Testimony before the U.S. House Committee on the Judiciary, Subcommittee on Courts, Intellectual Property, Artificial Intelligence, and the Internet

For the hearing titled

"Protecting Our Edge: Trade Secrets and the Global AI Arms Race"

May 7, 2025

Helen Toner

Director of Strategy and Foundational Research Grants

Center for Security and Emerging Technology, Georgetown University

---

Chair Issa, Ranking Member Johnson, members of the Subcommittee: Thank you for the opportunity to testify before you today.

I have spent the last 6 years working on AI and national security policy at Georgetown University's Center for Security and Emerging Technology (CSET). U.S.-China competition in AI is a major focus of my research, as are questions of AI safety, security, and governance. I served on OpenAI's board of directors from 2021 to 2023.

Context and Focus

My testimony will focus on trade secrets and competition issues relevant to so-called "**frontier AI**," which refers to cutting-edge, general-purpose AI systems such as Google's Gemini 2.5, OpenAI's o3, Anthropic's Claude 3.7, xAI's Grok 3, and Meta's Llama 4.

Frontier AI is only one part of the larger AI ecosystem, but from a strategic perspective, it is an especially important part. The companies at the frontier are actively working to build artificial general intelligence (AGI), i.e. AI that is as capable as human experts across a wide range of fields. The CEOs of these companies claim that this goal will likely be reached within the next 2-5 years,[1] a view shared by top researchers and engineers both inside and outside of these companies. Once they reach AGI, they plan to push ahead with building "superintelligence," i.e.

[1] Anthropic CEO Dario Amodei: "Making AI that is smarter than almost all humans at almost all things [...] is most likely to happen in 2026-2027." OpenAI CEO Sam Altman: "I think AGI will probably get developed during this president's term." Google DeepMind CEO Demis Hassabis: "I think we're probably three to five years away [from AGI]."

AI that is far smarter and more capable than humans.[2] Even if their projected timelines are overly optimistic, it is not an exaggeration to say that this level of AI would reshape the economy, upend the political system, and transform the international order.

The idea that AI is hard to govern because the technology is moving so fast is cliched, but true. But how AI is changing is predictable in at least one important way: it is getting more capable, more powerful, and more strategically critical.

Within the scope of this hearing, this has two important implications:

- **The U.S. government has a critical national security interest in protecting the trade secrets and IP of leading U.S. AI companies.**
- **The U.S. government has a critical national security interest in understanding and monitoring the technology being built inside leading U.S. AI companies.**

These two interests do not need to be in tension with each other, and can be pursued in parallel. I elaborate on how below.

The Importance of Securing Frontier AI Companies

As leading AI companies' technology becomes increasingly strategically sensitive, the national security implications of inadequate cybersecurity grow more acute than for a typical company. If we believe that it's vitally important for the world's most advanced AI systems to be developed and deployed by U.S. companies, then we must also believe that it's vitally important to prevent that AI technology from being stolen by our adversaries.

Unfortunately, current security practices at leading AI companies are insufficient to prevent compromise by advanced persistent threat (APT) actors such as state-based cyber operations. Even the best-defended companies in America are widely acknowledged to be unable to resist penetration by APTs, and AI companies are no exception. In most sectors, this vulnerability entails risks to U.S. citizens' personal data and companies' commercially sensitive IP. As AI advances, the stakes for frontier AI companies grow even higher than this—more akin to threats to critical infrastructure providers or companies in the defense industrial base. Compromise of a frontier AI system, theft of key trade secrets involved in creating it, or even the wholesale exfiltration of a frontier model itself would seriously harm U.S. prospects in the global AI competition, with implications for future military competitiveness, economic competitiveness, and national power more broadly.

[2] E.g. per Sam Altman: "We are beginning to turn our aim beyond AGI, to superintelligence in the true sense of the word."

What's more, we know that foreign actors (including but not limited to China and Russia) are extremely capable of stealing U.S. intellectual property and extremely interested in boosting their domestic AI ecosystems.[3] This means that the leading U.S. AI companies are sitting ducks for major state-sponsored operations to steal critical trade secrets that could eat away at U.S. AI leadership.

One way of breaking down the trade secrets of concern in frontier AI development is to consider AI models themselves, plus the triad of inputs used to make them: algorithms, data, and hardware.

- **AI models** are frontier AI companies' crown jewels. To steal a model, the attacker needs to download a file containing the set of numbers known as "model weights" or "parameters." For today's leading models, this file is likely on the order of several terabytes in size, meaning exfiltration is non-trivial but possible. If an adversary is able to steal a trained model, they have access to the capabilities of that model without needing the world-class team, months of research time, and tens or hundreds of millions of dollars used to create ("train") the model. They can also flexibly modify or further build on the model. For these reasons, the weights of frontier models are broadly considered the highest priority type of intellectual property for frontier AI companies to protect.

- **Algorithmic secrets** is an overarching term used to describe privately held insights and techniques used to design and train models. They include model architectures (the design of the model), training "recipes" (the sequence of steps used to optimize the model), nuggets of practical know-how, and new research ideas. Algorithmic secrets often take the form of a few sentences—perhaps a single sentence—that could be extracted from internal documents or messaging platforms, or even let slip by an incautious employee.

