As Six-Month Pause Letter Expires, Experts Call for Regulation of Advanced AI Development

Sept. 21, CAMBRIDGE, MA – On Friday, the Future of Life Institute (FLI) will mark six months since they released their open letter calling for a six month pause on giant AI experiments, which kicked off the global conversation about AI risk. It was signed by more than 30,000 experts, researchers, industry figures and other leaders.

Since then, the EU strengthened its draft AI law, the U.S. Congress has held hearings on the large-scale risks, emergency White House meetings have been convened, and polls show widespread public concern about the technology's catastrophic potential - and Americans' preference for a slowdown. Yet much remains to be done to prevent the harms that could be caused by uncontrolled and unchecked AI development.

"AI corporations are recklessly rushing to build more and more powerful systems, with no robust solutions to make them safe. They acknowledge massive risks, safety concerns, and the potential need for a pause, yet they are unable or unwilling to say when or even how such a slowdown might occur," said **Anthony Aguirre**, FLI's Executive Director.

Critical Questions

FLI has created a list of questions that must be answered by AI companies in order to inform the public about the risks they represent, the limitations of existing safeguards, and their steps to guarantee safety. We urge policymakers, press, and members of the public to consider these - and address them to AI corporations wherever possible.

It also includes **quotes from AI corporations** about the risks, and **polling data** that reveals widespread concern.
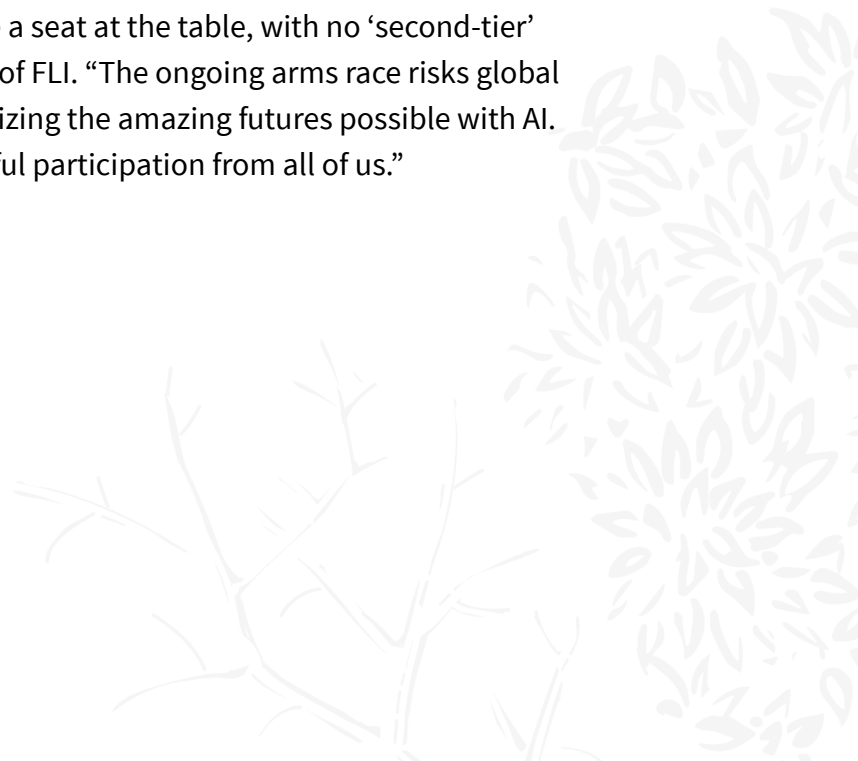
Policy Recommendations

FLI has published policy [recommendations](#) to steer AI toward benefiting humanity and away from extreme risks. They include: requiring registration for large accumulations of computational resources, establishing a rigorous process for auditing risks and biases of powerful AI systems, and requiring licenses for the deployment of these systems that would be contingent upon developers proving their systems are safe, secure, and ethical.

"Our letter wasn't just a warning; it proposed policies to help develop AI safely and responsibly. 80% of Americans don't trust AI corporations to self-regulate, and a bipartisan majority support the creation of a federal agency for oversight," said Aguirre. "We need our leaders to have the technical and legal capability to steer and halt development when it becomes dangerous. The steering wheel and brakes don't even exist right now".

Bletchley Park

Later this year, global leaders will convene in the United Kingdom to discuss the safety implications of advanced AI development. FLI has also released a [set of recommendations](#) for leaders leading up to and after the event.

"Addressing the safety risks of advanced AI should be a global effort. At the upcoming [UK summit](#), every concerned party should have a seat at the table, with no 'second-tier' participants" said **Max Tegmark**, President of FLI. "The ongoing arms race risks global disaster and undermines any chance of realizing the amazing futures possible with AI. Effective coordination will require meaningful participation from all of us."

