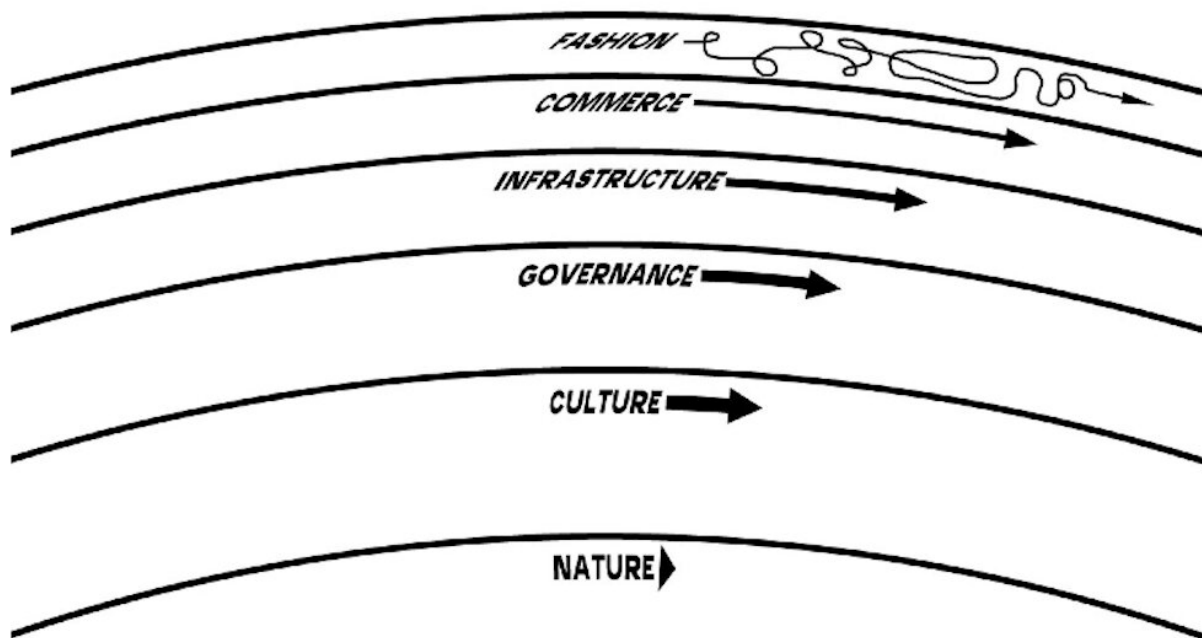# Opinion: I used to work at Google and now I'm an AI researcher. Here's why slowing down AI development is wise

April 4 2023, by Rodolfo Ocampo

According to the 'pace layers' model, different layers of a healthy civilisation move at different speeds, from the slow movement of nature to the rapid shifts of fashion. Credit: [Stewart Brand / Journal of Design and Science](#)

Is it time to put the brakes on the development of artificial intelligence (AI)? If you've quietly asked yourself that question, you're not alone.

In the past week, a host of AI luminaries signed an open letter calling for a six-month pause on the development of more powerful models than GPT-4; European researchers called for tighter AI regulations; and long-time AI researcher and critic Eliezer Yudkowsky demanded a complete shutdown of AI development in the pages of TIME magazine.

Meanwhile, the industry shows no sign of slowing down. In March, a senior AI executive at Microsoft reportedly spoke of "very, very high" pressure from chief executive Satya Nadella to get GPT-4 and other new models to the public "at a very high speed."

I worked at Google until 2020, when I left to study responsible AI development, and now I research human-AI creative collaboration. I am excited about the potential of artificial intelligence, and I believe it is already ushering in a new era of creativity. However, I believe a temporary pause in the development of more powerful AI systems is a good idea. Let me explain why.

What is GPT-4 and what is the letter asking for?

The open letter published by the US non-profit Future of Life Institute makes a straightforward request of AI developers:

"We call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4."

So what is GPT-4? Like its predecessor GPT-3.5 (which powers the popular ChatGPT chatbot), GPT-4 is a kind of generative AI software called a "large language model," developed by OpenAI.

GPT-4 is much larger and has been trained on significantly more data. Like other large language models, GPT-4 works by guessing the next word in response to prompts—but it is nonetheless incredibly capable.

In tests, it passed legal and medical exams, and can write software better than professionals in many cases. And its full range of abilities is [yet to be discovered](#).

Good, bad, and plain disruptive

GPT-4 and models like it are likely to have huge effects across many layers of society.

On the upside, they could enhance human creativity and scientific discovery, lower barriers to learning, and be used in [personalized educational tools](#). On the downside, they could facilitate personalized phishing attacks, produce disinformation at scale, and be used to hack through the network security around computer systems that control [vital infrastructure](#).

OpenAI's [own research](#) suggests models like GPT-4 are "general-purpose technologies" which will impact some 80% of the US workforce.

Layers of civilization and the pace of change

The US writer Stewart Brand has [argued](#) that a "healthy civilization" requires different systems or layers to move at different speeds:

"The fast layers innovate; the slow layers stabilize. The whole combines learning with continuity."

In Brand's "pace layers" model, the bottom layers change more slowly than the top layers.

Technology is usually placed near the top, somewhere between fashion

and commerce. Things like regulation, economic systems, security guardrails, ethical frameworks, and other aspects exist in the slower governance, infrastructure and culture layers.

Right now, technology is accelerating much faster than our capacity to understand and regulate it—and if we're not careful it will also drive changes in those lower layers that are too fast for safety.

The US sociobiologist E.O. Wilson described the dangers of a mismatch in the different paces of change like so:

"The real problem of humanity is the following: we have Paleolithic emotions, medieval institutions, and god-like technology."

Are there good reasons to maintain the current rapid pace?

Some argue that if top AI labs slow down, other unaligned players or countries like China will outpace them.

However, training complex AI systems is not easy. OpenAI is ahead of its US competitors (including Google and Meta), and developers in China and other countries also lag behind.

It's unlikely that "rogue groups" or governments will surpass GPT-4's capabilities in the foreseeable future. Most AI talent, knowledge, and computing infrastructure is concentrated in a handful of top labs.

Other critics of the Future of Life Institute letter say it relies on an overblown perception of current and future AI capabilities.

However, whether or not you believe AI will reach a state of general

superintelligence, it is undeniable that this technology will impact many facets of human society. Taking the time to let our systems adjust to the pace of change seems wise.

Slowing down is wise

While there is plenty of room for disagreement over specific details, I believe the Future of Life Institute letter points in a wise direction: to take ownership of the pace of technological change.

Despite what we have seen of the disruption caused by social media, Silicon Valley still tends to follow Facebook's infamous motto of "move fast and break things."

I believe a wise course of action is to slow down and think about where we want to take these technologies, allowing our systems and ourselves to adjust and engage in diverse, thoughtful conversations. It is not about stopping, but rather moving at a sustainable pace of progress. We can choose to steer this technology, rather than assume it has a life of its own that we can't control.

After some thought, I have added my name to the list of signatories of the open letter, which the Future of Life Institute says now includes some 50,000 people. Although a six-month moratorium won't solve everything, it would be useful: it sets the right intention, to prioritize reflection on benefits and risks over uncritical, accelerated, profit-motivated progress.

This article is republished from The Conversation under a Creative Commons license. Read the original article.

Provided by The Conversation