conjunto de dados, muitas vezes, mas nem sempre, composto de “linguagem humana”.



Google
Developer
Groups

O que é LLM?



Google
Developer
Groups

Klift Kloft Still,



Google
Developer
Groups



Klift Kloft Still, a porta se abriu



Google
Developer
Groups

Olá

Insira um comando para o Gemini



Ferramentas

2.5 Flash ▾



Google
Developer
Groups

O que é Gemma?

- O Gemma é baseado na mesma pesquisa do Gemini e, por ser um modelo aberto, você pode baixá-lo e implantá-lo onde quiser.
- O Gemma é a resposta do Google para **profissionais de IA que buscam mais controle**.
- **Com pesos abertos e uso comercial responsável.**
- Permite ajuste e implantação em projetos e aplicativos próprios.



Google
Developer
Groups

2024

Feb 21, 2024
Gemma
2B, 7B

Apr 5, 2024
Gemma 1.1
2B, 7B

Jun 27, 2024
Gemma 2
9B, 27B

Jul 31, 2024
Gemma 2
2B

Oct 3, 2024
Gemma 2 for Japan
2B

2025

Mar 12, 2025
Gemma 3
270M, 1B, 4B, 12B,
27B

June 26, 2025
Gemma 3n
E2B, E4B

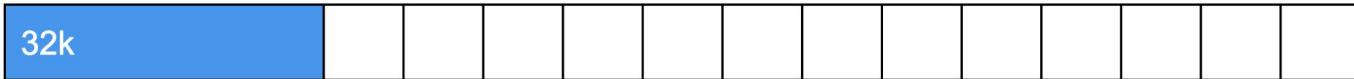


Google
Developer
Groups

Gemma 1, Gemma 2



Gemma 3 1B



Gemma 3 4B, 12B, 27B



Equivale a aproximadamente **200 a 300 páginas**,
considerando que cada página padrão contém
cerca de **400 a 500 palavras**.



Google
Developer
Groups

Gemma: Construído de forma responsável



Segurança por Design



Avaliações robustas



Responsible Generative AI Toolkit



Google
Developer
Groups

Gemma 3

- Recursos **multimodais** para analisar dados visuais.
- Contexto de token de 128K
- Crie interfaces de linguagem natural para trabalhar com APIs.
- Suporte a mais de **140 idiomas** para expandir recursos de IA.
- Versões amigáveis de 270M, 1B, 4B, 12B, 27B