Signatory Statements

Some of the letter's most prominent signatories, Apple co-founder **Steve Wozniak**, AI 'godfather' **Yoshua Bengio**, Skype co-founder **Jaan Tallinn**, political scientist **Danielle Allen**, historian **Yuval Noah Harari**, psychologist **Gary Marcus**, journalist **Rachel Bronson**, and leading expert **Stuart Russell** also made statements about the expiration of the six month pause letter.

"The last six months have seen a groundswell of alarm about the pace of unchecked, unregulated AI development," writes **Bengio**. "This is the correct reaction. Governments and lawmakers have shown great openness to dialogue and must continue to act swiftly to protect lives and safeguard our society from the many threats to our collective safety and democracies."

Russell reflects: "In 1951, Alan Turing warned us that success in AI would mean the end of human control over the future. AI as a field ignored this warning, and governments too. To express my frustration with this, I made up a fictitious email exchange, where a superior alien civilization sends an email to humanity warning of its impending arrival, and humanity sends back an out-of-office auto-reply. After the pause letter, humanity and its governments returned to the office and, finally, read the email from the aliens. Let's hope it's not too late."

"The out-of-control development and proliferation of increasingly powerful AI systems could inflict terrible harms, either deliberately or accidentally, and will be weaponized by the worst actors in our society," **Wozniak** explains. "Leaders must step in to help ensure they are developed safely and transparently, and that creators are accountable for the harms they cause. Crucially, we desperately need an AI policy framework that holds human beings responsible, and helps prevent horrible people from using this incredible technology to do evil things."

"It's been encouraging to see public sector leaders step up to the enormous challenge of

governing the AI-powered social and economic revolution we find ourselves in the midst of," says **Allen**. "We need to mitigate harms, block bad actors, steer toward public goods, and equip ourselves to see and maintain human mastery over emergent capabilities to come. We humans know how to do these things—and have done them in the past—so it's been a relief to see the acceleration of effort to carry out these tasks in these new contexts. We need to keep the pace up and cannot slacken now."

"Suppose we were told that a fleet of spaceships with highly intelligent aliens has been spotted, heading for Earth, and they will be here in a few years. Suppose we were told these aliens might solve climate change and cure cancer, but they might also enslave or even exterminate us. How would we react to such news?" **Harari** asks."Well, six months ago some of the world's leading AI experts warned us that an alien intelligence is indeed heading our way – only that this alien intelligence isn't coming from outer space, it is coming from our own laboratories. Make no mistake: AI is an alien intelligence. It can make decisions and create ideas in a radically different way than human intelligence. AI has enormous positive potential, but it also poses enormous threats. We must act now to ensure that AI is developed in a safe way, or within a few years we might lose control of our planet and our future to an alien intelligence."

"The *Bulletin of the Atomic Scientists*, the organization that I run, was founded by Manhattan Project scientists like J. Robert Oppenheimer who feared the consequences of their creation. AI is facing a similar moment today, and, like then, its creators are sounding an alarm," recollects **Bronson**. "In the last six months we have seen thousands of scientists - and society as a whole - wake up and demand intervention. It is heartening to see our governments starting to listen to the two thirds of American adults who want to see regulation of generative AI. Our representatives must act before it is too late."

"I supported this letter to make the growing fears of more and more AI experts known to the world. We wanted to see how people responded, and the results were mindblowing. The public are very, very concerned, as confirmed by multiple subsequent surveys," said **Tallinn**.

"People are justifiably alarmed that a handful of companies are rushing ahead to build and deploy these advanced systems, with little-to-no oversight, without even proving that they are safe. People, and increasingly the AI experts, want regulation even more than I realized. It's time they got it."

"In the six months since the pause letter, there has been a lot of talk, and lots of photo opportunities, but not enough action. No new laws have passed. No major tech company has committed to transparency into the data they use to train their models, nor to revealing enough about their architectures to others to mitigate risks," **Marcus** writes. "Nobody has found a way to keep large language models from making stuff up, nobody has found a way to guarantee that they will behave ethically. Bad actors are starting to exploit them. I remain just as concerned now as I was then, if not more so."

About the Future of Life Institute

Founded in 2014, the Future of Life Institute is a nonprofit working to steer transformative technologies toward benefiting humanity and avoiding existential risks. An organization composed of scientists, academics, and other leading experts, FLI is currently focused on guiding the impact of technology development around three major risk areas: artificial intelligence, nuclear weapons, and biotechnology.

Contact: Ben Cumming, Communications Director
Email: press@futureoflife.org
Phone: +44 7891 551 514