OpenAI, Google and Anthropic Are Struggling to Build More Advanced AI

Three of the leading artificial intelligence companies are seeing diminishing returns from their costly efforts to develop newer models.

The ChatGPT virtual assistant logo. *Photographer: Andrey Rudakov/Bloomberg*

By Rachel Metz, Shirin Ghaffary, Dina Bass, and Julia Love
November 13, 2024 at 5:00 AM EST

OpenAI was on the cusp of a milestone. The startup finished an initial round of training in September for a massive new artificial intelligence model that it hoped would significantly surpass prior versions of the technology behind ChatGPT and move closer to its goal of powerful AI that outperforms humans.

But the model, known internally as Orion, did not hit the company's desired performance, according to two people familiar with the matter, who spoke on condition of anonymity to discuss company matters. As of late summer, for example, Orion fell short when trying to answer coding questions that it hadn't been trained on, the people said. Overall, Orion is so far not considered to be as big a step up from OpenAI's existing models as GPT-4 was from GPT-3.5, the system that originally powered the company's flagship chatbot, the people said.

OpenAI isn't alone in hitting stumbling blocks recently. After years of pushing out increasingly sophisticated AI products at a breakneck pace, three of the leading AI companies are now seeing diminishing returns from their costly efforts to build newer models. At Alphabet Inc.'s Google, an upcoming iteration of its Gemini software is not living up to internal expectations, according to three people with knowledge of the matter. Anthropic, meanwhile, has seen the timetable slip for the release of its long-awaited Claude model called 3.5 Opus.

The companies are facing several challenges. It's become increasingly difficult to find new, untapped sources of high-quality, human-made training data that can be used to build more advanced AI systems. Orion's unsatisfactory coding performance was due in part to the lack of sufficient coding data to train on, two people said. At the same time, even modest improvements may not be enough to justify the tremendous costs associated with building and operating new models, or to live up to the expectations that come with branding a product as a major upgrade.

There is plenty of potential to make these models better. OpenAI has been putting Orion through a months-long process often referred to as post-training, according to one of the people. That procedure, which is routine before a company releases new AI software publicly, includes incorporating human feedback to improve responses and refining the tone for how the model should interact with users, among other things. But Orion is still not at the level OpenAI would want in order to release it to users, and the company is unlikely to roll out the system until early next year, one person said.

These issues challenge the gospel that has taken hold in Silicon Valley in recent years, particularly since OpenAI released ChatGPT two years ago. Much of the tech industry has bet on so-called scaling laws that say more computing power, data and larger models will inevitably pave the way for greater leaps forward in the power of AI.

The recent setbacks also raise doubts about the heavy investment in AI and the feasibility of reaching an overarching goal these companies are aggressively pursuing: artificial general intelligence. The term typically refers to hypothetical AI systems that would match or exceed humans on many intellectual tasks. The chief executives of OpenAI and Anthropic have previously said AGI may be only several years away.

"The AGI bubble is bursting a little bit," said Margaret Mitchell, chief ethics scientist at AI startup Hugging Face. It's become clear, she said, that "different training approaches" may be needed to make AI models work really well on a variety of tasks — an idea a number of experts in artificial intelligence echoed to Bloomberg News.

Margaret Mitchell, chief ethics scientist at AI startup Hugging Face, said "the AGI bubble is bursting." *Photographer: Chona Kasinger/Bloomberg*

In a statement, a Google DeepMind spokesperson said the company is "pleased with the progress we're seeing on Gemini and we'll share more when we're ready." OpenAI declined to comment. Anthropic declined to comment, but referred Bloomberg News to a five-hour [podcast](#) featuring Chief Executive Officer Dario Amodei that was released Monday.

"People call them scaling laws. That's a misnomer," he said on the podcast. "They're not laws of the universe. They're empirical regularities. I am going to bet in favor of them continuing, but I'm not certain of that."

Amodei said there are "lots of things" that could "derail" the process of reaching more powerful AI in the next few years, including the possibility that "we could run out of data." But Amodei said he's optimistic AI companies will find a way to get over any hurdles.

Plateauing Performance

The technology that underpins ChatGPT and a wave of rival AI chatbots was built on a trove of social media posts, online comments, books and other data freely scraped from around the web. That was enough to create products that can spit out clever essays and poems, but building AI systems that are smarter than a Nobel laureate — as some companies hope to do — may require data sources other than Wikipedia posts and YouTube captions.

OpenAI, in particular, has inked deals with publishers to fill some of the need for high-quality data, and also adapt to growing legal pressure from publishers and artists over the data used to build generative AI products. Some tech companies are also hiring people with graduate degrees that can label data related to their own subject expertise, such as math and coding. The goal is to make these systems better at responding to queries about certain topics.

These efforts are slower going and costlier than simply scraping the web. Tech companies are also turning to synthetic data, such as computer-generated images or text meant to mimic content created by real people. But here, too, there are limits.

