

# GPT-4's Secret Has Been Revealed



The new GPT-4

GPT-4 was the most anticipated AI model in history. Yet when OpenAI released it in March they didn't tell us anything about its size, data, internal structure, or how they trained and built it. A true black box.

As it turns out, they didn't conceal those critical details because the model was too innovative or the architecture too moat-y to share. The opposite seems to be true if we're to believe the latest rumors:

GPT-4 is, technically and scientifically speaking, hardly a breakthrough.

That's not necessarily bad – GPT-4 is, after all, the best language model in existence – just... somewhat underwhelming. Not what people were expecting after a 3-year wait.

This news, yet to be officially confirmed, reveals key insights about GPT-4 and OpenAI and raises questions about AI's true state-of-the-art – and its future.

## GPT-4: A mixture of smaller models

On June 20th, George Hotz, founder of self-driving startup Comma.ai leaked that GPT-4 isn't a single monolithic dense model (like GPT-3 and GPT-3.5) but a mixture of 8 x 220-billion-parameter models. Later that day, Soumith Chintala, co-founder of PyTorch at Meta, reaffirmed the leak. Just the day before, Mikhail Parakhin, Microsoft Bing AI lead, had also hinted at this.

GPT-4 is not one big >1T model but eight smaller ones cleverly put together. The mixture of experts paradigm OpenAI supposedly used for this "hydra"

model is neither new nor invented by them. In this article, I'll explain why this is very relevant for the field and how OpenAI masterfully executed its plan to achieve three key goals.

#### Two caveats.

First, this is a rumor. The explicit sources (Hotz and Chintala) are robust but not OpenAI staff. Parakhin holds an executive position at Microsoft but he never confirmed it explicitly. For these reasons, it's worth taking this with a grain of salt. The story is nevertheless very plausible.

Second, let's give credit where credit's due. GPT-4 is exactly as impressive as users say. The details of the internal architecture can't change that. If it works, it works. It doesn't matter whether it's one model or eight tied together. Its performance and ability on writing and coding tasks are legit. This article is not a dunk on GPT-4 - just a warning that we may want to update our priors.

This article is a selection from The Algorithmic Bridge, an educational newsletter whose purpose is to bridge the gap between AI, algorithms, and people. It will help you understand the impact AI has in your life and develop the tools to better navigate the future.

#### Subscribe to The Algorithmic Bridge

Bridging the gap between AI and people. Join 13,000 others, including professionals from the Big Five: Google, Amazon... [thealgorithmicbridge.substack.com](http://thealgorithmicbridge.substack.com)

#### The secrecy around GPT-4

I have to applaud OpenAI's mastery in dealing with the unreasonably high expectations that surrounded GPT-4 by covering up the more unsatisfactory aspects of the model while remaining at the top of the conversation.

In January, when Connie Loizos of StrictlyVC mentioned the ridiculous 100-trillion GPT-4 graphs that were making the rounds on Twitter, Altman told her that "people are begging to be disappointed and they will be." He knew GPT-4, which had finished training in the summer of 2022, wouldn't meet people's expectations.

But he didn't want to kill OpenAI's almost-mystical reputation. So they hid GPT-4 from public scrutiny, further fueling its mysterious aura.

OpenAI had already crystallized its status with ChatGPT by then. They were leaders in the space in the eyes of the majority (despite Google's longer and richer history of AI R&D). As such, they couldn't admit explicitly that GPT-4 wasn't the anticipated breakthrough - and the huge leap from GPT-3 - that people wanted.

So they focused on hinting and implying it was really powerful (e.g., sparks of AGI, superintelligence is near, and all that) and defended their decision to not disclose GPT-4's specs by alluding to increased competitive pressures, as Ilya Sutskever told The Verge.

With this on the table, the mainstream reading of OpenAI's secrecy was along the lines of: "They won't disclose the specs because they can't afford Google or open source initiatives to copy them due to business survival and safety reasons. Also, GPT-4's SOTA performance implies the architecture must be a scientific feat."

OpenAI got what it wanted. Altman was honest - GPT-4 would've been disappointing - but, at the same time, the subliminal signals suggested something else: GPT-4 is magical. And people believed it. It is magical in a way, though. We've all seen it in

action. It's just not what most people would perceive as a revolutionary achievement. It seems to be just an old trick reimagined. Combining several expert models into one, with each expert trained to specialize in separate areas, tasks, or data was a technique first successfully implemented in 2021. Two years ago. Who did it? You guessed it, Google engineers (some of them, like William Fedus and Trevor Cai, were later hired by OpenAI).

OpenAI surely added engineering ingenuity on top (otherwise Google would have their own GPT-4, or better), but the very key to the model's absolute dominance across benchmarks is simply that it's not one model but eight.

So, yes, GPT-4 is magic, but OpenAI made it into the kind we see in shows. A clever mix of skillful misdirection and smooth sleight of hand. And the trick is merely a remake.

### **The 3 goals OpenAI achieved by hiding GPT-4**

First, they freed people's imagination. Although skeptics saw this as an unscientific practice, it fueled speculation about the model's power. This, in turn, allowed them to establish their preferred narrative – AGI and the need to plan for it – convincing the government that safety requirements (especially for others) and regulation (that which fits their goals) are paramount. The illusion was complete: GPT-4 had a shiny appearance so it had to be equally shiny inside – and shiny can be dangerous. In actuality, if we go for the snarky analogy, GPT-4 is better portrayed as a gaze of “raccoons in a trenchcoat.”

Second, they effectively prevented open source initiatives, as well as competitors like Google or Anthropic, from copying the techniques they had

supposedly invented or discovered. But OpenAI had no moat in GPT-4. LLaMA is unable to compete with GPT-4, but maybe 8 LLaMAs tied together could – people were comparing apples to oranges but they didn't know. So maybe I was mistaken and open source wasn't so far behind after all.

The moat was making GPT-4 appear more impressive than it was.

Finally, they concealed the truth that GPT-4 is actually not that much of an AI breakthrough, effectively preventing witnesses, outsiders, and users from losing faith in the apparently breakneck pace of progress in the field. If we're nitpicky, GPT-4 is the result of having, on the one hand, enough money and GPUs to train and run eight ~GPT-3.5 models stacked together and, on the other, the audacity to dust an old technique invented by another company without telling anyone.

GPT-4 was a business marketing masterclass.

### **A final thought**

Maybe OpenAI – and the industry at large – are out of ideas, as Hotz suggests. Maybe AI isn't really going that fast milestone after milestone as companies, media, marketers, and arXiv make it seem. Maybe GPT-4 isn't as huge a leap from GPT-3 as it should've been.

A rumor is still a rumor until we get an official version (I reached out to OpenAI but didn't hear back yet). It's hard to deny the plausibility of the story, though. Besides the value of the sources, there's an overall coherence to it. That's why I'm giving this news a high credibility.

Quoting Hotz's conclusion: “Whenever a company is secretive is because they are hiding something that's not that cool.” Maybe GPT-4 is not that cool after all