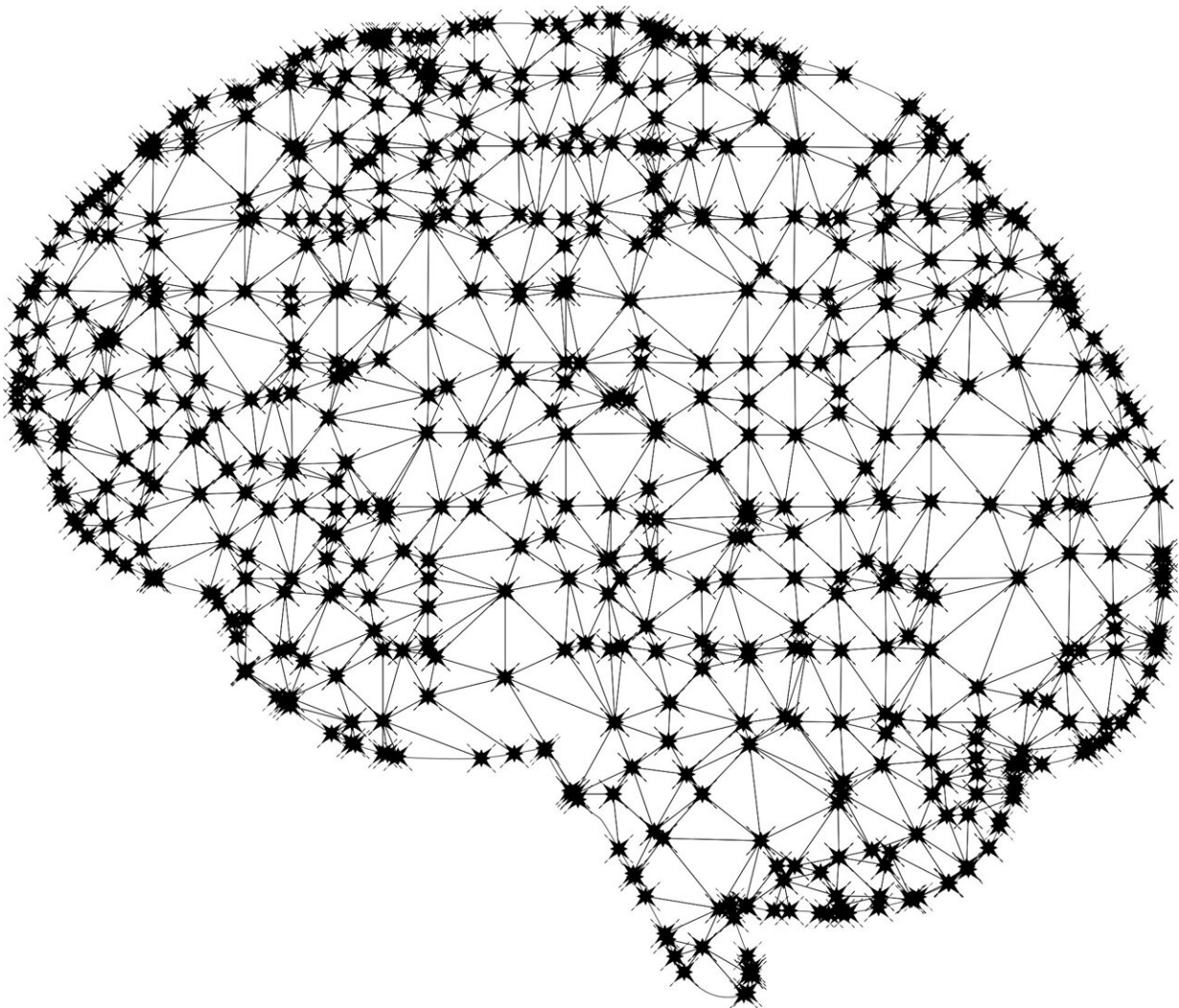


AI exemplifies the 'free rider' problem—here's why that points to regulation

May 5 2023, by Tim Juvshik



Credit: Pixabay/CC0 Public Domain

On March 22, 2023, thousands of researchers and tech leaders—including Elon Musk and Apple co-founder Steve Wozniak—published an [open letter](#) calling to slow down the artificial intelligence race. Specifically, the letter recommended that labs pause training for technologies stronger than OpenAI's GPT-4, the [most sophisticated generation](#) of today's language-generating AI systems, for at least six months.

Sounding the alarm on [risks posed by AI](#) is nothing new—academics have issued warnings about the risks of [superintelligent machines](#) for decades now. There is still no consensus about [the likelihood of creating artificial general intelligence](#), autonomous AI systems that match or [exceed humans](#) at most economically valuable tasks. However, it is clear that current AI systems already pose plenty of dangers, from racial bias in [facial recognition technology](#) to the increased threat of misinformation and [student cheating](#).

