

# GPT-4 has arrived. It will leave ‘ChatGPT’ in the dust.

**OpenAI has released GPT-4, the latest version of its hugely popular artificial intelligence chatbot ChatGPT.**

The new model can respond to images - providing recipe suggestions from photos of ingredients, for example, as well as writing captions and descriptions . It can also process up to 25,000 words, about eight times as many as ChatGPT.

OpenAI has created GPT-4, the latest milestone in OpenAI’s effort in scaling up deep learning. GPT-4 is a large multimodal model (accepting image and text inputs, emitting text outputs) that, while less capable than humans in many real-world scenarios, exhibits human-level performance on various professional and academic benchmarks.

For example, it passes a simulated bar exam with a score around the top 10% of test takers; in contrast, GPT-3.5’s score was around the bottom 10%. They spent 6 months iteratively aligning GPT-4 using lessons from their adversarial testing program as well as ChatGPT, resulting in their best-ever results on factuality, steerability.

Over the past two years, OpenAI has rebuilt their entire deep learning stack and together with MS Azure, co-designed a supercomputer from the ground up for the workload. In 2022, they trained GPT-3.5 as a first “test run” of the system. GPT-4 training run was unprecedentedly stable, becoming the first large model whose training performance were able to accurately predict ahead of time.

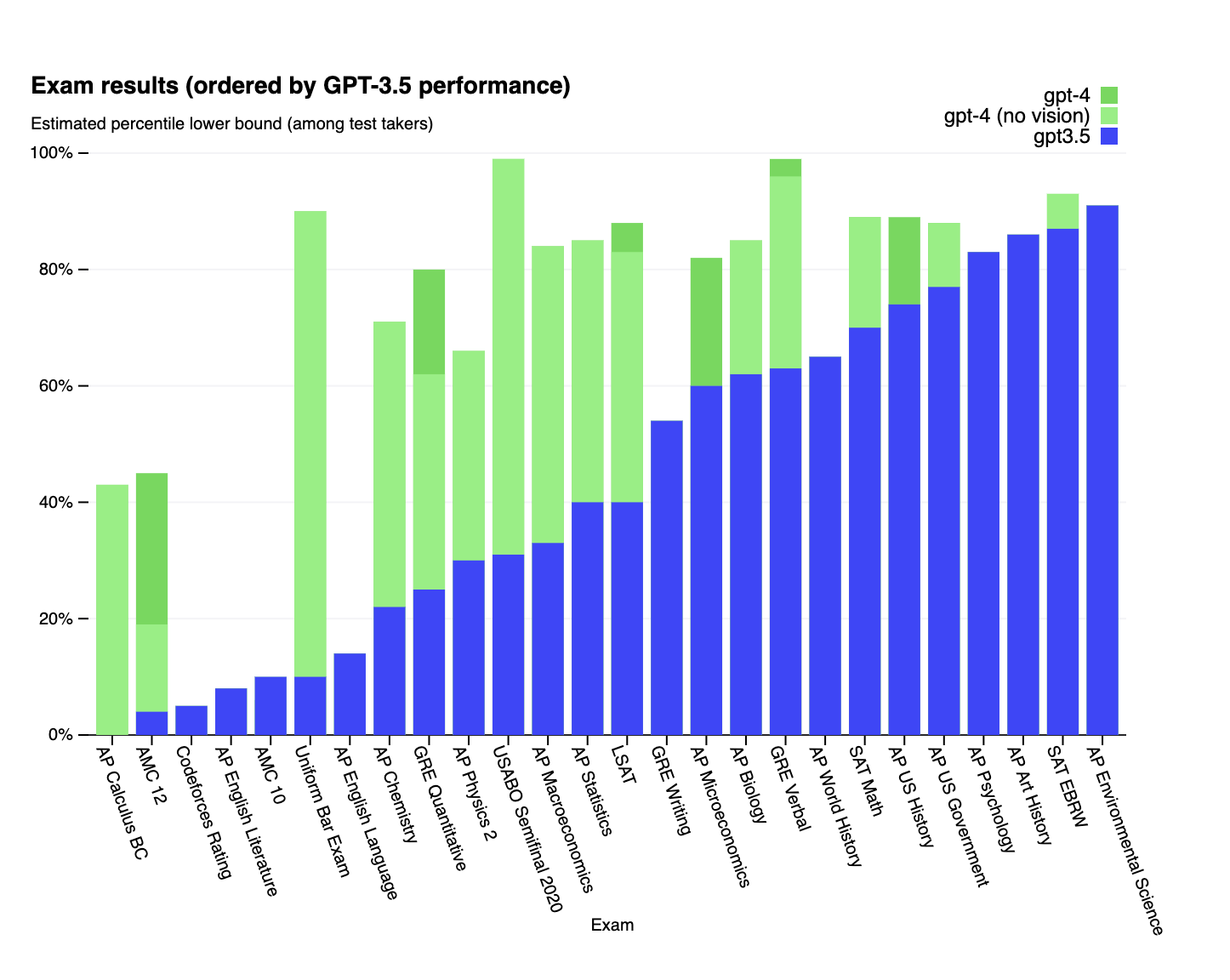


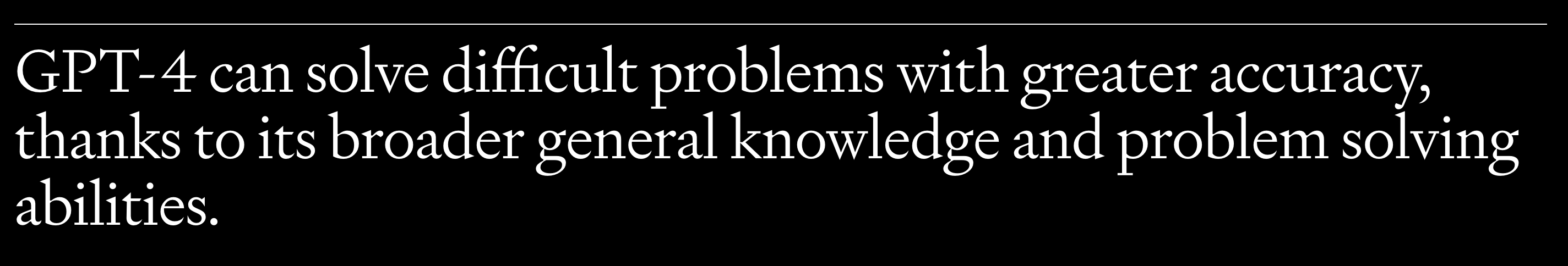
Accessibility of GPT-4’s text input capability is opened via ChatGPT and the API.

