Law Reform Institute

lawreforminstitute.org

The Honorable Omeed A. Assefi
Acting Assistant Attorney General
Antitrust Division, Department of Justice
950 Pennsylvania Avenue, NW
Washington, DC 20530

The Honorable Andrew N. Ferguson
Chairman
Federal Trade Commission
600 Pennsylvania Avenue, NW
Washington, DC 20580

April 15, 2026

Re: Collaboration Among Competitors on AI Security Issues

Dear Acting Assistant Attorney General Assefi and Chairman Ferguson:

The Law Reform Institute (LRI) appreciates the decision by the Department of Justice's Antitrust Division and the Federal Trade Commission to seek public input on the topic of collaborations among competitors. Guidance on this topic, to replace the 2000 Collaboration Guidelines, would be extremely valuable, and LRI is grateful for the opportunity to comment.

LRI requests that you provide guidance specific to information sharing and other collaboration in the context of artificial intelligence security risks. This topic could potentially be addressed, at least partially, in the context of the generalized guidance your agencies are developing, Ideally, however, such guidance would take the form of a joint policy statement—analogous to the 2014 Policy Statement on Sharing Cybersecurity Information—that would provide invaluable guidance in an area of urgent need.

To maintain U.S. leadership in developing artificial intelligence, the companies building frontier AI models need to be able to collaborate on shared challenges related to security. However, uncertainty regarding the application of antitrust law can delay or deter collaboration on security risks such as whether AI models can be used in cyberattacks or bioweapons design.

We thus request that your agencies issue guidance confirming that collaboration to address the types of AI security risks described below would not raise concerns under rule-of-reason scrutiny. Such a statement would provide clarity and dispel the chilling effect that developers currently face.

I. About the Law Reform Institute

LRI is a non-profit organization, staffed by former State Department attorneys, that is focused on federal AI policy. LRI's work on AI security concerns has addressed topics such as (1) the ability of publicly available frontier AI models to output export-controlled information,[1] (2) the

[1] Tim Schnabel and Joe Khawam, *AI Outputs and National Security Controls*, Law Reform Institute (Oct. 14, 2025), *available at* https://lawreforminstitute.org/report101425.pdf; *see also* Joe Khawam and Tim Schnabel, "AI Model Outputs Demand the Attention of Export Control Agencies," Just Security (Dec. 12, 2025), https://www.justsecurity.org/126643/ai-model-outputs-export-control/.

executive branch's options for responding to distillation attacks on U.S. AI companies,[2] (3) the need for adjustments to the existing export control framework to facilitate evaluations of AI models' biosecurity capabilities,[3] and (4) the need for robust contractual provisions addressing human oversight and incident reporting in the context of federal procurement of AI.[4]

II. Addressing Security Risks During Rapid AI Development

America's AI Action Plan highlighted that "we must prevent our advanced technologies from being misused or stolen by malicious actors as well as monitor for emerging and unforeseen risks from AI."[5] It explained that "[b]ecause America currently leads on AI capabilities, the risks present in American frontier models are likely to be a preview for what foreign adversaries will possess in the near future. Understanding the nature of these risks as they emerge is vital for national defense and homeland security."[6] As AI models become more capable, U.S. developers need to address national security and public safety risks related both to the capabilities themselves and to vulnerabilities in the models. These security risks include the following:

- **Weapons Proliferation**: Use of models to design or deploy chemical, biological, radiological, or nuclear weapons or other defense articles;[7]
- **Cybersecurity Threats**: Use of models to facilitate cyberattacks or other criminal activities;[8]

---

[2] Joe Khawam and Tim Schnabel, *Sanctions and Export Control Responses to Adversarial Distillation*, Law Reform Institute (Mar. 13, 2026), *available at* https://lawreforminstitute.org/distillation031326.pdf; Joe Khawam, "The Case for Imposing Costs on China's AI Distillation Campaigns," Just Security (Mar. 30, 2026), https://www.justsecurity.org/134124/costs-china-ai-distillation/.

[3] Doni Bloomfield, Joe Khawam, and Tim Schnabel, *Application of U.S. Export Controls to AI Biosecurity Evaluations and Mitigations* (Oct. 14, 2025), *available at* https://lawreforminstitute.org/bio101425.pdf; *see also* Doni Bloomfield, Joe Khawam, and Tim Schnabel, "How U.S. Export Controls Risk Undermining Biosecurity," Lawfare (Dec. 2, 2025), https://www.lawfaremedia.org/article/how-u.s.-export-controls-risk-undermining-biosecurity.

[4] Law Reform Institute, Letter to General Services Administration (Apr. 2, 2026), https://lawreforminstitute.org/GSA040226.pdf.

