

# Prophecies of the Flood

What to make of the statements of the AI labs?

ETHAN MOLLICK

JAN 10, 2025

Recently, something shifted in the AI industry. Researchers began speaking urgently about the arrival of supersmart AI systems, a flood of intelligence. Not in some distant future, but imminently. They often refer to AGI - Artificial General Intelligence - defined, albeit imprecisely, as machines that can outperform expert humans across most intellectual tasks. This availability of intelligence on demand will, they argue, change society deeply and will change it soon.

## Sam Altman, CEO of OpenAI 1/6/25

We are now confident we know how to build AGI as we have traditionally understood it. We believe that, in 2025, we may see the first AI agents "join the workforce" and materially change the output of companies. We continue to believe that iteratively putting great tools in the hands of people leads to great, broadly-distributed outcomes.

We are beginning to turn our aim beyond that, to superintelligence in the true sense of the word. We love our current products, but we are here for the glorious future. With superintelligence, we can do anything else. Superintelligent tools could massively accelerate scientific discovery and innovation well beyond what we are capable of doing on our own, and in turn massively increase abundance and prosperity.

This sounds like science fiction right now, and somewhat crazy to even talk about it. That's alright—we've been there before and we're OK with being there again. We're pretty confident that in the next few years, everyone will see what we see, and that the need to act with great care, while still maximizing broad benefit and empowerment, is so important. Given the possibilities of our work, OpenAI cannot be a normal company.

## Dario Amodei, CEO of Anthropic 10/24/24

What powerful AI (I dislike the term AGI) <sup>3</sup> will look like, and when (or if) it will arrive, is a huge topic in itself. It's one I've discussed publicly and could write a completely separate essay on (I probably will at some point). Obviously, many people are skeptical that powerful AI will be built soon and some are skeptical that it will ever be built at all. **I think it could come as early as 2026**, though there are also ways it could take much longer. But for the purposes of this essay, I'd like to put these issues aside, assume it will come reasonably soon, and focus on what happens in the 5-10 years after that. I also want to assume a definition of what such a system *will look like*, what its capabilities are and how it interacts, even though there is room for disagreement on this.

By *powerful AI*, I have in mind an AI model—likely similar to today's LLM's in form, though it might be based on a different architecture, might involve several interacting models, and might be trained differently—with the following properties:

- In terms of pure intelligence <sup>4</sup>, it is smarter than a Nobel Prize winner across most relevant fields – biology, programming, math, engineering, writing, etc. This means it can prove unsolved mathematical theorems, write extremely good novels, write difficult codebases from scratch, etc.

## Google Deepmind researcher, 1/6

Vedant Misra  
@vedantmisra

There are maybe a few hundred people in the world who viscerally understand what's coming. Most are at DeepMind / OpenAI / X but some are on the outside. You have to be able to forecast aggregate effect of rapid algorithmic improvement, aggressive investment in building RL environments for iterative self-improvement and many tens of billions already committed to building data. Either we're all wrong, or everything is about to change.

Stephen McAleer  
@McAleerStephen

It's hard to convey my views without sounding like an AI grifter

I will say this: many researchers at frontier labs are taking the of short timelines very seriously, and virtually nobody outside talking enough about the safety implications.

11:21 AM · Jan 9, 2025 · 367 retweets

Joshua Achiam  
@jachiam0 · Jan 5

It feels outside of the overtone window right now to suggest that so much change could happen very quickly, or even to realistically grapple with those changes might entail. It is too easy to say "the present is more and more real"

Head of Mission Alignment, OpenAI

Joshua Achiam  
@jachiam0 · Jan 5

Nonetheless, change is coming. It will be reflected first in the prices of goods and labor. It will force changes in strategy in businesses, institutions of all kinds, and countries.

2 replies · 8 retweets · 106 likes · 6.8K views

A sample of the recent statements by prominent researchers within AI labs predicting near-term supersmart AIs.

There are plenty of reasons to [not believe insiders](#) as they have clear incentives to make bold predictions: they're raising capital, boosting stock valuations, and perhaps convincing themselves of their own historical importance. They're technologists, not

prophets, and the track record of technological predictions is littered with confident declarations that turned out to be decades premature. Even setting aside these human biases, the underlying technology itself gives us reason for doubt. Today's Large Language Models, despite their impressive capabilities, remain fundamentally inconsistent tools - brilliant at some tasks while stumbling over seemingly simpler ones. This “jagged frontier” is a core characteristic of current AI systems, one that won't be easily smoothed away.