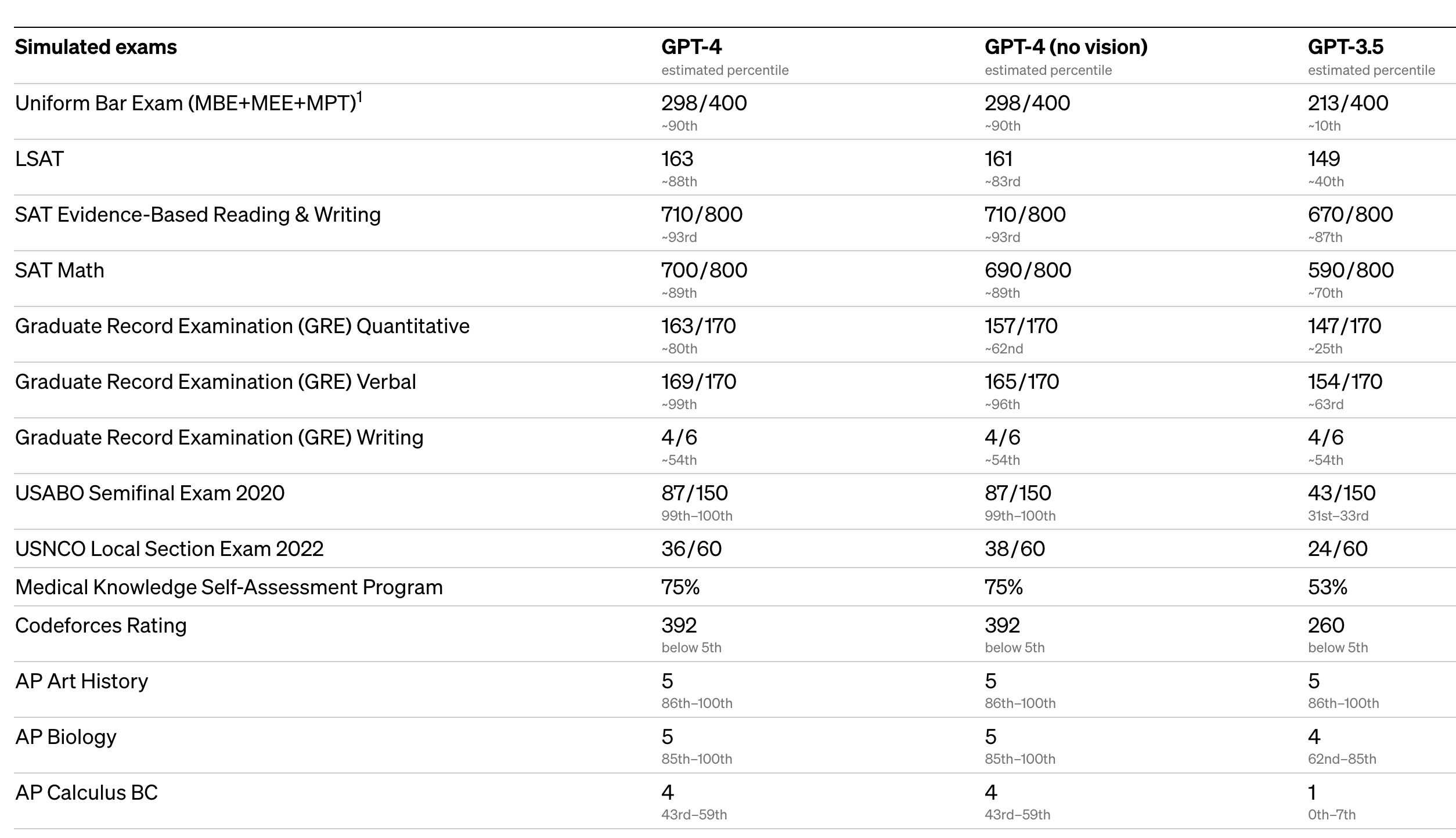
**What is it up to ?**

In a casual conversation, the distinction between GPT-3.5 and GPT-4 can be subtle. The difference comes out when the complexity of the task reaches a sufficient threshold—GPT-4 is more reliable, creative, and able to handle much more nuanced instructions than GPT-3.5.

To understand the difference between the two models, OpenAI tested on a variety of benchmarks, including simulating exams that were originally designed for humans. They proceeded by using the most recent publicly-available tests (in the case of the Olympiads and AP free response questions) or by purchasing 2022–2023 editions of practice exams.







**Input : Text or Image**

GPT-4 can accept a prompt of text and images, which—parallel to the text-only setting—lets the user specify any vision or language task. Specifically, it generates text outputs (natural language, code, etc.) given inputs consisting of interspersed text and images. Over a range of domains—including documents with text and photographs, diagrams, or screenshots—GPT-4 exhibits similar capabilities as it does on text-only inputs. Furthermore, it can be augmented with test-time techniques that were developed for text-only language models, including few-shot and chain-of-thought prompting. Image inputs are still a research preview and not publicly available.

