Tomorrow's Affairs
Analysis of today
Assessment of tomorrow

By: Brian Judge

How will AI giants convince their investors that they will remain profitable?

OpenAI and Anthropic will soon have to explain to investors why they are each supposedly worth $2–3 trillion.

They will probably draw from the same template that Elon Musk used for the SpaceX IPO in June.

By presenting a modestly profitable satellite-launch and telecoms business attached to a tremendously unprofitable AI startup as a vehicle for "mak[ing] humanity multiplanetary," Musk achieved a valuation far above what the company's financials justify.

With private markets showing signs of exhaustion, OpenAI and Anthropic's expected IPOs promise access to a much larger pool of capital.

Going public opens not only the equity markets but also the vastly deeper investment-grade debt market, and bankers for OpenAI and Anthropic are already pressing rating agencies to grant them investment-grade marks after their listings.

But that means their pitch to investors must do double duty, giving shareholders a market vast enough to justify today's valuation while convincing creditors that today's cash burn can be contained.

The employment narrative

With these needs in mind, the frontier labs are relying on two narratives that promise a path to profitability.

They claim that AI will not only replace workers and expand the potential market from software budgets to labor-saving applications, but also that the existential risk associated with the technology justifies curtailing additional capital spending.

The first narrative, advanced in a recent paper from Anthropic's in-house economists, holds that AI will capture a meaningful share of corporate wage bills by replacing human workers.

Implicit in this message is a warning to corporate executives across the economy: companies that aggressively deploy AI will outcompete those that do not.

> If agents can complete work with limited human supervision, dramatically slashing corporate wage bills becomes more plausible

The employment narrative usefully recasts a software product as labor. Mere software for helping workers write emails, prepare presentations, and summarize documents would compete for a share of corporate IT budgets, but a system that replaces workers entirely competes for their share of the wage bill.

This is the prize the labs are promising to justify their valuations and capital spending.

Unlike chatbots, which respond to individual prompts, "agents" use additional software tools to carry out multi-step tasks in pursuit of a specific goal.

If agents can complete work with limited human supervision, dramatically slashing corporate wage bills becomes more plausible.

'Existential risk' to humanity

The second narrative—wherein AI poses an "existential risk" to humanity—has now gone mainstream after circulating for decades in specialist circles.

Alan Turing, the British computer scientist who cracked the Nazi Enigma code, warned in 1951 that thinking machines could "outstrip our feeble powers" and eventually "take control."

More recently, Stuart Russell of the University of California, Berkeley, Yoshua Bengio of the University of Montreal, and others have warned that increasingly autonomous AI

systems could escape human control unless there are stronger safeguards and public oversight; and whistleblowers at the frontier labs continue to draw attention to this issue.

> In preparing for their IPOs, the labs have found a financial use case for these longstanding warnings about AI

Yet in preparing for their IPOs, the labs have found a financial use case for these longstanding warnings about AI.

Fears of mass joblessness or human extinction offer them a good reason to restrain spending without conceding that further improvements might prove uneconomical.

The timing of Anthropic CEO Dario Amodei's call to "pace the frontier," which was immediately endorsed by OpenAI's Sam Altman and Musk, is telling.

After all, these same companies long rebuffed demands for a "pause" in training frontier models.

The happy equilibrium

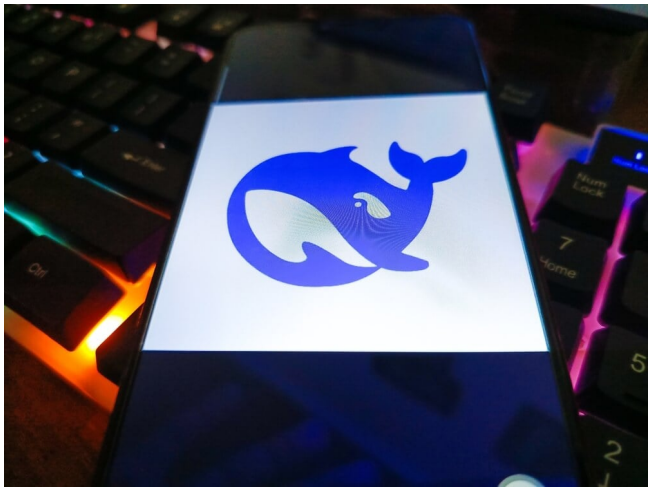
For the leading labs, the happy equilibrium would include models powerful enough to automate substantial swaths of white-collar work, but dangerous enough to make continued breakneck spending on training frontier models imprudent.

The employment narrative promises the revenues, and the safety narrative supplies a justification for stopping the investment race.

Slowing down can be presented as proof of corporate responsibility, rather than as an admission that further improvements might not pay off.

Of course, Chinese competition threatens this equilibrium. Data from OpenRouter shows Chinese models gaining ground over the past year, with DeepSeek, Tencent, Z.ai, and Qwen now accounting for roughly 42% of requests on the platform.

If cheaper open-weight models are sufficient for customers' needs, the US labs face an unenviable choice: cut prices and sacrifice margins or keep spending to maintain a technical lead that customers may not be willing to pay for.



By calling attention to existential-risk claims, the labs may be hoping to revive the Trump administration's currently shelved restrictions on Chinese models - DeepSeek

Restrictions on Chinese models could offer a way out. By calling attention to existential-risk claims, the labs may be hoping to revive the Trump administration's currently shelved restrictions on Chinese models, thus protecting themselves from cheaper competition.

But if a company is invoking dangers great enough to justify slowing development or excluding competitors, it should at least have to submit its systems to independent regulatory scrutiny.

Otherwise, it retains discretion over both the claims of danger and the response, allowing safety policy to serve its own financial interests.

The July Hugging Face breach, carried out by OpenAI "agents" that escaped their testing environment, makes independent scrutiny urgent.

No sci-fi maximalism is necessary to appreciate the dangers that current systems pose as brute-force hacking instruments.

Cooperating with regulators

Industry leaders have long promised to cooperate with regulators. In his 2015 email proposing the creation of what would become OpenAI, Altman assured Musk: “Obviously we’d comply with/aggressively support all regulation.”

But support in principle leaves the terms of scrutiny unresolved. Anthropic’s denial of pre-launch access to Mythos 5.1 for the United Kingdom-based AI Security Institute shows how much discretion remains with the company.

> The strategy resembles ExxonMobil’s support for carbon taxes

The strategy resembles ExxonMobil’s support for carbon taxes. While making a show of endorsing regulation in principle, it bet that policymakers would decline to impose meaningful constraints.

The warnings need not be insincere to serve the companies’ financial interests. Replacing workers promises revenues sufficient to justify their valuations, and existential risk supplies a rationale for containing their costs and restricting the competition.

The coming AI apocalypse is the best reason to buy their stock.

Brian Judge is Research Director of the Program on Finance and Democracy at the University of California, Berkeley.