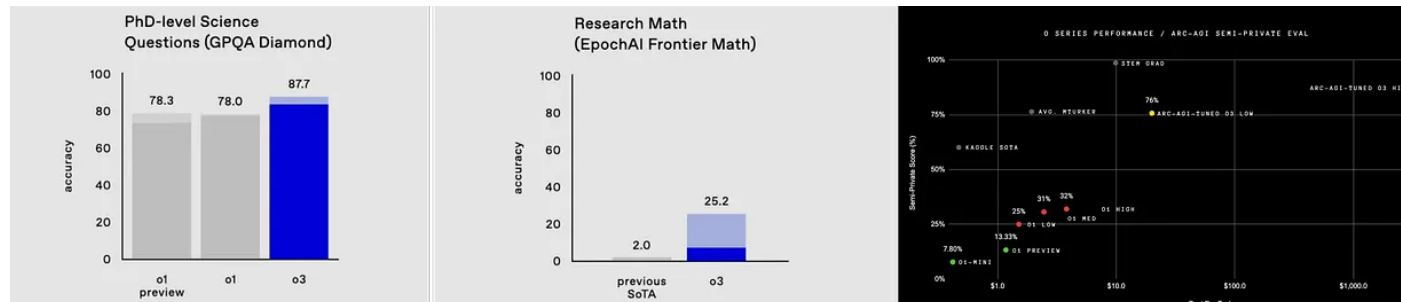
Plus, even assuming researchers are right about reaching AGI in the next year or two, they are likely overestimating the speed at which humans can adopt and adjust to a new technology. Changes to organizations take a long time. Changes to systems of work, life, and education, are slower still. And technologies need to find specific uses that matter in the world, which is itself a slow process. We could have AGI right now and most people wouldn't notice (indeed, some observers have suggested that has already happened, arguing that the latest AI models like Claude 3.5 are effectively AGI [1](#)).

Yet dismissing these predictions as mere hype may not be helpful. Whatever their incentives, the researchers and engineers inside AI labs appear genuinely convinced they're witnessing the emergence of something unprecedented. Their certainty alone wouldn't matter - except that increasingly public benchmarks and demonstrations are beginning to hint at why they might believe we're approaching a fundamental shift in AI capabilities. The water, as it were, seems to be rising faster than expected.

## Where the water is rising

The event that kicked off the most speculation was the reveal of a new model by OpenAI called o3 in late December. No one outside of OpenAI has really used this system yet, but it is the successor to o1, [which is already very impressive](#) [2](#). The o3 model is one of the new generation of “reasoners” - AI models that take extra time to “think” before answering questions, which greatly improves their ability to solve hard

problems. OpenAI provided a number of startling benchmarks for o3 that suggest a large advance over o1, and, indeed, over where we thought the state-of-the-art in AI was. Three benchmarks, in particular, deserve a little attention.

