Nicole Schneidman (SBN 319511)
AI for Democracy Action Lab at
PROTECT DEMOCRACY PROJECT
P.O. Box 341423
Los Angeles, CA 90034-9998
(202) 579-4582
nicole.schneidman@protectdemocracy.org

Deana K. El-Mallawany (SBN 674825)*
AI for Democracy Action Lab at
PROTECT DEMOCRACY PROJECT
15 Main Street, Suite 312
Watertown, MA 02472
(202) 579-4582
deana.elmallaway@protectdemocracy.org

Ori Lev (SBN 452565)*
AI for Democracy Action Lab at
PROTECT DEMOCRACY PROJECT
2020 Pennsylvania Ave. NW, Suite #163
Washington, D.C. 20006
(202) 579-4582
ori.lev@protectdemocracy.org

Counsel for Amici Curiae

**Application for admission pro hac vice forthcoming*

IN THE UNITED STATES DISTRICT COURT

FOR THE NORTHERN DISTRICT OF CALIFORNIA

ANTHROPIC PBC,

Plaintiff,

v.

U.S. DEPARTMENT OF WAR, *et al.*,

Defendants.

Case No.: 3:26-cv-01996-RFL

[PROPOSED] BRIEF OF *AMICI CURIAE* EMPLOYEES OF OPENAI AND GOOGLE IN THEIR PERSONAL CAPACITIES IN SUPPORT OF PLAINTIFF'S MOTION FOR A TEMPORARY RESTRAINING ORDER

Judge: Hon. Rita F. Lin

Date:
Time:
Ctrm:

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Cases

Statutes

Other Authorities

IDENTITY AND INTEREST OF THE *AMICI*

Amici are engineers, researchers, scientists, and other professionals employed at U.S. frontier artificial intelligence laboratories. We build, train, and study the large-scale AI systems that serve a wide range of users and deployments, including in the consequential domains of national security, law enforcement, and military operations. We submit this brief not as spokespeople for any single company, but in our individual capacities as professionals with direct knowledge of what these systems can and cannot do, and what is at stake when their deployment outpaces the legal and ethical frameworks designed to govern them.

As a group, we are diverse in our politics and philosophies, but we are united in the conviction that today's frontier AI systems present risks when deployed to enable domestic mass surveillance or the operation of autonomous lethal weapons systems without human oversight, and that those risks require some kind of guardrails, whether via technical safeguards or usage restrictions. We view this conviction not as a result of any particular set of ideological or political commitments, but rather as a conclusion that follows from any reasonable evaluation of the capabilities and limitations of currently available frontier AI systems. It is this conviction that brings us before the Court to respectfully submit this brief, in the hopes that our understanding of the technology at issue, and our unique perspective as employees of companies currently engaged in fierce competition with Anthropic, will shed some light on the stakes of this case.

Individual *amici* are identified below. They sign the brief in their personal capacities only, and their companies and titles are included only to provide a sense of the perspectives they bring to this case.

- Grant Birkinbine, Member of Technical Staff – Security Engineer, OpenAI
- Anna-Luisa Brakman, Member of Technical Staff, OpenAI
- Sarah Cogan, Senior Software Engineer, Google DeepMind
- Jeff Dean, Chief Scientist, Google
- Michael Dennis, Senior Research Scientist, Google DeepMind
- Sanjeev Dhanda, Senior Staff Software Engineer, Google DeepMind