**Steerability**

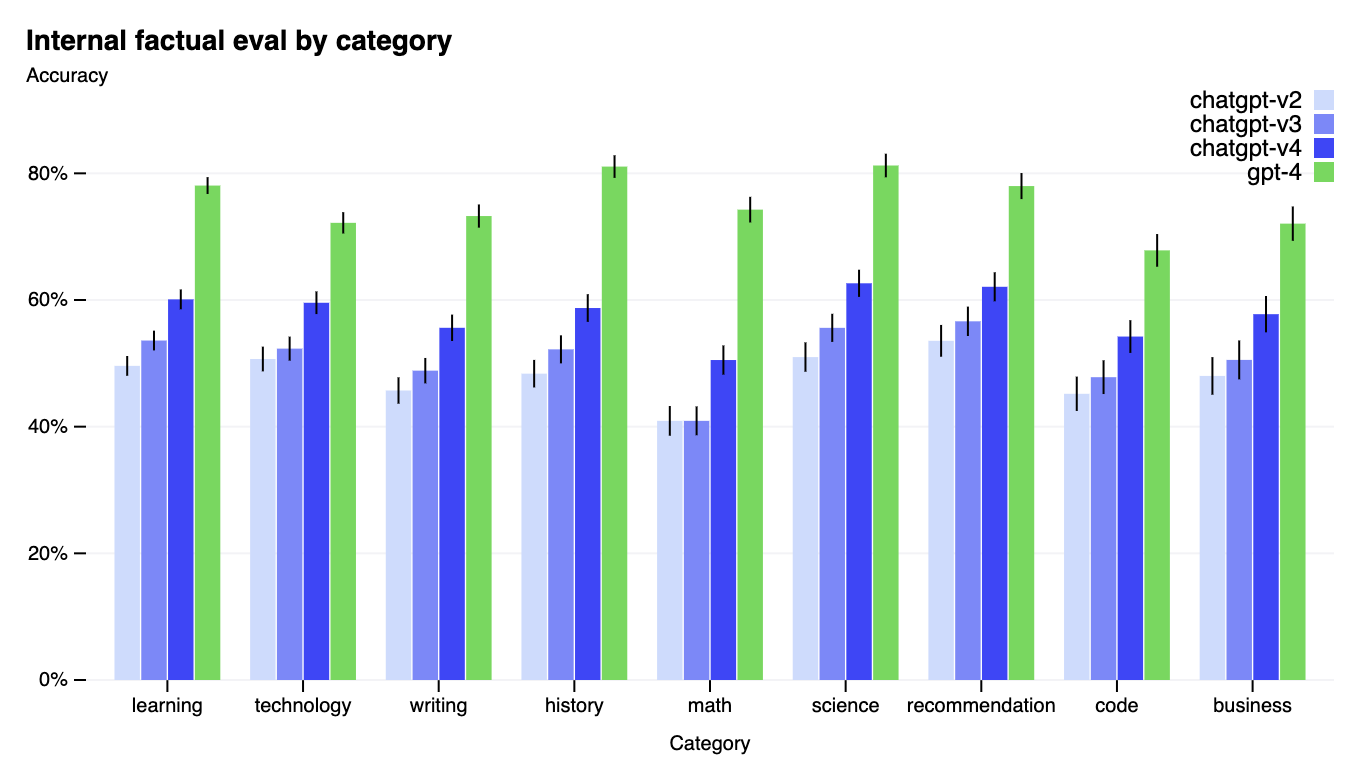
Rather than the classic ChatGPT personality with a fixed verbosity, tone, and style, developers (and soon ChatGPT users) can now prescribe their AI’s style and task by describing those directions in the “system” message. System messages allow API users to significantly customize their users’ experience within bounds.



**Limitations**

Despite its capabilities, GPT-4 has similar limitations as earlier GPT models. Most importantly, it still is not fully reliable (it “hallucinates” facts and makes reasoning errors). Great care should be taken when using language model outputs, particularly in high-stakes contexts, with the exact protocol (such as human review, grounding with additional context, or avoiding high-stakes uses altogether) matching the needs of a specific use-case.

While still a real issue, GPT-4 significantly reduces hallucinations relative to previous models (which have themselves been improving with each iteration). GPT-4 scores 40% higher than our latest GPT-3.5 on our internal adversarial factuality evaluations:



**ChatGPT Plus**

ChatGPT Plus subscribers will get GPT-4 access on *chat.openai.com* with a usage cap. OpenAI will adjust the exact usage cap depending on demand and system performance in practice, but it expect to be severely capacity constrained (though scale up and optimize over upcoming months).

**API : For Developers**

As of today to get access to the GPT-4 API (which uses the same Chat Completions API as gpt-3.5-turbo), signing up for their waitlist is must.

