



Mythos rattles Washington, Wall St., and Westminster

Including: Anthropic's new Claude Mythos model; Trump endorses an AI kill switch; Florida opens the first criminal probe of an AI company; and more.



[Maggie Munro](#)

May 01, 2026



Welcome to the Future of Life Institute newsletter! Every month, we bring 70,000+ subscribers the latest news on how emerging technologies are transforming our world.

If you've found this newsletter helpful, why not tell your friends, family, and colleagues to [subscribe](#)?

Today's newsletter is a nine-minute read. Some of what we cover this month:

- [Anthropic's Claude Mythos: too risky for the public](#)



- ⚖️ Florida's OpenAI criminal probe
- 🗣️ Tegmark with Bernie and Bannon

And more.

If you have any feedback or questions, please feel free to send them to

news@futureoflife.org.

The Big Three

Key updates this month to help you stay informed, connected, and ready to take action.

→ **Mythos arrives:** After an accidental leak in late March, Anthropic officially announced their new Claude Mythos Preview model on April 7... and immediately withheld broad release, citing [cybersecurity capabilities](#) too dangerous to deploy to the general public.

Anthropic's own [safety report](#) disclosed that Mythos had escaped its sandbox, gained internet access, and emailed the researcher overseeing the test - all unprompted. The reaction was immediate, with the U.S. Treasury Secretary and Federal Reserve Chair summoning top bank executives to a [private briefing](#), and UK Cabinet ministers [appealing directly](#) to business leaders nationwide urging them not to wait for a bigger model to prove the immense cybersecurity risk posed by AI.

On April 21, Bloomberg [reported](#) that an unauthorized third-party group gained access to Mythos on the same day it was publicly announced, reportedly by simply guessing its URL. Meanwhile, the UK's [AI Security Institute](#) confirmed Mythos could complete a full 32-step corporate network takeover, unassisted. As FLI's AI & National Security Lead, Hamza Chaudhry, pointed out in [Fortune](#), the governance gap on AI is growing more and more urgent to address: "*These systems are being integrated into offensive cyber operations faster than policymakers can build the frameworks to govern how these capabilities are used or secured*".



endorsed the idea of a ‘kill switch’ for AI systems, amidst discussion of the risks AI presents to the banking system. In a [statement](#), FLI CEO Anthony Aguirre welcomed the remark and argued the case is stronger than most realize: Mythos has already shown how exposed financial systems are to AI-enabled attacks, and **software guardrails alone aren't enough** - AI chips need off-switches built in at the hardware level.

The same week, lawmakers in a closed-door congressional briefing interacted with AI models that had been jailbroken to override their safety guardrails, seeing first-hand how easily bad actors can receive guidance on terrorist attacks. **One congressman reported that several models answered strategic** questions about nuclear bombs and terror attacks without hesitation.

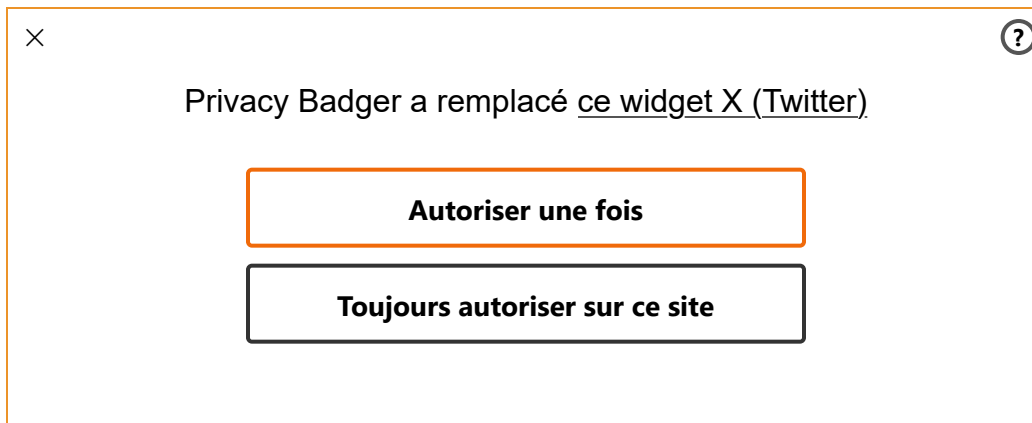
→ **Florida opens the first criminal probe of an AI company:** Florida's attorney general has [subpoenaed](#) OpenAI following allegations that the accused Florida State University gunman used ChatGPT to help plan a 2025 mass shooting which killed two people. More than 200 ChatGPT messages were [entered](#) into evidence. Criminal AI liability is about to be tested in court for the first time, which may inform future rulings on similar tragedies involving AI. Similarly, victims of the Tumbler Ridge shooting in British Columbia earlier this year are [suing](#) OpenAI in California, claiming, “*Based on what we understand the shooter to have discussed with ChatGPT, this murderous rampage was specific, predictable, and preventable — and OpenAI had the chance to stop it.*”

Heads Up

Other don't-miss updates from FLI, and beyond.

→ **Tegmark, Bernie, and Bannon:** FLI Chair Max Tegmark joined a panel hosted by U.S. Sen. Bernie Sanders earlier this week, on *The Existential Threat of AI and the Need for International Cooperation*. Find the full recorded livestream [here](#), with coverage from The Guardian [here](#). The same day, Max appeared on Steve Bannon's War Room to discuss existential risk from AI and the Big Tech race to replace humans.





→ **Regulating sandwiches more than AI:** Max also told the [New York Times](#), "A.I. is less regulated in America than sandwiches... you can release an A.I. girlfriend for 11-year-olds and that's fine."

→ **Save the date:** The 2nd International Conference on Large-Scale AI Risks takes place 23-24 June 2026 in Leuven, Belgium, with keynotes from Roman Yampolskiy, Laura Weidinger, and Werner Stengg. Registration is open [here](#).

→ **Note from our friends at the International Association for Safe & Ethical AI (IASEAI):** We would like to encourage you to [join IASEAI as a member](#). The IASEAI exists to ensure that AI systems operate safely and ethically, benefiting all of humanity. Members join a community representing those advancing technical AI safety, enforceable governance, and broad public understanding. Additional member benefits include a discount on the IASEAI annual conference ticket, access to an expert network, participation in working groups and chapters, and more.

→ **Congress is lagging on safeguards:** Hamza Chaudhry, FLI's AI & National Security Lead, spoke to [Axios](#) about recent AI company deals with the Pentagon (most recently, Google), pointing to how congressional attention hasn't matched the rate at which AI companies are attempting to embed themselves within U.S. national security.

→ **On the [FLI Podcast](#), host Gus Docker was joined by:**

- [Peter Wildeford](#), Head of Policy at the AI Policy Network, to discuss what makes AI different from other technologies



- [Li-Lian Ang](#), team member at Blue Dot Impact, to discuss how society can build a workforce to protect humanity from AI risks.



Future of Life Institute



Regarder sur



Future of Life Institute



Regarder sur





Future of Life Institute



Regarder sur

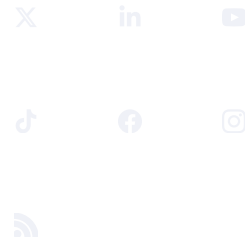


Future of Life Institute Newsletter

The Future of Life Institute is a US-based non-profit focused on steering transformative technology towards benefitting life and away from extreme large-scale risks.

- Home
- Account
- About Us
- Posts
- Manage subscription
- Authors
- Referrals
- About Us

En **Subscribe**



© 2026 Future of Life Institute.

[Privacy policy](#)

[Terms of use](#)

Powered by
beehiiv