- **Training datasets** are the data used to create AI models. For frontier AI models, these include multi-trillion-word corpora of text and other data scraped from the internet and other sources; carefully curated human-written examples of priority use cases; collections of human-submitted or AI-generated rankings and ratings comparing different AI-generated options; AI-generated text; and other types of data.

- **AI hardware** provides the computing power (often referred to as "compute") needed to train frontier models. Access to state-of-the-art AI chips, which are subject to U.S. export controls, is a major competitive advantage for U.S. frontier AI firms. The most

[3] See, e.g., Hannas, Mulvenon, and Puglisi 2013, *Chinese Industrial Espionage.*

sensitive trade secrets associated with AI hardware are chip designs, though these are of limited utility unless the thief can access leading-edge chip production facilities. In most cases, chip designs are not held by AI companies themselves, since almost all frontier AI companies use graphics processing units (GPUs) designed by NVIDIA. The exception is Google, which uses its in-house tensor processing units (TPUs). The GPUs and TPUs used for frontier AI training can currently only be produced by the Taiwanese Semiconductor Manufacturing Corporation (TSMC).

Trade secrets in these categories have already been stolen from top firms. In 2023, an attacker gained access to internal communication channels at OpenAI and was able to extract proprietary information on how they develop their AI technology, i.e. algorithmic secrets.[4] In 2024, the FBI arrested a Google employee for systematically exfiltrating large volumes of confidential information from Google, including information relating to Google's TPU AI chips.[5]

A 2024 RAND analysis found that there were many areas where frontier AI companies could immediately improve their security, and that there was also a need for significant investments over the longer term to increase their capacity to resist more sophisticated attacks.[6]

The Importance of Government Visibility Into Frontier AI Development

U.S. frontier AI companies claim that they are on track to build extraordinarily powerful—and potentially extraordinarily dangerous—technology in the next 2-5 years. Materials produced by frontier AI companies describe a range of severe risks that their own AI systems might soon pose, including aiding in sophisticated offensive cyber operations, enabling amateurs to carry out bioterror attacks, and potentially evading human control entirely.[7] Even in scenarios where all goes to plan and these catastrophic risks do not materialize, increasingly advanced AI will have extremely disruptive effects on society more broadly.

Historically, technologies with such major strategic implications have been developed under the auspices of the U.S. government and closely associated firms. This time, it is being developed entirely within private industry, with little visibility available to Congress or the executive branch. This threatens the U.S. government's ability to appropriately manage and

[4] New York Times 2024, "A Hacker Stole OpenAI Secrets, Raising Fears That China Could, Too."
[5] Department of Justice 2024, "Chinese National Residing in California Arrested for Theft of Artificial Intelligence-Related Trade Secrets from Google."
[6] Nevo et al. 2024, "Securing AI Model Weights."
[7] See Anthropic's Responsible Scaling Policy, OpenAI's Preparedness Framework, and Google's Frontier Safety Framework.

respond to continued developments in frontier AI, up to and including the possible development of AGI and superintelligence.

Efforts to develop a robust regulatory framework for frontier AI have largely stalled out at the federal level in the United States. In the absence of such a framework, there is broad agreement among AI policy experts with widely varying political views that transparency and disclosure requirements are a minimal, light-touch approach that should be pursued.[8] Increasing the information flow between frontier companies and the outside world has a slew of benefits—it reduces information asymmetries, empowers government to understand and respond to advances, and equips the public to weigh in on a technology that will profoundly impact them.

Transparency requirements for AI development can take different forms, depending on the information of interest and the tradeoffs involved in sharing it. Three major types of transparency are disclosure to the *public*, disclosure to *government*, and disclosure to a *third-party auditor.* Each of these balances different pros and cons. Public disclosure goes furthest in reducing information gaps, but may be undesirable for information that is sensitive from a commercial or national security perspective. Disclosure directly to USG is well suited for information directly related to national security, such as results of tests on AI models' capabilities in areas including cyber operations, bioweapon development, and nuclear weapons. Disclosure to third-party auditors is a flexible option that can allow a neutral, independent organization to verify or assess sensitive information while keeping it largely under the AI company's control, e.g. by having the auditor work within the AI company's own facilities under a non-disclosure agreement.

Forcing U.S. companies to disclose information that would damage their competitiveness would not be a good approach. Fortunately, most of the information that is of greatest interest from a national security and public interest perspective would not damage competitiveness if disclosed appropriately. Types of information that would be valuable to share include:

- **Results of testing for AI models' capabilities and risks.** A lack of clear, up-to-date information about what the world's best AI systems are capable of puts the U.S. government at a huge disadvantage in understanding how the AI frontier is progressing and what kinds of responses might be needed. At present, information about how rapidly AI is advancing is shared (or not shared) at the discretion of companies, on the timeline and in the format that they find most convenient.

