

All Open Letters

Pause Giant AI Experiments: An Open Letter

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

Signatures

26156

Add your signature

PUBLISHED

March 22, 2023



AI systems with human-competitive intelligence can pose profound risks to society and humanity, as shown by extensive research^[1] and acknowledged by top AI labs.^[2] As stated in the widely-endorsed Asilomar AI Principles, *Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources*. Unfortunately, *this level of planning and management is not happening*, even though recent months have seen AI labs locked in an out-of-control race to develop and deploy ever more powerful digital minds that no one – not even their creators – can understand, predict, or reliably control.

Contemporary AI systems are now becoming human-competitive at general tasks,^[3] and we must ask ourselves: *Should we let machines flood our information channels with propaganda and untruth? Should we automate away all the jobs, including the fulfilling ones? Should we develop nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us? Should we risk loss of control of our civilization? Such decisions must not be delegated to unelected tech leaders. Powerful AI systems should be developed only once we are confident that their effects will be positive and their risks will be manageable*. This confidence must be well justified and increase with the magnitude of a system's potential effects. OpenAI's recent statement regarding artificial general intelligence, states that *"At some point, it may be important to get independent review before starting to train future systems, and for the most advanced efforts to agree to limit the rate of growth of compute used for creating new models."* We agree. That point is now.

Therefore, we call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4. This pause should be public and verifiable, and include all key actors. If such a pause cannot be enacted quickly, governments should step in and institute a moratorium.

AI labs and independent experts should use this pause to jointly develop and implement a set of shared safety protocols for advanced AI design and development that are rigorously audited and overseen by independent outside experts. These protocols should ensure that systems adhering to them are safe beyond a reasonable doubt.^[4] This does *not* mean a pause on AI development in general, merely a stepping back from the dangerous race to ever-larger unpredictable black-box models with emergent capabilities.

AI research and development should be refocused on making today's powerful, state-of-the-art systems more accurate, safe, interpretable, transparent, robust, aligned, trustworthy, and loyal.

In parallel, AI developers must work with policymakers to dramatically accelerate development of robust AI governance systems. These should at a minimum include: new and capable regulatory authorities dedicated to AI; oversight and tracking of highly capable AI systems and large pools of computational capability; provenance and watermarking systems to help distinguish real from synthetic and to track model leaks; a robust auditing and certification ecosystem; liability for AI-caused harm; robust public funding for technical AI safety research; and well-resourced institutions for coping with the dramatic economic and political disruptions (especially to democracy) that AI will cause.

Humanity can enjoy a flourishing future with AI. Having succeeded in creating powerful AI systems, we can now enjoy an "AI summer" in which we reap the rewards, engineer these systems for the clear benefit of all, and give society a chance to adapt. Society has hit pause on other technologies with potentially catastrophic effects on society.^[5] We can do so here. Let's enjoy a long AI summer, not rush unprepared into a fall.

We have prepared some FAQs in response to questions and discussion in the media and elsewhere. You can find them here.

In addition to this open letter, we have published a set of policy recommendations which can be found here:

Policymaking in the Pause

View paper

12th April 2023

Notes and references

Down Arrow

Add your name to the list

Demonstrate your support for this open letter by adding your own signature to the list:

Full Name *

Email *

An email will be sent to this address to validate the signature.

Job Title / Position

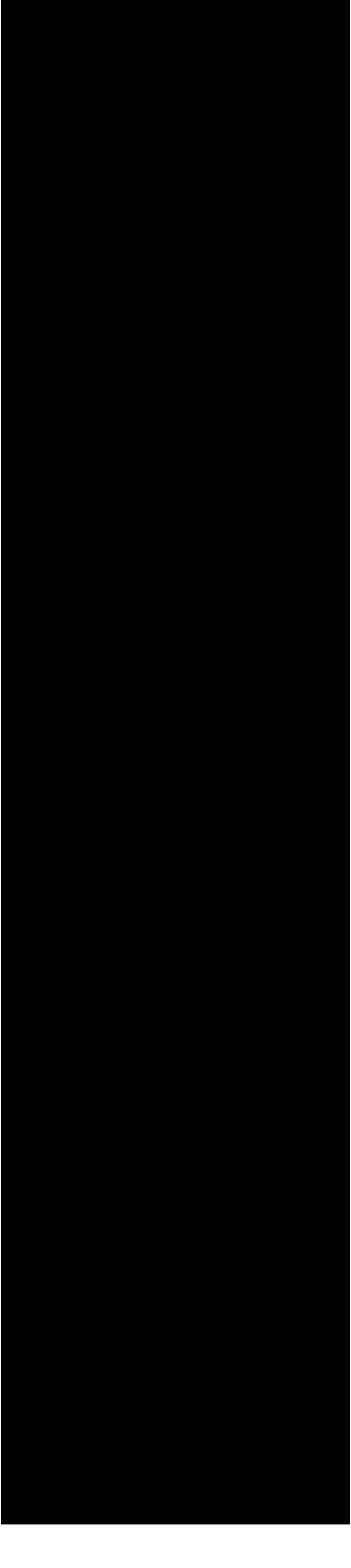
Affiliation

Next

Signature corrections

If you believe your signature has been added in error or have other concerns about its appearance, please contact us at letters@futureoflife.org.

Signatories



Read More

OPEN LETTERS

Related posts

If you enjoyed this, you also might like:

Open Letter Against Reckless Nuclear Escalation and Use

The abhorrent Ukraine war has the potential to escalate into an all-out NATO-Russia nuclear conflict that would be the greatest catastrophe in human history. More must be done to prevent such escalation.

Taylor Jones October 18, 2022

Forensic AI in AI Regulation Open Letter

The emergence of artificial intelligence (AI) promises dramatic changes in our economic and social structures as well as everyday life [...]

Anna Yelizarova June 14, 2020

Autonomous Weapons Open Letter: Global Health Community

Given our commitment to do no harm, the global health community has a long history of successful advocacy against inhumane weapons, and the World and American Medical Associations have called for bans on nuclear, chemical and biological weapons. Now, recent advances in artificial intelligence have brought us to the brink of a new arms race in lethal autonomous weapons.

admin March 13, 2019

2018 Statement to United Nations on Behalf of LAWS Open Letter Signatories

The following statement was read on the floor of the United Nations during the August, 2018 COW meeting, in which [...]

Ariel Conn September 4, 2018

Sign up for the Future of Life Institute newsletter

Join 20,000+ others receiving periodic updates on our work and cause areas.

View previous editions

Your name

Email address

Subscribe

☐ I consent to Future of Life Institute storing my information for the purpose of sending me email newsletters.*

☐ I'm not a robot

reCAPTCHA

Privacy - Terms