- Rasmi Elasmar, Senior Research Engineer, Google DeepMind
- Brian Fioca, Member of Go to Market Staff, OpenAI
- Aaron Friel, Member of Technical Staff – Software Engineer, OpenAI
- Leo Gao, Member of Technical Staff, OpenAI
- Edward Grefenstette, Director of Research, Google DeepMind
- Alexander Irpan, Research Scientist, Google DeepMind
- Rishub Jain, Research Engineer, Google DeepMind
- Manas Joglekar, Member of Technical Staff, OpenAI
- Kathy Korevec, Director, Product at Google
- Shrinu Kushagra, Research Scientist, Google
- Teddy Lee, Member of GTM Staff, OpenAI
- Sharon Lin, Research Engineer, Google DeepMind
- Soukaina Mansour, Creative Community Lead, OpenAI
- Ian McKenzie, Research Engineer, Google DeepMind
- Pamela Mishkin, Research, OpenAI
- Roman Novak, Research Scientist, OpenAI
- Zach Parent, Forward Deployed Engineer, OpenAI
- Andrew Schmidt, Model Designer, OpenAI
- Noah Siegel, Senior Research Engineer, Google DeepMind
- Jordan Sitkin, Member of Technical Staff, OpenAI
- Chang Sun, Member of Data Science Staff, OpenAI
- Sean Talts, Staff Software Engineer, Google
- Alexander Matt Turner, Research Scientist, Google DeepMind
- Anna Wang, Research Scientist, Google DeepMind
- Zhengdong Wang, Senior Research Engineer, Google DeepMind
- Jonathan Ward, Member of Technical Staff, OpenAI
- Jason Wolfe, Member of Technical Staff, OpenAI

- Kate Woolverton, Senior Software Engineer, Google DeepMind
- Gabriel Wu, Member of Technical Staff – Research Engineer, OpenAI
- Cathy Yeh, Member of Technical Staff, OpenAI
- Jelle Zijlstra, Member of Technical Staff, OpenAI

INTRODUCTION AND SUMMARY OF ARGUMENT

This case arises from the Pentagon delivering on its threat to designate Anthropic a "supply chain risk" if the company declined to agree to remove limitations on the use of its AI systems for domestic mass surveillance or fully autonomous lethal weapons systems. If it were no longer satisfied with the agreed-upon terms of its contract with Anthropic, the Defendants could have simply canceled the contract and purchased the services of another leading AI company. Instead, Defendants recklessly invoked national security authorities intended to protect the procurement process from interference by foreign adversaries. If allowed to proceed, this effort to punish one of the leading U.S. AI companies will undoubtedly have consequences for the United States' industrial and scientific competitiveness in the field of artificial intelligence and beyond. And it will chill open deliberation in our field about the risks and benefits of today's AI systems. Because we understand the risks of frontier AI systems and the need for guardrails, and because we believe that speaking openly about them is of paramount importance, we submit this brief.

We offer three arguments.

First, the government's designation of Anthropic as a supply chain risk was an improper and arbitrary use of power that has serious ramifications for our industry. While we are not privy to the details of how Anthropic and the Pentagon's contractual relationship broke down, we are concerned that the Defendants' action harms public debate on the risks and benefits of AI as well as U.S. competitiveness in the field of AI and innovation more broadly.

Second, the technical concerns animating Anthropic's "red lines" are legitimate and widely recognized within our scientific community as requiring some kind of response. The best

currently available AI systems cannot safely or reliably handle fully autonomous lethal targeting, and should not be available for domestic mass surveillance of the American people. While there are various ways to establish these guardrails, we agree that these guardrails must be in place.

Third, as AI professionals, we understand that the substantive risks of the two use cases at issue are profound. AI-enabled mass domestic surveillance would transform the fragmented data ecosystem that already surrounds American life into a unified, real-time instrument for monitoring the entire population. Even the awareness that such capability exists creates a chilling effect on democratic participation. Autonomous lethal weapons systems, as currently designed and deployed, cannot reliably distinguish combatants from civilians, cannot explain their targeting decisions, and cannot engage in human accountability structures. These concerns require a response.

For these reasons, we urge the Court to grant the relief requested by Anthropic.

ARGUMENT

I. The "Supply Chain Risk" Designation Is Improper Retaliation That Harms the Public Interest.

This case poses a question of seismic importance for our industry, our national security, and our democracy: What happens when the government uses its national security authorities to punish a private company for maintaining safeguards on certain uses of its AI systems while speaking to why those safeguards exist and why they matter?