"It is less about quantity and more about quality and diversity of data," said Lila Tretikov, head of AI strategy at New Enterprise Associates and former deputy chief technology officer at Microsoft. "We can generate quantity synthetically, yet we struggle to get unique, high-quality datasets without human guidance, especially when it comes to language."

Dates of Key Model Releases

Date	Model	Release
Nov. 2018	OpenAI	GPT
Feb. 2019	OpenAI	GPT-2 (partial release)
June 2020	OpenAI	GPT-3
Nov. 2022	OpenAI	GPT-3.5 (first model to power ChatGPT)
March 2023	OpenAI	GPT-4
March 2023	Anthropic	Claude
July 2023	Anthropic	Claude 2
Nov. 2023	Anthropic	Claude 2.1
March 2024	Anthropic	Claude 3 Opus, Sonnet, Haiku
May 2024	OpenAI	GPT-4o
June 2024	Anthropic	Claude 3.5 Sonnet
Sept. 2024	OpenAI	o1-preview
Nov. 2024	Anthropic	Claude 3.5 Haiku

Source: Bloomberg

Still, AI companies continue to pursue a more-is-better playbook. In their quest to build products that approach the level of human intelligence, tech firms are increasing the amount of computing power, data and time they use to train new models — and driving up costs in the process. Amodei has said companies will spend $100 million to train a bleeding-edge model this year and that amount will hit $100 billion in the coming years.

As costs rise, so do the stakes and expectations for each new model under development. Noah Giansiracusa, an associate professor of mathematics at Bentley University in Waltham, Massachusetts, said AI models will keep improving, but the rate at which that will happen is questionable.

"We got very excited for a brief period of very fast progress," he said. "That just wasn't sustainable."

Silicon Valley's Conundrum

This conundrum has come into focus in recent months inside Silicon Valley. In March, Anthropic released a set of three new models and [said](#) the most powerful option, called Claude Opus, outperformed OpenAI's GPT-4 and Google's Gemini systems on key benchmarks, such as graduate-level reasoning and coding.

Over the next few months, Anthropic pushed out updates to the other two Claude models – but not Opus. "That was the one everyone was excited about," said Simon Willison, an independent AI researcher. By October, Willison and other industry watchers noticed that wording related to 3.5 Opus, including an indication that it would arrive "later this year" and was "coming soon," was removed from [some](#) [pages](#) on the company's website.

Similar to its competitors, Anthropic has been facing challenges behind the scenes to develop 3.5 Opus, according to two people familiar with the matter. After training it, Anthropic found 3.5 Opus performed better on evaluations than the older version but not by as much as it should, given the size of the model and how costly it was to build and run, one of the people said.

An Anthropic spokesperson said the language about Opus was removed from the website as part of a marketing decision to only show available and benchmarked models. Asked whether Opus 3.5 would still be coming out this year, the spokesperson pointed to Amodei's podcast remarks. In the interview, the CEO said Anthropic still plans to release the model but repeatedly declined to commit to a timetable.

Anthropic CEO Dario Amodei said there are "lots of things" that could "derail" the process of reaching more powerful AI in the next few years. "We could run out of data," he said. *Photographer: Benjamin Girette/Bloomberg*

Tech companies are also beginning to wrestle with whether to keep offering their older AI models, perhaps with some additional improvements, or to shoulder the costs of supporting hugely expensive new versions that may not perform much better.

Google has released updates to its flagship AI model Gemini to make it more useful, including restoring the ability to generate images of people, but introduced few major breakthroughs in the quality of the underlying model. OpenAI, meanwhile, has focused on a number of comparatively incremental updates this year, such as a new version of a voice assistant feature that lets users have more fluid spoken conversations with ChatGPT.

More recently, OpenAI rolled out a preview version of a model called o1 that spends extra time computing an answer before responding to a query, a process the company refers to as reasoning. Google is working on a similar approach, with the goal of handling more complex queries and yielding better responses over time.

Tech firms also face meaningful tradeoffs with diverting too much of their coveted computing resources to developing and running larger models that may not be significantly better.

"All of these models have gotten quite complex and we can't ship as many things in parallel as we'd like to," OpenAI CEO Sam Altman wrote in response to a question on a recent Ask Me Anything session on Reddit. The ChatGPT-maker faces "a lot of limitations and hard decisions," he said, about how it decides what to do with its available computing power.

Altman said OpenAI will have some "very good releases" later this year, but that list won't include GPT-5 — a name many in the AI industry would expect the company to use for a big release following GPT-4, which was introduced more than 18 months ago.

Like Google and Anthropic, OpenAI is now shifting attention from the size of these models to newer use cases, including a crop of AI tools called agents that can book flights or send emails on a user's behalf. "We will have better and better models," Altman wrote on Reddit. "But I think the thing that will feel like the next giant breakthrough will be agents."