While the letter calls for industry and policymakers to cooperate, there is currently no mechanism to enforce such a pause. As [a philosopher who studies technology ethics](#), I've noticed that AI research exemplifies the "[free rider problem](#)." I'd argue that this should guide how societies respond to its risks—and that [good intentions](#) won't be enough.

Riding for free

Free riding is a common consequence of what philosophers call "collective action problems." These are situations in which, as a group, everyone would benefit from a particular action, but as individuals, each member would [benefit from not doing it](#).

Such problems most commonly involve [public goods](#). For example,

suppose a city's inhabitants have a collective interest in funding a subway system, which would require that each of them pay a small amount through taxes or fares. Everyone would benefit, yet it's in each individual's best interest to save money and avoid paying their fair share. After all, they'll still be able to enjoy the subway if most other people pay.

Hence the "free rider" issue: Some individuals won't contribute their fair share but will still get a "free ride"—literally, in the case of the subway. If every individual failed to pay, though, no one would benefit.

Philosophers tend to argue that [it is unethical to "free ride,"](#) since free riders fail to reciprocate others' paying their fair share. Many philosophers also argue that free riders fail in their responsibilities as part of [the social contract](#), the collectively agreed-upon cooperative principles that govern a society. In other words, they fail to uphold their duty to be contributing members of society.

Hit pause, or get ahead?

Like the subway, AI is a public good, given its potential to complete tasks far more efficiently than human operators: everything from [diagnosing patients](#) by analyzing medical data to taking over [high-risk jobs in the military](#) or [improving mining safety](#).

But both its benefits and dangers will affect everyone, even people who don't personally use AI. To reduce [AI's risks](#), everyone has an interest in the industry's research being conducted carefully, safely and with proper oversight and transparency. For example, misinformation and fake news already pose serious threats to democracies, but AI has the potential to [exacerbate the problem](#) by spreading "fake news" faster and more effectively than people can.

Even if some [tech companies](#) voluntarily halted their experiments, however, other corporations would have a monetary interest in continuing their own AI research, allowing them to get ahead in the AI arms race. What's more, voluntarily pausing AI experiments would allow other companies to get a free ride by eventually reaping the benefits of safer, more transparent AI development, along with the rest of society.

Sam Altman, CEO of OpenAI, has acknowledged that the company [is scared of the risks](#) posed by its chatbot system, ChatGPT. "We've got to be careful here," he said in an interview with ABC News, mentioning the potential for AI to produce misinformation. "I think people should be happy that we are a little bit scared of this."

In a letter published April 5, 2023, OpenAI said that the company believes powerful AI systems [need regulation](#) to ensure thorough safety evaluations and that it would "actively engage with governments on the best form such regulation could take." Nevertheless, OpenAI is continuing with the [gradual rollout](#) of GPT-4, and the rest of the industry is also continuing to develop and train advanced AIs.

Ripe for regulation

Decades of [social science research](#) on collective action problems has shown that where trust and goodwill are insufficient [to avoid free riders](#), regulation is often the only alternative. Voluntary compliance is the key factor that creates free-rider scenarios—and [government action](#) is at times the way to nip it in the bud.

Further, [such regulations must be enforceable](#). After all, would-be subway riders might be unlikely to pay the fare unless there were a threat of punishment.

Take one of the most dramatic free-rider problems in the world today:

[climate change](#). As a planet, we all have a high-stakes interest in maintaining a habitable environment. In a system that allows free riders, though, the incentives for any one country to actually follow greener guidelines are slim.

[The Paris Agreement](#), which is currently the most encompassing global accord on climate change, is voluntary, and the United Nations has no recourse to enforce it. Even if the European Union and China voluntarily limited their emissions, for example, the United States and India could "free ride" on the reduction of carbon dioxide while continuing to emit.

Global challenge

Similarly, the free-rider problem grounds arguments to regulate AI development. In fact, [climate change](#) is a particularly close parallel, since neither the risks posed by AI nor greenhouse gas emissions are restricted to a program's country of origin.

Moreover, the race to develop more advanced AI is an international one. Even if the U.S. introduced federal regulation of AI research and development, China and Japan could ride free and continue their own domestic [AI programs](#).

Effective regulation and enforcement of AI would require global collective action and cooperation, just as with [climate change](#). In the U.S., [strict enforcement](#) would require federal oversight of research and the ability to impose hefty fines or shut down noncompliant AI experiments to ensure responsible development—whether that be through regulatory oversight boards, whistleblower protections or, in extreme cases, laboratory or research lockdowns and criminal charges.

Without enforcement, though, there will be free riders—and free riders mean the AI threat won't abate anytime soon.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: AI exemplifies the 'free rider' problem—here's why that points to regulation (2023, May 5) retrieved 19 April 2024 from <https://techxplore.com/news/2023-05-ai-exemplifies-free-rider-problemhere.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.