[8] See, e.g., Ball and Kokotajlo 2024, "4 Ways to Advance Transparency in Frontier AI Development."

- **Information about the goals or specifications AI models are being trained to pursue.** This creates visibility into how companies are making crucial, politically loaded decisions about what their models should and shouldn't do, as well as how to prioritize different values (e.g. freedom of speech vs. avoiding hate speech, supporting users vs. avoiding sycophancy, etc.). As frontier AI models are integrated into citizens' lives, businesses' software, and government's infrastructure, knowing what they are designed to optimize for will become increasingly important. OpenAI and Anthropic voluntarily share a version of this information, to their credit.

- **Analysis of why AI developers believe their current risk management practices are sufficient** (sometimes referred to as a safety case). Developing and releasing frontier AI models involves making a large number of judgment calls about how to test for risks, what results are acceptable, and how to manage uncertainty. Making at least a high-level version of that thinking transparent to a wider audience (perhaps with especially sensitive details redacted) would bring these critical decisions out from closed meeting rooms into the sunlight.

- **Data on internal usage of AI.** Frontier AI companies are increasingly relying on their own, sometimes unreleased, AI models to automate their own work. It is a widely held view among AI researchers and CEOs that using AI to accelerate AI research could lead to a runaway feedback loop of increasingly advanced AI. Such a feedback loop is known as an "intelligence explosion," and is at the center of some of AI experts' greatest fears about AI catastrophe. Greater transparency about the extent to which AI is in fact accelerating research would prevent this phenomenon from transpiring in secret.

- **Whistleblower reports.** If AI companies want to claim that they are building the most world-changing technology to ever exist, their employees should be able to share concerns about risks to public safety along the way without fearing the repercussions. AI whistleblowing is largely unprotected by existing whistleblower laws, which focus on outright illegal activity.

Note how little these categories overlap with the trade secrets outlined in the previous section. The categories immediately above are areas where the public or governmental interest in transparency is much higher than the downside of making the information available to competitors. This cost-benefit is different for other types of information. In particular, as outlined in the previous section, information about how AI models are built (e.g. architectural details, training recipes, or full datasets) is more commercially sensitive, as are the models themselves.

In weighing the pros and cons of requiring any particular kind of information be disclosed, however, it is important to recall that even commercially sensitive information is already stored inside AI companies themselves, where it is already vulnerable to espionage. This baseline should be considered alongside the sensitivity of the information, the benefit of sharing it, and the security of where it would be stored (e.g. in a classified USG environment vs. on a public website).

Recommendations

There are a range of actions available to Congress, in its legislative and oversight capacities, to pursue the twin priorities of keeping U.S. frontier AI secure while ensuring visibility into its development:

1. **Expand and resource security-focused collaborative arrangements between the U.S. government and frontier AI companies.** Several small-scale, voluntary programs currently exist, such as threat intelligence sharing with the intelligence community and voluntary testing for national security-relevant capabilities at a dedicated institute within NIST. These initiatives are highly valuable and should be strengthened.

2. **Use existing authorities, or create new ones, to promote transparency from frontier AI companies about their most advanced systems.** Without new legislation, Congress can ask AI companies about their compliance with their own voluntary commitments to share safety and security information. The U.S. government can also use existing authorities, such as the Defense Production Act, to require companies to share information via secure channels. In addition to these options, new legislation could enshrine transparency requirements and/or create whistleblower protections for AI company employees, for example by shielding them from retaliation, identifying clear channels to report concerns, and creating monetary incentives.

3. **Invest in the technologies and infrastructure needed to ensure the security of strategically critical AI systems of the future.** Confidential computing techniques could allow model weights to be kept secure at the chip level, even while in use, but would need to be adapted and adopted at scale. Federal funding could help accelerate the availability of this technology. Likewise, it may be necessary to rebuild advanced AI data centers from the ground up to maximize security; if so, public funding or public-private partnerships could be helpful.

4. **Consider creating legal space to allow companies to prioritize security in ways that are currently legally fraught.** One example would be creating safe harbor protections for companies to collaborate on AI safety and security issues without triggering

antitrust concerns. This would of course need to be structured in a way that would not permit collusion. Another possibility would be authorizing AI companies to engage in intensive personnel vetting and monitoring, practices that may be discouraged under state employment law but that would be essential in any serious effort against insider threats. Crucially, any personnel-focused efforts must be implemented thoughtfully in order to protect the enormous contributions of foreign nationals to U.S. AI competitiveness.[9] What is needed is a scalpel—the ability to carefully identify individuals who should not have access to highly sensitive trade secrets, and keep tabs on those who do have access—not a sledgehammer of anti-immigrant measures.

5. **Push frontier AI companies to improve their cybersecurity practices**, and emphasize to them that this must be a high priority. Make clear that the security of critical frontier AI IP is as core to U.S. competitiveness as rapid innovation, since the United States cannot lead if adversaries have easy access to our best technology.

Thank you, and I look forward to your questions.

[9] See Zwetsloot 2021, "Winning the Tech Talent Competition," and Oschinski et al. 2025, "Strengthening America's AI Workforce."