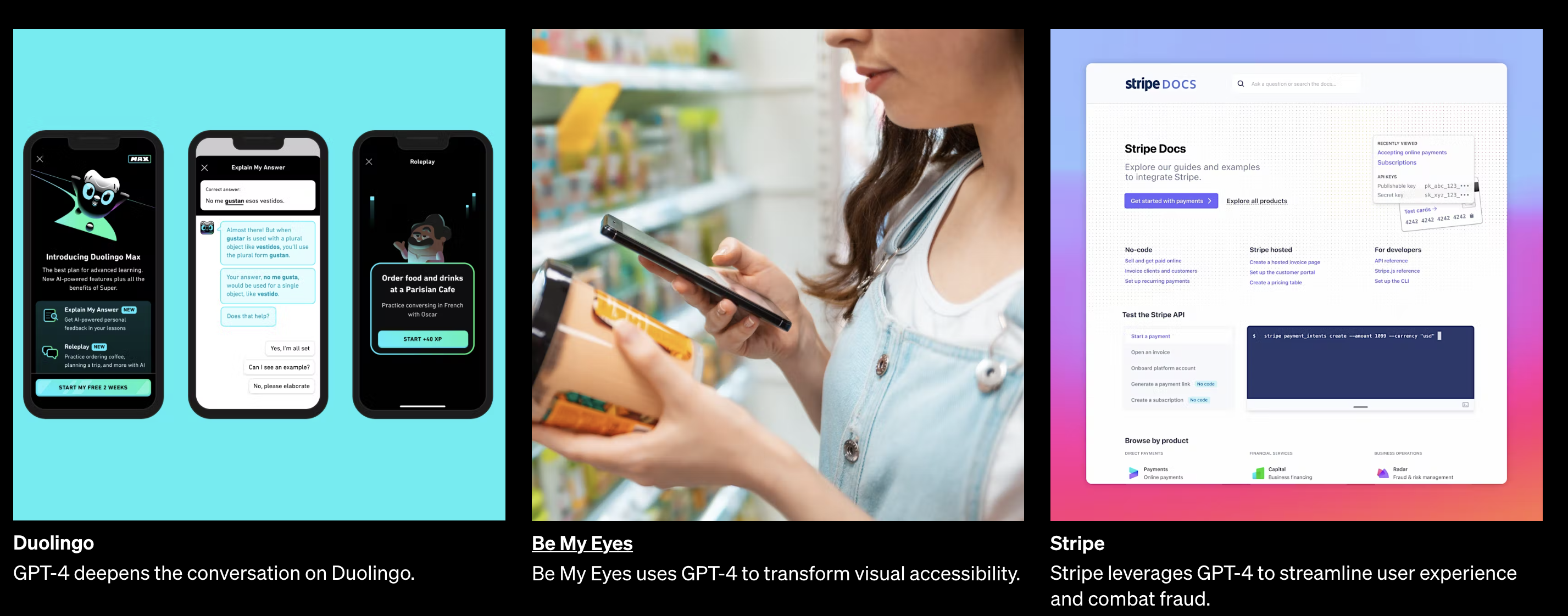
Once you have access, you can make text-only requests to the gpt-4 model (image inputs are still in limited alpha).And the current build will be discontinued after a new stable release. Pricing is $0.03 per 1k prompt tokens and $0.06 per 1k completion tokens. Default rate limits are 40k tokens per minute and 200 requests per minute.