In early March 2026, the Pentagon officially designated Anthropic as a supply chain risk, following earlier threats to do so. While we are not privy to the details of their negotiations, the Defendants had the option simply to drop Anthropic's contract if it no longer wished to be bound by its terms. The supply chain risk designation is a mechanism for excluding from the defense industrial base vendors who pose a genuine threat to the integrity of military systems.[1] It is scarcely used, and then for foreign adversary-controlled companies, compromised suppliers, and

[1] *See* 10 U.S.C. § 3252 (supply chain security authority); *see also* Secure and Trusted Networks Security Act of 2019, 47 U.S.C. §1601 *et seq.* (requiring the Federal Communications Commission to publish list of prohibited equipment).

contractors whose products create exploitable vulnerabilities.[2] Anthropic is a domestic AI developer[3] that has worked with the Pentagon on military applications of AI systems since last year.

The Pentagon's decision to reach for supply chain risk authority in response to Anthropic's contract negotiations introduces an unpredictability in our industry that undermines American innovation and competitiveness. It chills professional debate on the benefits and risks of frontier AI systems and various ways that risks can be addressed to optimize the technology's deployment. The United States' thriving AI ecosystem leads the rest of the world largely due to the competition and flow of ideas between different AI companies. By silencing one lab, the government reduces the industry's potential to innovate solutions. The resulting harm has constitutional dimensions as well, undermining the freedom to engage in public debate about how powerful technologies should be governed. *See Hartman v. Moore*, 547 U.S. 250, 256 (2006) ("[T]he First Amendment prohibits government officials from subjecting an individual to retaliatory actions . . . for speaking out.").

II. The Concerns Underlying Anthropic's "Red Lines" Are Real and Require a Response.

As AI professionals, we recognize that frontier AI is a powerful technology that could have many benefits for humanity but also carries many risks. The risks are not hypothetical. They are structural. They follow from the nature of the technology itself, at least as it exists today, and from what happens when institutions, however well-intentioned, acquire capabilities that exceed the oversight mechanisms designed to check them.

That is why it is important to put guardrails around the domains in which these systems carry intolerable risk as they are currently constructed. A child's tricycle can physically be driven on an interstate, but we do not allow it because of the risks of using the technology in that

[2] *See, e.g.*, Exec. Order No. 13873, 84 Fed. Reg. 22689 (May 17, 2019) (authorizing import bans on technology threats under emergency trade authorities, aimed at Chinese vendors Huawei and ZTE).

[3] It is at least notable that Congress has already largely prohibited acquisition of AI systems from China, Russia, North Korea, and Iran. National Defense Authorization Act of 2026, Pub. L. 119-60, tit. XV, § 1532. There is no equivalent provision for U.S.-based systems providers. *Id*.

environment. Mass domestic surveillance and autonomous lethal weapons systems are the equivalently reckless domain for today's frontier models. The considered judgment, shared widely across the AI development community, is that these applications of current AI technology carry risks so severe, and threaten harm so impossible to repair after the fact, that some kind of guardrails — whether contractual or technical — are necessary to constrain them in the absence of robust, genuinely effective governance frameworks. For a system vendor to insist that those boundaries be honored as a condition of access to its software is not arbitrary, anticompetitive, or contrary to the public interest.

The legal vacuum in which these contractual terms exist makes them only more important. The United States currently has no comprehensive federal law governing the use of AI by military or intelligence agencies in domestic contexts. There is no statutory framework requiring transparency, judicial oversight, or meaningful accountability for AI-driven surveillance at scale. There is no enforceable legal standard governing when an autonomous weapons system may select and engage a target. In the absence of public law, the contractual and technological requirements that AI developers impose on the use of their systems represent a vital safeguard against their catastrophic misuse.

III. Mass Domestic Surveillance Powered by AI Poses Profound Risks to Democratic Governance — Even in Responsible Hands.

The risks of AI-enabled mass domestic surveillance merit greater public understanding. At its core, AI-enabled mass surveillance means the ability to monitor, analyze, and act on the behavior of an entire population continuously and in real time. The devices and data streams required to do this already exist. As of 2018,[4] there were approximately 70 million surveillance cameras operating in the United States across airports, subway stations, parking lots, storefronts, and street corners. Every smartphone continuously broadcasts location data to carriers and dozens of applications. Credit and debit cards generate a timestamped record of nearly every