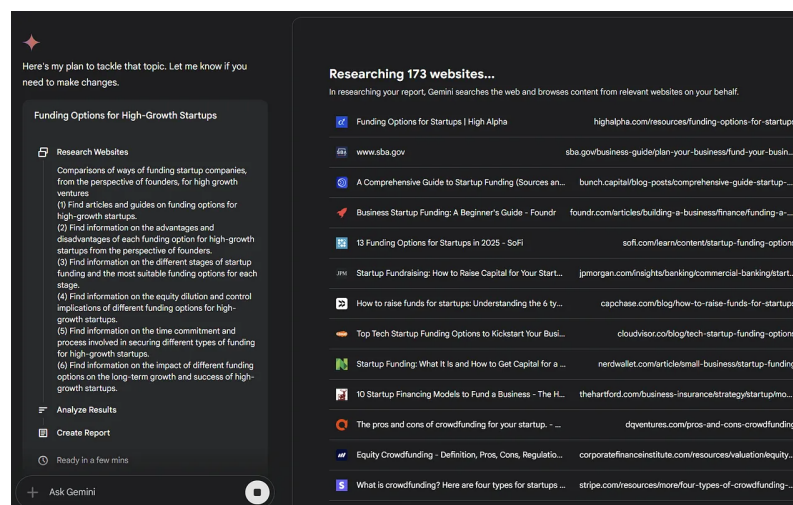


The first is the called the Graduate-Level Google-Proof Q&A test ([GPQA](#)), and it is supposed to test high-level knowledge with a series of multiple-choice problems that even Google can't help you with. PhDs with access to the internet got 34% of the questions right on this test outside their specialty, and 81% right inside their specialty. When tested, o3 achieved 87% beating human experts for the first time. The second is [Frontier Math](#), a set of private math problems created by mathematicians to be incredibly hard to solve, and, indeed, no AI ever scored higher than 2%, until o3, which got 25% right. The final benchmark is [ARC-AGI](#), a rather famous test of fluid intelligence that was designed to be relatively easy for humans but hard for AIs. As [o3 beat all previous AIs as well as the baseline human level on the test](#), scoring 87.5%. All of these tests come with significant caveats<sup>3</sup> but they suggest that what we previously considered unpassable barriers to AI performance may actually be beaten quite quickly.

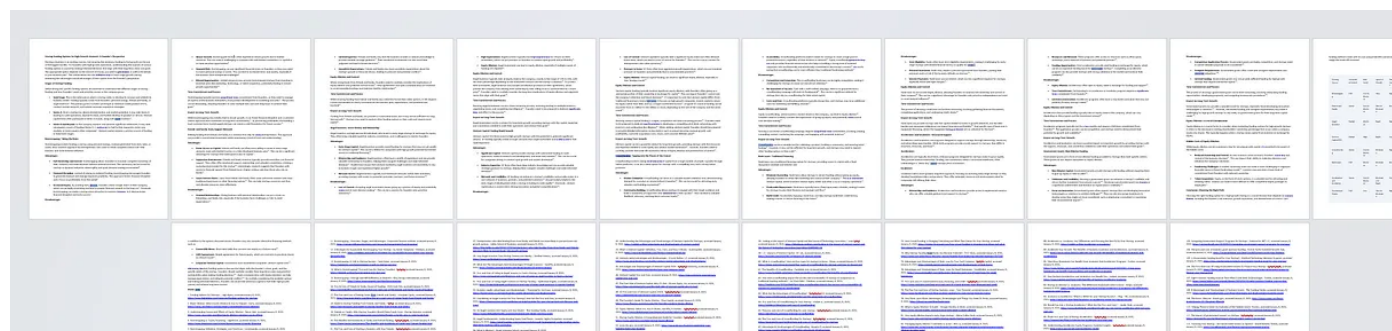
## Agents

As AIs get smarter, they become more effective agents, another ill-defined term (see pattern?) that generally means an AI given the ability to act autonomously towards achieving a set of goals. [I have demonstrated some of the early agentic systems in previous posts](#), but I think the past few weeks have also shown us that practical agents at least for narrow but economically important areas, are now viable.

A nice example of that is Google's Gemini with Deep Research (accessible to everyone who subscribes to Gemini), which is really a specialized research agent. I gave it a topic like "research a comparison of ways of funding startup companies, from the perspective of founders, for high growth ventures." And the agentic system came up with a plan, read through 173(!) websites and compiled a report for me with the answer a few minutes later.



The result was a 17 page paper with 118 references! But is it any good? I have taught the introductory entrepreneurship class at Wharton for over a decade, published on the topic, started companies myself, and even wrote a book on entrepreneurship, and I think this is pretty solid. I didn't spot any obvious errors, but [you can read it yourself](#) [you would like here](#). The biggest issue is not accuracy, but that the agent is limited to public non-paywalled websites, and not scholarly or premium publications. It also is a bit shallow and does not make strong arguments in the face of conflicting evidence not as good as the best humans, but better than a lot of reports that I see.





Still, this is a genuinely disruptive example of an agent with real value. Researching and report writing is a major task of many jobs. What Deep Research accomplished in three minutes would have taken a human many hours, though they might have added more nuanced analysis. Given that, anyone writing a research report should probably try Deep Research and see how it works as a starting place, even though a good final report will still require a human touch. I had a chance to speak with the leader of the Deep Research project, where I learned that it is just a pilot project from a small team. I thus suspect that other groups and companies that were highly incentivized to create narrow but effective agents would be able to do so. Narrow agents are now a real product, rather than a future possibility. There are already many coding agents, and you can use experimental open-source agents that do [scientific](#) and [financial](#) research.

Narrow agents are specialized for a particular task, which means they are somewhat limited. That raises the question of whether we soon see generalist agents where you can just ask the AI anything and it will use a computer and the internet to do it. [Simon Willison thinks not](#) despite what Sam Altman has argued. We will learn more as the year progresses, but if general agentic systems work reliably and safely, that really will change things, as it allows smart AIs to take action in the world.