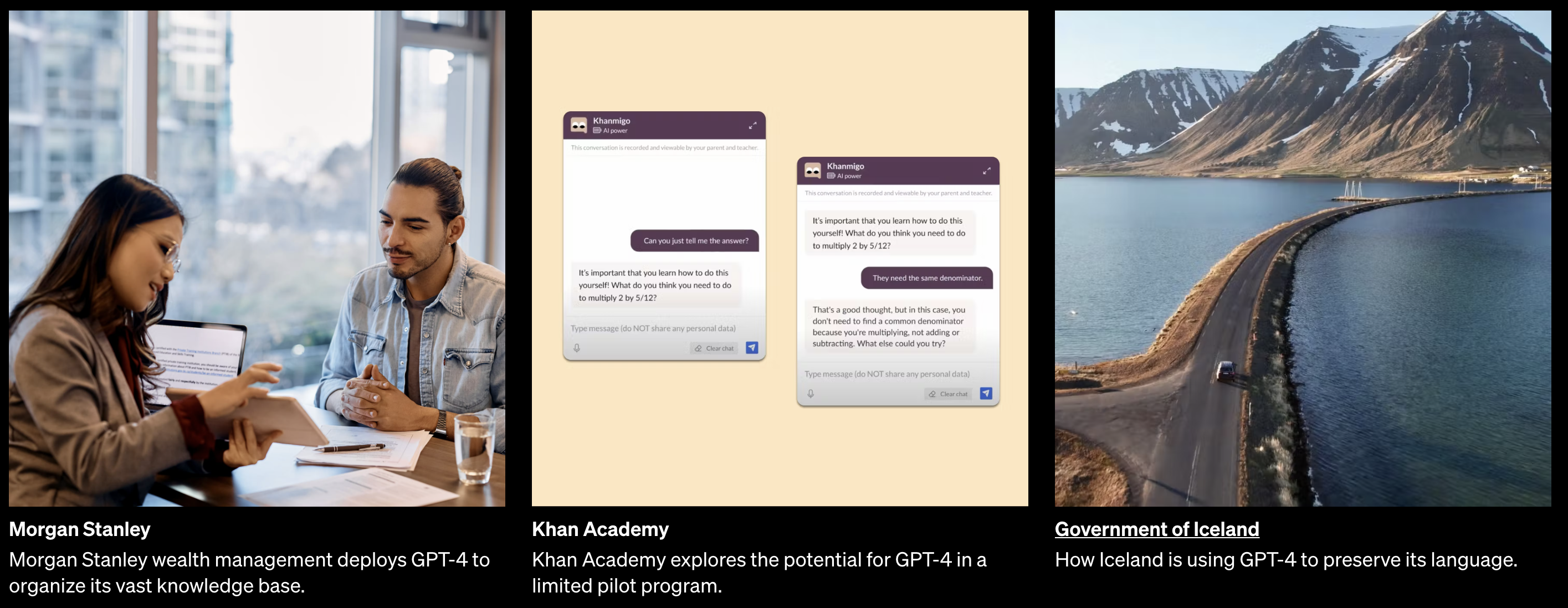
GPT-4 has a context length of 8,192 tokens. They are providing limited access to 32,768–context (about 50 pages of text) version, gpt-4-32k, which will also be updated automatically over time (current version gpt-4-32k-0314, also supported until June 14). Pricing is $0.06 per 1K prompt tokens and $0.12 per 1k completion tokens.

OpenAI has incorporated more human feedback, including feedback submitted by ChatGPT users, to improve GPT-4’s behaviour. They also worked with over 50 experts for early feedback in domains including AI safety and security.

