

GPT-4o

Mohan Ramadoss



LLMs (Large Language Models)

A large language model is a model trained on a large amount of text data that is capable of **understanding** human text and/or **generating** human-like text



ChatGPT

LLaMA
by  **Meta**




Gemini



GPT

GPT stands for [Generative Pre-trained Transformer](#). It refers to a series of large language models developed by OpenAI that use transformer-based neural network architectures.



ChatGPT



The Gemini logo, featuring the word "Gemini" in a blue and purple gradient sans-serif font, with a small purple four-pointed star above the letter "i".

LLaMA
by  **Meta**



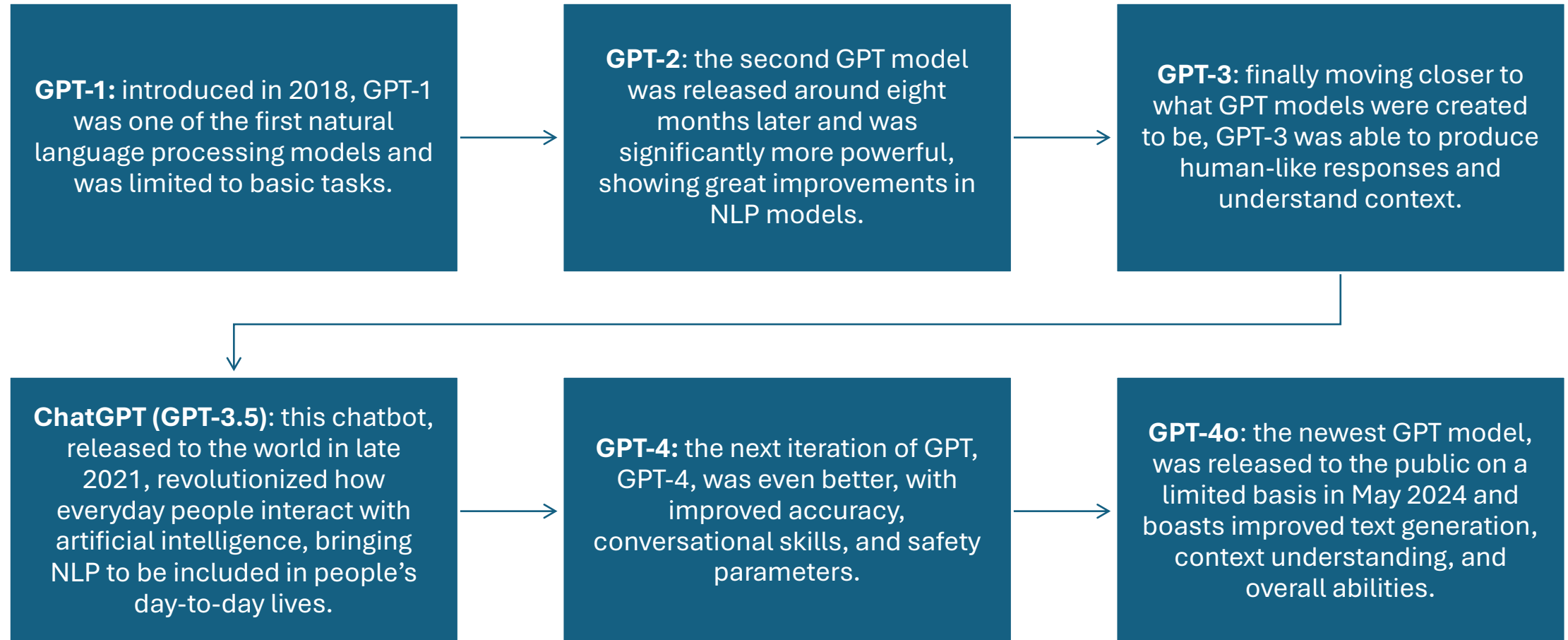
Generative AI

- A generative model is one that can create new content based on data it's been trained on

- Image generation
- Video generation
- Music generation
- Text generation



Quick refresher on the various GPT models



What is GPT-4o?

Features of GPT-4o



GPT-4o is OpenAI's latest LLM. The 'o' in GPT-4o stands for "omni"—represents a significant advancement in AI. **Omniscient (knowing everything) & Omnipotent (having unlimited power)**

Unlike GPT-4, which only handles text.

GPT-4o is a multimodal model that processes and generates text, audio, and visual data.