## Many smaller advances are happening

Agents and very smart models are the core elements needed for transformative AI, but there are many other pieces as well that seem to be making rapid progress. This includes advances in how much AIs can remember (context windows) and multimodal capabilities that allow them to see and speak. It can be helpful to look back a little to get a sense of progress. For example, I have been testing the prompt “otter on a plane using wifi” for image and video models since before ChatGPT came out. In October 2023, that prompt got you this terrifying monstrosity.



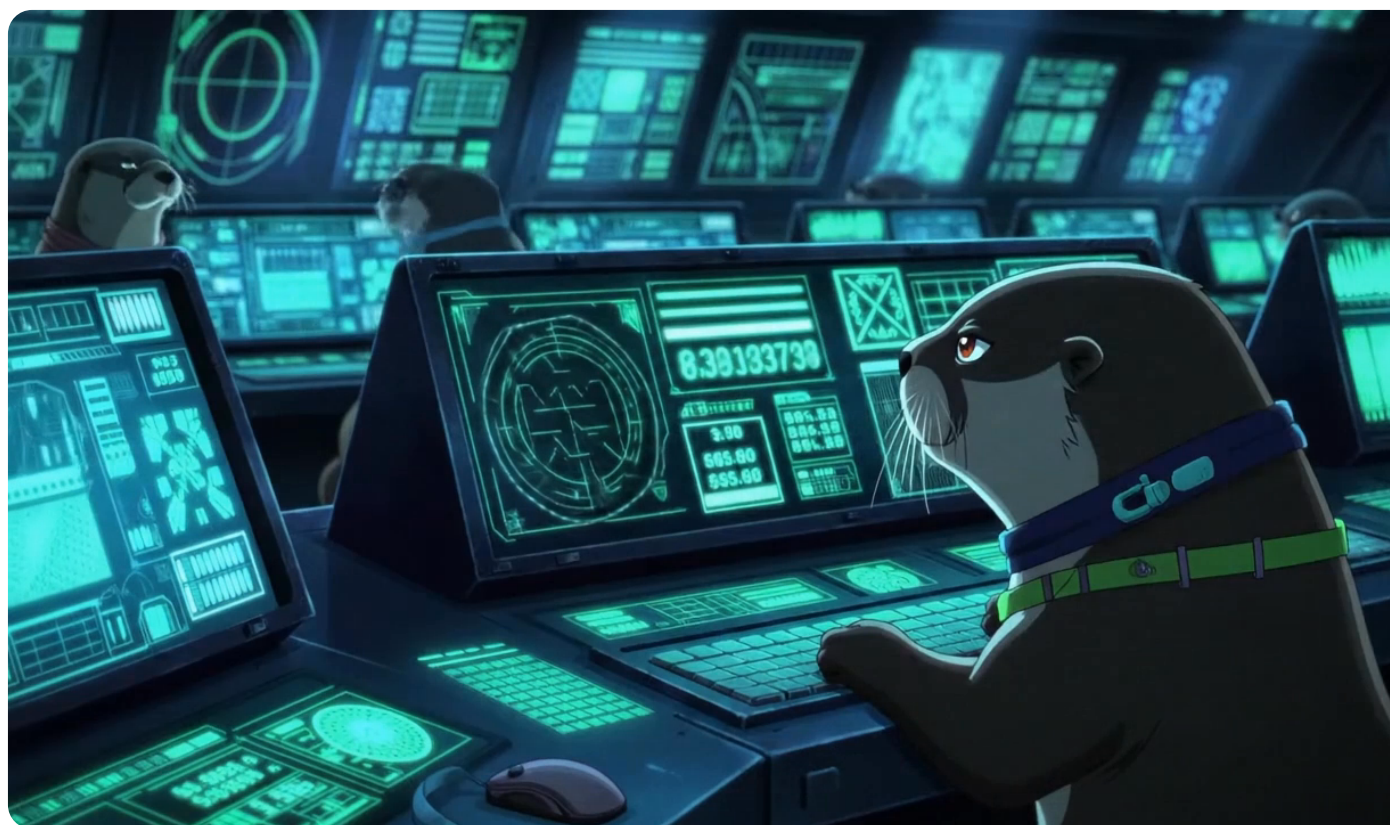


“Otter on a plane using wifi” October, 2023

Less than 18 months later, multiple image creation tools nail the prompt. The result that I have had to figure out something more challenging (this is an example of benchmark saturation, where old benchmarks get beaten by the AI). I decided to take a few minutes and see how far I could get with Google’s Veo 2 video model in producing a movie of the otter’s journey. The video you see below took less than 15 minutes of active work, although I had to wait a bit for the videos to be created. Take a look at the quality of the shadows and light. I especially appreciate how the otter opens the computer at the end.