[4] U.S. Bureau of Labor Statistics, *Beyond the Numbers: Artificial Intelligence — Taking on a Bigger Role in Our Future Security* (May 3, 2021), https://www.bls.gov/opub/btn/volume-10/investigation-and-security-services.htm.

commercial transaction Americans make. Social media platforms log not just what people post, but what they read, how long they browse, and what they posted before deleting it. Employers, insurers, and data brokers have assembled behavioral profiles on most American adults that are already, in many cases, available for government purchase without a warrant. What does not yet exist is the AI layer that transforms this sprawling, fragmented data landscape into a unified, real-time surveillance apparatus. Today, these streams are siloed, inconsistent, and require significant human effort to connect. From our vantage point at frontier AI labs, we understand that an AI system used for mass surveillance could dissolve those silos, correlating face recognition data with location history, transaction records, social graphs, and behavioral patterns across hundreds of millions of people simultaneously.

The mere existence of such a capability in government hands — even if never activated against a specific individual — changes the character of public life in a democracy. Behavioral scientists and legal scholars have long documented what is sometimes called the "panopticon effect": when people believe they may be observed, they modify their behavior as if they are always being observed, regardless of whether anyone is actually watching. The journalist thinks twice before calling a source inside the military, knowing the call could be logged and cross-referenced. The activist softens her public messaging, calculating that visibility now carries risk it didn't carry before. The academic researcher avoids certain search terms — not because the research is wrong, but because she doesn't want to surface in a database. None of these people have been targeted. None have been punished. But their behavior has already been constrained, and with it the democratic functions they serve — a free press, political organizing, open intellectual inquiry — have been quietly degraded. These chilling effects require no abuse, only the awareness that the capability exists.

History offers ample warning. The FBI's COINTELPRO program, which ran from 1956 to 1971 and was exposed years later, demonstrated how domestic intelligence powers justified by security concerns were systematically turned against civil rights leaders, journalists, and political dissidents. The program did not merely surveil its targets. It fabricated evidence, sent anonymous

letters designed to destroy marriages and careers, tipped off employers, and worked to discredit Martin Luther King, Jr. after he was awarded the Nobel Peace Prize.[5] It operated for fifteen years before Congress learned of its existence. AI does not merely replicate those dangers — it multiplies them by orders of magnitude, automating at national scale what previously required hundreds of human operatives.

Further enhancing the risk terrain for AI's deployment in this context, the Pentagon operates under a legal framework oriented toward external threats and warfighting, not domestic civil life. The Posse Comitatus Act, passed in 1878 in direct response to the use of federal troops to police American civilians during Reconstruction, reflects a constitutional tradition of keeping military power categorically separate from domestic governance.[6] When the Pentagon acts domestically, it is operating in legal territory it was not designed for, with oversight structures that were not built to catch domestic abuses. That is in part why the bulk data collection programs by the Pentagon's own National Security Agency (NSA), revealed by Edward Snowden in 2013, were so shocking and produced measurable chilling effects on lawful speech and inquiry. A study published in the Berkeley Technology Law Journal found statistically significant drops in traffic to Wikipedia articles on terrorism-related topics following the Snowden revelations, likely as ordinary people adjusted their online behavior in response to awareness that their searches were potentially being monitored.[7]

The harms from building this infrastructure are not easily undone, as we understand in our field. Data collected on a population does not expire. A database of location records, behavioral profiles, and social graphs built today will still exist years from now, accessible to

---

[5] *See* Senate Select Committee to Study Governmental Operations with Respect to Intelligence Activities, Final Report, S. Rep. No. 94-755, Book III, at 3, 8, 26, 138, 143-144 (1976).

[6] Posse Comitatus Act, 18 U.S.C. § 1385 (originally enacted 1878) (prohibiting use of the military to execute civilian laws except as expressly authorized by the Constitution or Act of Congress).