Google
Developer
Groups

- Esta tabela mostra a memória aproximada de GPU ou TPU

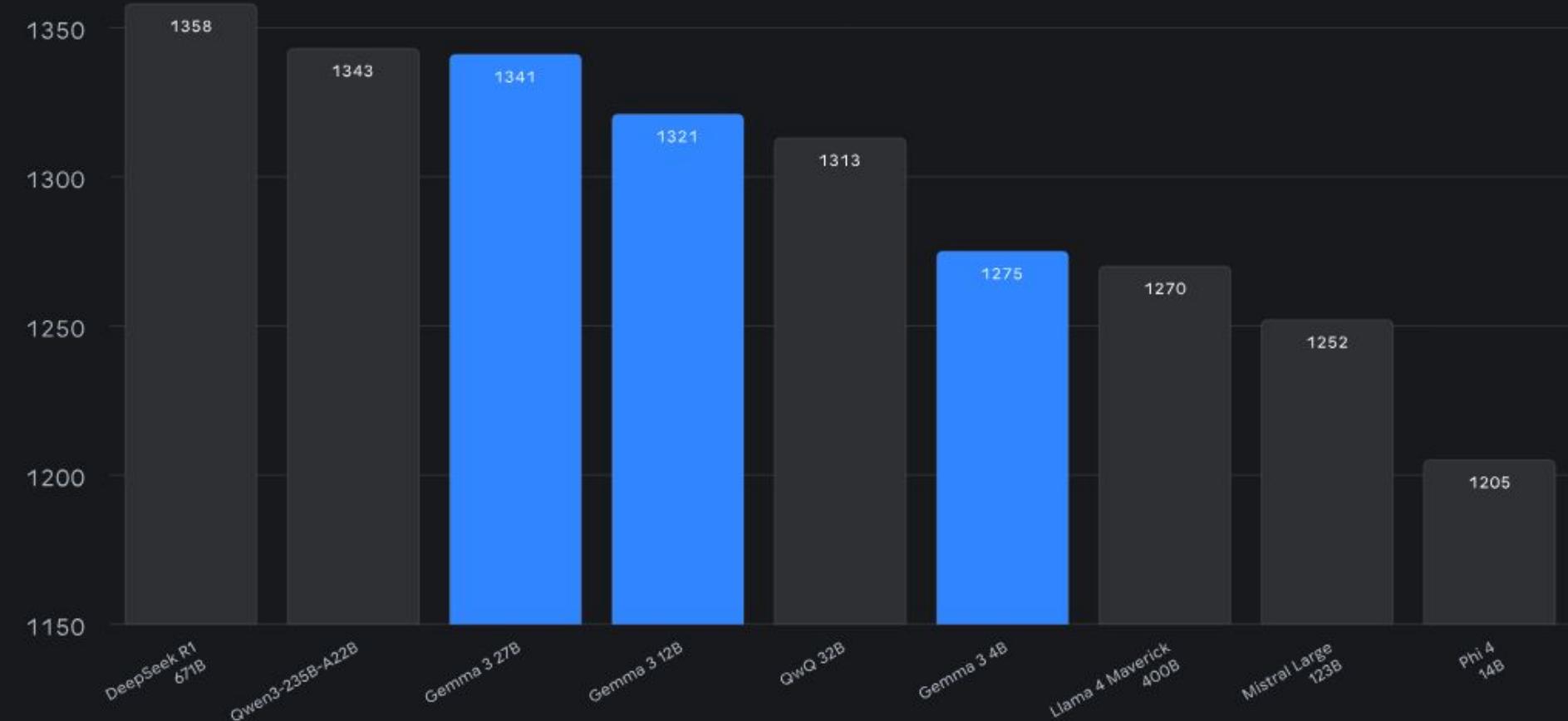
Parâmetros (Tamanho do Modelo)	32 bits completo	BF16 (16 bits)	INT4 (4 bits)
Gemma 3 1B (texto)	4 GB	1,5 GB	861 MB
Gemma 3 4B	16 GB	6,4 GB	3,2 GB
Gemma 3 12B	48 GB	20 GB	8,2 GB
Gemma 3 27B	108 GB	46,4 GB	19,9 GB



Google
Developer
Groups

LMArena Leaderboard Elo scores

as of May 14, 2025



Modelos Abertos: Quando Usar

- Quando você tem uma opinião formada sobre onde deseja que sua IA seja executada.
- Quando você não quer enviar seus dados para um servidor externo.
- Quando você quer treinar sua IA com seus próprios dados