**Collaboration:**

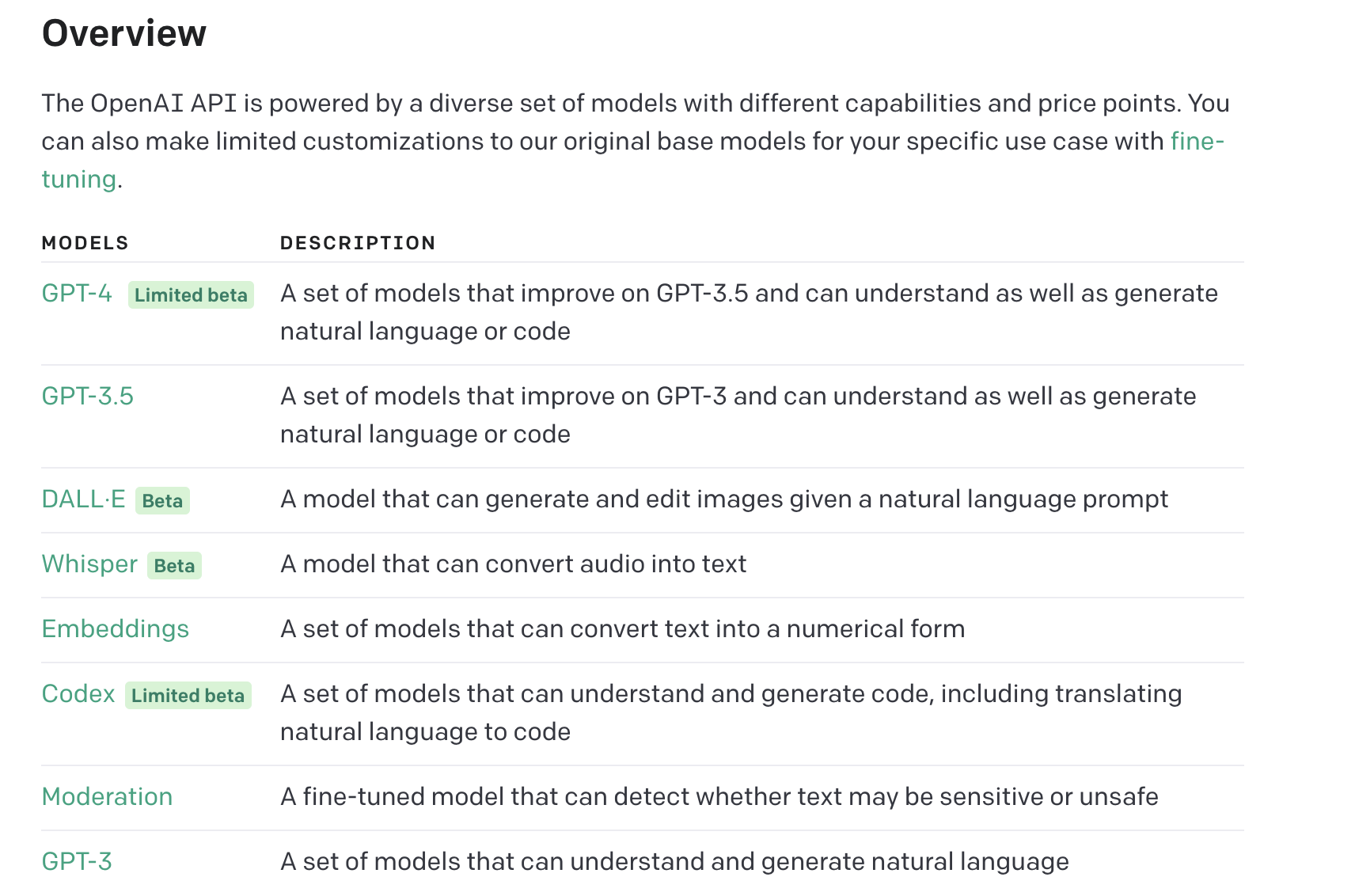
OpenAI has collaborated with organizations building innovative products with GPT-4.





**Where are we now ?**

The difference of different GPT models.



**Ideas for any firm to become a Product Based Unicorn !**

Since GPT is an evolving technology with great potential ,there are a lot of opportunities by building products based on GPT. Nearby 400 or more apps and platforms are already in market based on GPT.

Here are some examples of how GPT can be used to enhance the capabilities of apps:

An easy innovative product that can be realised by any technology solution leaders is Smart Apps based on GPT potential. Either you can *increase your revenue or fame*, a positive result is assured. Let’s look forward to new GPT releases and its impact in tech market.

**Motivation**

I would like to conclude by a wonder story of a very young start up, **KAITO AI**.

The Seattle start-up was founded in 2022 by Yu Hu, a former international equities trader at Citadel and investment banking analyst at Deutsche Bank.

At its core, Kaito is an AI start-up that is trying to solve the information fragmentation issue in crypto. Data and information is often dispersed across a number of sources such as Discord, Medium, Mirror, podcast transcripts as well as news and research platforms, the company said in the release. Kaito brings this information all in the one place through its AI-powered search engine.

The start-up leverages AI not only to aggregate information but also for ranking, recommendations and topic mining. Over the past 12 months Kaito has built one of the most extensive information databases in crypto. By combining this database and their in-house AI technologies with the advanced language models of ChatGPT/GPT-3, they aim to offer a far superior search experience compared to current alternatives in the market.

Whats the big news…

**Kaito AI raises $5.3 million to build ChatGPT-powered search engine for crypto**