- **Free** for Everyone though free users will have capacity Limits
- Speaks **in 50 different** languages and multiple tone
- **2 x faster**
- **50 % cheaper**
- **5 x higher** rate limits compared to GPT-4 Turbo
- Rumors claim that **GPT-4 has 1.76 trillion parameters (GPT4 is 8 x 220B params = 1.7T params)**

With Use Cases That Will Melt Your Brain



**It can act as a
tutor**



**It can act as a real
time translator**



**It can act as a
coding assistant**



**It can act as a fair
host**



**Can help blind
and low vision
individuals**

New Features of GPT-4o

Enhanced Artificial Intelligence:

- Provides better creativity, reasoning, problem-solving, and knowledge.
- Handles greater context (more than 25,000 words).

Data Analysis:

- Upload data to receive insights, patterns, trends, and create charts/tables.
- Supports 512MB per file and up to 10 files in a conversation.

Image and Voice Capabilities:

- Understands and generates responses based on images and voice inputs.
- Can solve math problems from images or suggest recipes from fridge contents.
- Allows voice conversations for a more interactive experience.

File Uploads:

- Analyze, transform, and extract information from uploaded files.
- Identify tone, style, and provide summaries or feedback.

Longer Memory:

- Retains guidelines and preferences across conversations, improving user experience.

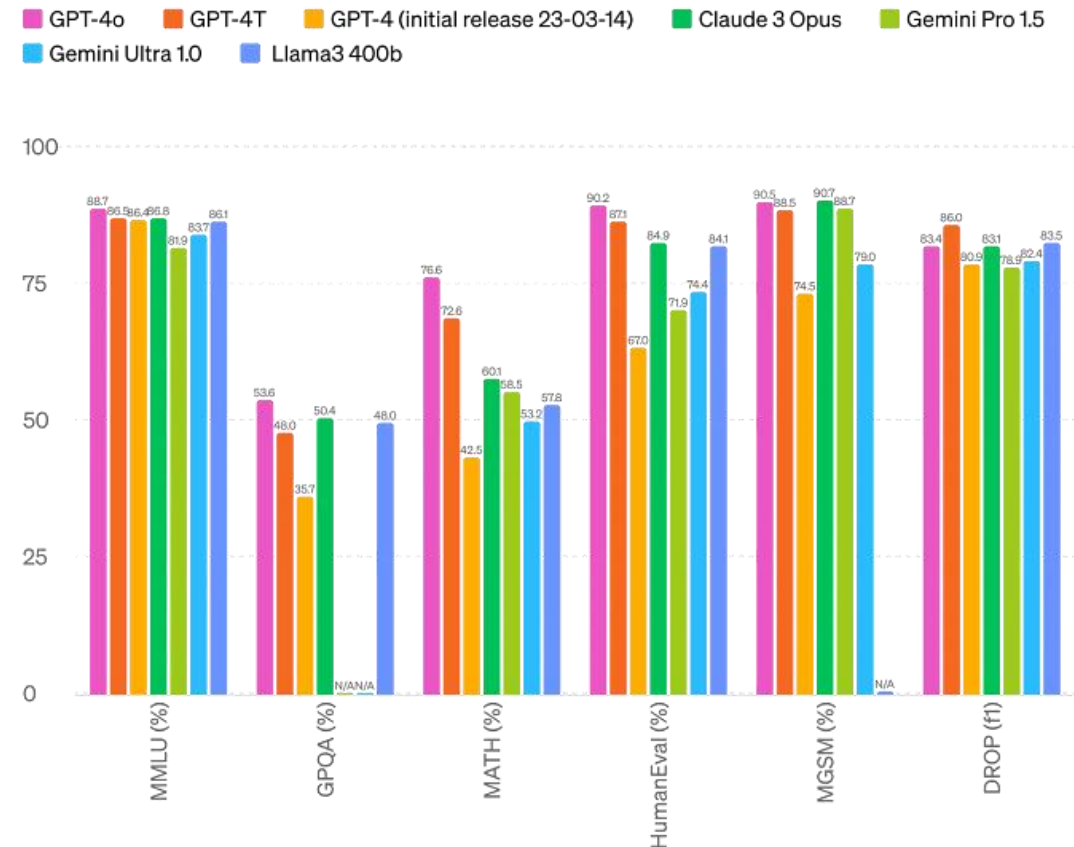


GPT-4o has surpassed other AI models

Text Evaluation

The image shared by OpenAI clearly shows that GPT-4o outperforms other models, particularly in areas such as Math and HumanEval — attributes highly valued by users for enabling smoother and more human-like conversations