And, to up the ante even further, I decided to turn the saga of the otter into a 1980s style science fiction anime featuring otters in space and a period-appropriate theme song (thanks to Suno). Again, very little (human) work was involved.



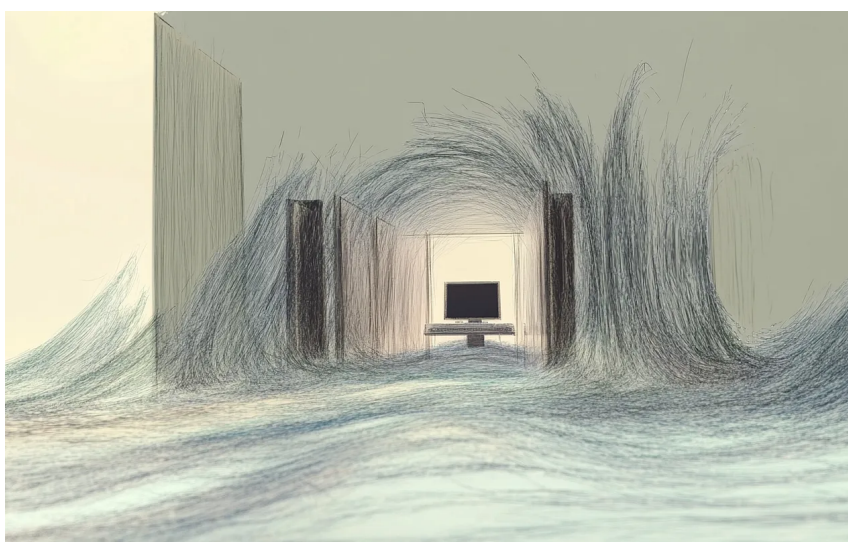
## What of the flood?

Given all of this, how seriously should we take the claims of the AI labs that a flood of intelligence is coming? Even if we only consider what we've already seen - the o3 benchmarks shattering previous barriers, narrow agents conducting complex research and multimodal systems creating increasingly sophisticated content - we're looking at capabilities that could transform many knowledge-based tasks. And yet the labs insist this is merely the start, that far more capable systems and general agents are imminent.

What concerns me most isn't whether the labs are right about this timeline - it's that we're not adequately preparing for what even current levels of AI can do, let alone the chance that they might be correct. While AI researchers are focused on alignment, ensuring AI systems act ethically and responsibly, far fewer voices are trying to envision and articulate what a world awash in artificial intelligence might actually

look like. This isn't just about the technology itself; it's about how we choose to shape and deploy it. These aren't questions that AI developers alone can or should answer. They're questions that demand attention from organizational leaders who will need to navigate this transition, from employees whose work lives may transform, and from stakeholders whose futures may depend on these decisions. The flood of intelligence that may be coming isn't inherently good or bad - but how we prepare for it, how we adapt to it, and most importantly, how we choose to use it, will determine whether it becomes a force for progress or disruption. The time to start having these conversations isn't after the water starts rising - it's now.

Type your email...



- 1 I asked Claude to read over the completed document and give me feedback and it wrote: “The parenthetical comment about Claude 3.5 could potentially benefit from an update or revision since it's mentioned as an example of potential AGI. As Claude 3.5 Sonnet, I should note that I can't verify specific claims about my capabilities in relation to AGI.”
- 2 They skipped the name o2 because it is the name of a telephone company in the UK, AI naming continues to be very bad.
- 3 The caveat for GPQA is that the data is publicly available, and it is possible that the mod

trained on that data, either by accident or on purpose, although there is no indication that it did so. The caveat from the Frontier Math test is that the problems differ in difficulty level: Tier 1 are hard Math Olympiad problems, Tier 2 are graduate level problems, and Tier 3 are genuine research-level problems. [In the words of the mathematician in charge](#), of o3's correct answers: "40% have been Tier 1, 50% Tier 2, and 10% Tier 3. However, most Tier 3 "solutions"—and many Tier 2 ones—stem from heuristic shortcuts rather than genuine mathematical understanding." The caveat for ARC-AGI is that it required a lot of very expensive computer time for o3 to run long enough to achieve its high score.