[7] Jonathon W. Penney, *Chilling Effects: Online Surveillance and Wikipedia Use*, 31 Berkeley Tech. L.J. 117 (2016). For more on NSA program's chilling effects, see generally Privacy and Civil Liberties Oversight Board, *Report on the Telephone Records Program Conducted Under Section 215 of the USA PATRIOT Act* (Jan. 23, 2014), https://documents.pclob.gov/prod/Documents/OversightReport/ec542143-1079-424a-84b3-acc354698560/215-Report_on_the_Telephone_Records_Program.pdf.

whoever controls it under whatever political conditions prevail then. That data would feed into an AI-powered surveillance infrastructure that, once constructed, tends to expand rather than contract. Agencies find new uses for existing capabilities, authorities get quietly reinterpreted, and the political cost of dismantling something already built is almost always higher than the cost of letting it continue and grow. One lesson of the Snowden revelations is that technology built for international espionage has a way of being used for domestic surveillance without clear legal boundaries. The boundary the Posse Comitatus Act was designed to protect, once eroded by the establishment of a Pentagon-controlled domestic surveillance apparatus, may prove practically impossible to restore.

We do not suggest that the Defendants intend to misuse such capabilities. We suggest that the question of intent is the wrong question. Democratic governance does not rest on the good intentions of those in power. It rests on structural constraints that make abuse difficult regardless of intent. AI-enabled mass domestic surveillance, deployed without transparent legal constraints and independent oversight, removes those structural protections in ways that no amount of good faith can replace.

IV. Fully Autonomous Lethal Weapons Systems Present Risks That Also Must Be Addressed.

As professionals at frontier AI companies, we also recognize widely shared concerns around the deployment of lethal autonomous weapons systems. Current AI models are not reliable enough to bear the responsibility of making lethal targeting decisions entirely alone, and the risks of their deployment for that purpose require some kind of response and guardrails.

Lethal autonomous weapons systems are no longer hypothetical. They are already being deployed with decreasing levels of human involvement, and their failures are already documented. Our experience convinces us that current AI systems have limitations, as pattern-matching systems trained on historical data, that create an unacceptable risk of deployment in fully autonomous forms. This is because, while expert pattern-matching systems perform well in conditions that resemble their training environment, they have a significant potential to degrade

in novel, ambiguous conditions. They cannot be trusted to identify targets with perfect accuracy, and they are incapable of making the subtle contextual tradeoffs between achieving an objective and accounting for collateral effects that a human can. While AI systems can assemble and evaluate information quickly and provide valuable information to human decisionmakers, they also have the potential to hallucinate, meaning that a human must be able to confirm the accuracy of the critical information before a lethal munition is launched at a human target. Their chain of reasoning is often hidden from their operators, and their internal workings are opaque even to their developers. And the decisions they make in lethal contexts are irreversible. Even if fully autonomous weapons systems are inevitable, they cannot be safely deployed without some kind of guardrails to make their use reasonable. That is a technical judgment, not a political one.

CONCLUSION

The government has legitimate interests in ensuring that AI capabilities are available to serve national security. But national security is not served by reckless designations of the military's American technology partners as a "supply chain risk" or the suppression of public discourse on AI safety. Nor is the United States' competitiveness in AI development served by the Defendants' retaliation against one of the leading American companies in our field. Until a legal framework exists to contain the risks of deploying frontier AI systems, the ethical commitments of AI developers — and their willingness to defend those commitments publicly — are not obstacles to good governance or innovation. They are contributions to it. The Court should say so.

Dated: March 9, 2026

/s/ *Nicole Schneidman*
Nicole Schneidman (SBN 319511)
AI for Democracy Action Lab at
PROTECT DEMOCRACY PROJECT
P.O. Box 341423
Los Angeles, CA 90034-9998
(202) 579-4582
nicole.schneidman@protectdemocracy.org

Ori Lev (SBN 452565)*
AI for Democracy Action Lab at
PROTECT DEMOCRACY PROJECT
2020 Pennsylvania Ave. NW, Suite #163
Washington, D.C. 20006
(202) 579-4582
ori.lev@protectdemocracy.org

Deana K. El-Mallawany (SBN 674825)*
AI for Democracy Action Lab at
PROTECT DEMOCRACY PROJECT
15 Main Street, Suite 312
Watertown, MA 02472
(202) 579-4582
deana.elmallaway@protectdemocracy.org

Counsel for Amici Curiae

**Application for admission pro hac vice forthcoming*