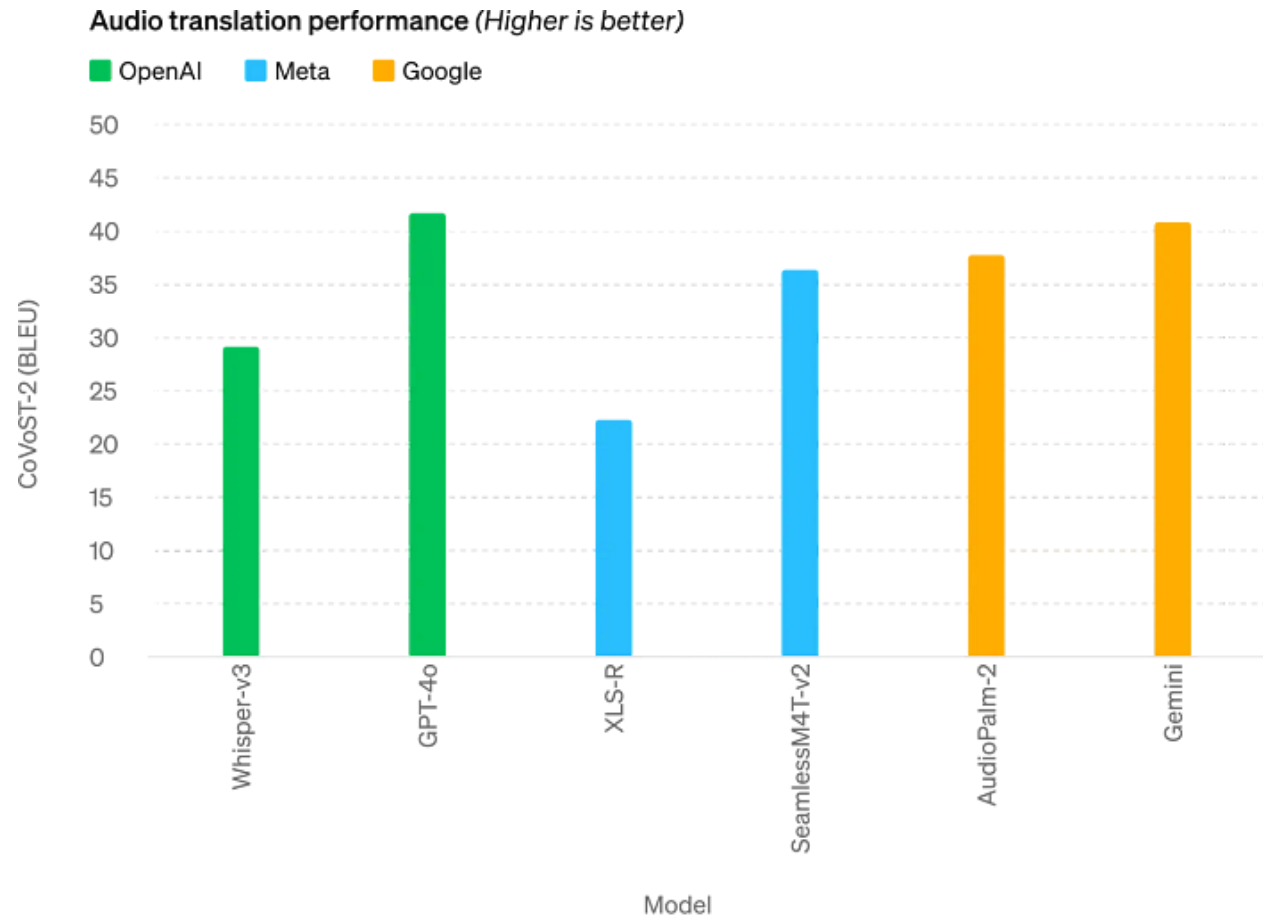
Text Evaluation



Audio translation performance

The enhanced capabilities of GPT-4o, coupled with its Text Evaluation feature, provide an opportunity to connect with more people, recognizing that language often serves as a barrier beyond just communication.

The chart clearly illustrates that GPT-4o has outperformed other AI systems, such as Gemini and Whisper-v3.



Vision understanding evals

GPT-4o's state-of-the-art performance in visual perception benchmarks through 0-shot evaluations, including advanced reasoning in MMMU, MathVista, and ChartQA, underscores its significant capabilities in understanding and reasoning across multi-modal information. Its proficiency in 0-shot Chain-of-Thought reasoning further highlights its robustness and adaptability in diverse visual tasks.