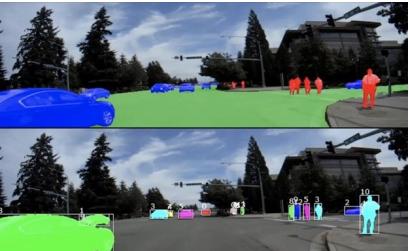
Google
Developer
Groups

Gemma é menor
que o Gemini, e
esse é o ponto



Modelos abertos: Pequeno vs Grande

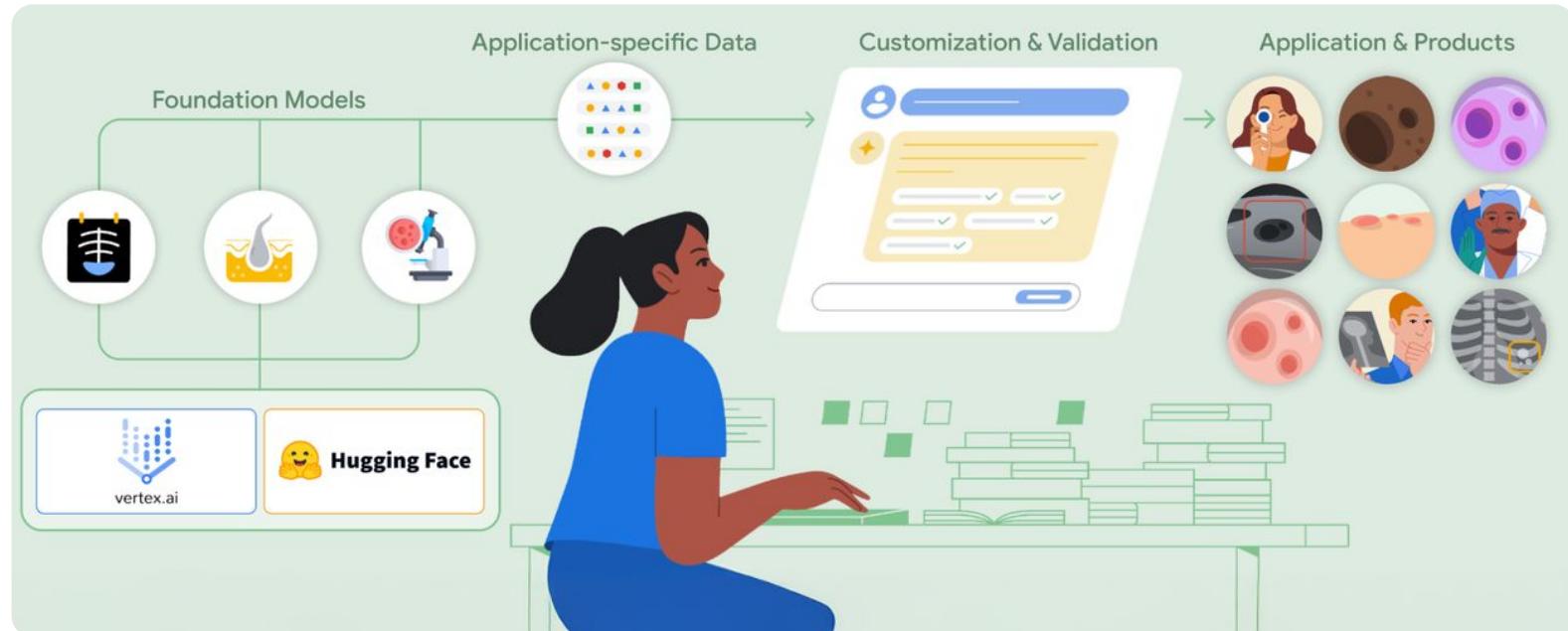
- Espera-se que modelos grandes forneçam conhecimento abrangente. Modelos menores **podem ser mais especializados.**
- Você pode expandir um modelo pequeno com tudo o que você precisa e nada do que não precisa.



Google
Developer
Groups

Health & Life Sciences

Health AI Developer Foundations (HAI-DEF)



Google
Developer
Groups

Health AI Developer Foundations (HAI-DEF)

TXGemma

Uma coleção de modelos para acelerar o desenvolvimento de terapias através da previsão de propriedades de moléculas, proteínas e muito mais.

MedGemma

Variantes do Gemma 3 treinadas para alto desempenho na compreensão de textos e imagens médicas.

Cell2Sentence

O modelo é treinado para entender o "idioma" das células individuais, permitindo-lhe capturar relações sutis entre genes, células e tecidos.



**É fácil aprender sobre
Gemma.**

Onde quer que bons modelos sejam encontrados

- Os modelos Gemma estão disponíveis em plataformas populares como **Kaggle, Hugging Face Hub e outras.**
- Basta solicitar acesso na plataforma de sua escolha, concordar com os Termos de Uso da Gemma e você estará pronto para começar!



Google
Developer
Groups

kaggle

Level up with the largest AI & ML community

Join over 17M+ machine learners to share, stress test, and stay up-to-date on all the latest ML techniques and technologies. Discover a huge repository of community-published models, data & code for your next project.



Google
Developer
Groups

Kaggle: Teste suas habilidades, ganhe prêmios e aprenda

- Resolva problemas reais com conjuntos de dados exclusivos!
Teste suas habilidades em IA contra os melhores, concorrendo a prêmios em dinheiro e reconhecimento do setor.
- As competições continuam relevantes pós-evento: participantes frequentemente publicam suas soluções, servindo como valioso recurso de aprendizado.



Google
Developer
Groups

Kaggle

Competições em destaque



Google - The Gemma 3n Impact Challenge

Explore the newest Gemma model and build your best products for a better world

August 6, 2025
submission deadline
\$150,000
in prizes



Google
Developer
Groups

Demonstração