Eval Sets	GPT-4o	GPT-4T 2024-04-09	Gemini 1.0 Ultra	Gemini 1.5 Pro	Claude Opus
MMMU (%) (val)	69.1	63.1	59.4	58.5	59.4
MathVista (%) (testmini)	63.8	58.1	53.0	52.1	50.5
AI2D (%) (test)	94.2	89.4	79.5	80.3	88.1
ChartQA (%) (test)	85.7	78.1	80.8	81.3	80.8
DocVQA (%) (test)	92.8	87.2	90.9	86.5	89.3
ActivityNet (%) (test)	61.9	59.5	52.2	56.7	
EgoSchema (%) (test)	72.2	63.9	61.5	63.2	

GPT-4o Use Cases

Modality	Use Cases	Description
Text	Text Generation, Text Summarization, Data Analysis & Coding	Content creation, concise summaries, code explanations, and coding assistance.
Audio	Audio Transcription, Real-Time Translation, Audio Generation	Convert audio to text, translate in real-time, create virtual assistants or language learning.
Vision	Image Captioning, Image Analysis & Logic, Accessibility for Visually Impaired	Describe images, analyze visual information, provide accessibility for visually impaired.
Multi	Multimodal Interactions, Roleplay Scenarios	Seamlessly combine modalities, create immersive experiences.

How ChatGPT Enhances Team Collaboration

Engineering	Marketing	Sales & Success	Finance & Accounting	IT	Operations	HR & Recruiting
Code reviews and debugging assistance.	Content creation and optimization.	Customer support and chatbots.	Financial data analysis and reporting.	Network monitoring and issue resolution.	Workflow automation and process optimization.	Candidate screening and interview scheduling.
Automated documentation generation.	Social media management and engagement.	Lead qualification and follow-ups.	Automated invoicing and expense tracking.	User support and helpdesk automation.	Inventory and supply chain management	Employee onboarding and training support.
Providing quick answers to technical questions.	Market research and trend analysis.	Personalized sales pitches and proposals.	Compliance and audit support.	Security threat detection and response.	Real-time performance monitoring and reporting.	Policy documentation and compliance tracking.

Future Potential - GPT5

Parameter size GPT -5 will be twice

Human like AI assistance

Human like reasoning abilities

Multimodality - Video processing emerges as a natural progression for GPT-5.

Better accuracy

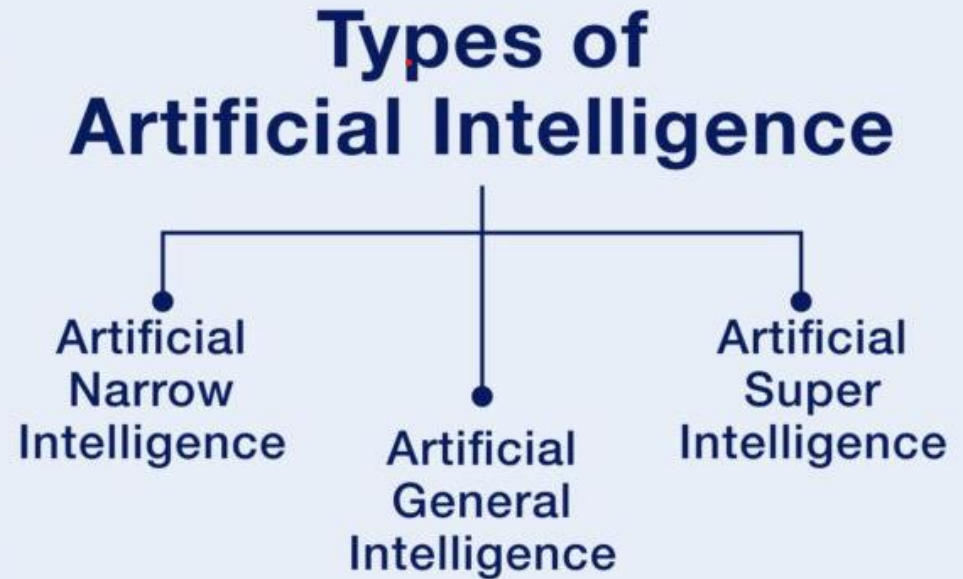
Increased context windows

Cost-effective use of the OpenAI API

From Chatbot to Agent.

Artificial General Intelligence (AGI)

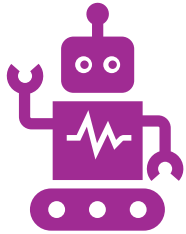
- **Artificial general intelligence (AGI)** is a type of artificial intelligence (AI) that matches or surpasses human capabilities across a wide range of cognitive tasks.
- Regarding whether modern large language models (LLMs) such as GPT-4 are early, incomplete forms of AGI



Future AI Generation details

- AI – 2G
- Generative AI – 4G
- Artificial General Intelligence - 5G
- Artificial Super Intelligence - 6G - End 2030
- 2032 Entire AI will be replaced by Organized Intelligence Most Emerging Multi intelligence

Conclusion



GPT-4o is user-friendly
and intelligent



Enhances communication
and task management



Encouragement to explore
GPT-4o's capabilities