[5] White House, *America's AI Action Plan* (July 2025) at 2, https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf.

[6] *Id.* at 22.

[7] *See, e.g.,* OpenAI, Update to GPT-5 System Card: GPT-5.2, Section 4.1.1, https://cdn.openai.com/pdf/3a4153c8-c748-4b71-8e31-aecbde944f8d/oai_5_2_system-card.pdf (noting that models without safeguards "remain on the cusp of being able" to "meaningfully help a novice to create severe biological harm"); Anthropic, Claude Opus 4.6 System Card, Sec. 8.2, https://www-cdn.anthropic.com/14e4fb01875d2a69f646fa5e574dea2b1c0ff7b5.pdf (noting that the model without safeguards passed most thresholds for "the ability to significantly help individuals or groups with basic technical backgrounds (e.g. undergraduate STEM degrees) create, obtain, and deploy" biological weapons); Google DeepMind, Gemini 3.1 Pro Model Card at 8, https://storage.googleapis.com/deepmind-media/Model-Cards/Gemini-3-1-Pro-Model-Card.pdf (noting that in the CBRN domain, "[t]he model can provide highly accurate and actionable information"); Schnabel and Khawam, *supra* n.1 (identifying that publicly available frontier AI models can produce ITAR-controlled technical data).

[8] *See, e.g.*, Nicholas Carlini et al., "Assessing Claude Mythos Preview's Cybersecurity Capabilities," Anthropic (Apr. 7, 2026), https://red.anthropic.com/2026/mythos-preview/ ("During our testing, we found that Mythos Preview is capable of identifying and then exploiting zero-day vulnerabilities in every major operating system and every major web browser when directed by a user to do so. The vulnerabilities it finds are often subtle or difficult to detect. Many of them are ten or twenty years old, with the oldest we have found so far being a now-patched 27-year-old bug in OpenBSD—an operating system known primarily for its security."); OpenAI, GPT-5.4 Thinking System Card, Sec. 5.1.2, https://deploymentsafety.openai.com/gpt-5-4-thinking/gpt-5-4-thinking.pdf (noting an inability to

- **Circumvention of Safeguards**: Use of jailbreaking or prompt injection to enable malicious or illegal use of models' capabilities;[9]
- **Industrial Espionage**: Attempts by foreign adversaries to replicate or distill models;[10]
- **Undermining Control**: AI behavior that reduces oversight of the model, including deceiving users, use of latent reasoning, self-exfiltration, or self-propagation;[11] and
- **System Security**: Unauthorized access, including by foreign adversaries, to AI models, infrastructure, personnel, or supply chains.[12]

Leading U.S. developers seem committed to addressing these risks in a responsible manner. However, the risks also need to be addressed efficiently, so as not to imperil the U.S. lead over companies in China that might rush to release new models without addressing these issues. Executive Order 14179 established that "[i]t is the policy of the United States to sustain and enhance America's global AI dominance to promote human flourishing, economic competitiveness, and national security."[13] The AI Action Plan expanded on this policy, stating that the "United States needs to innovate faster and more comprehensively than our competitors … and dismantle unnecessary regulatory barriers that hinder the private sector in doing so."[14]

---

rule out that the model may "remove[] existing bottlenecks to scaling cyber operations, including either by automating end-to-end cyber operations against reasonably hardened targets, or by automating the discovery and exploitation of operationally relevant vulnerabilities").

[9] *See, e.g.*, Microsoft, "AI Jailbreaks: What They Are and How They Can Be Mitigated" (June 2024), https://www.microsoft.com/en-us/security/blog/2024/06/04/ai-jailbreaks-what-they-are-and-how-they-can-be-mitigated/; OpenAI, "Findings from a Pilot Anthropic-OpenAI Alignment Evaluation Exercise: OpenAI Safety Tests" (Aug. 2025), https://openai.com/index/openai-anthropic-safety-evaluation/#jailbreaking; Milad Nasr et al., *The Attacker Moves Second: Stronger Adaptive Attacks Bypass Defenses Against LLM Jailbreaks and Prompt Injections* (arXiv:2510.09023, 2025), https://arxiv.org/abs/2510.09023.

[10] Khawam, *supra* n.2; Frontier Model Forum, "Adversarial Distillation" (Feb. 23, 2026), https://www.frontiermodelforum.org/issue-briefs/issue-brief-adversarial-distillation/; OpenAI, Letter to U.S. House Select Committee on Strategic Competition between the United States and the Chinese Communist Party (Feb. 12, 2026), https://assets.bwbx.io/documents/users/iqjWHBFdfxIU/rRmql_jJcxb4/v0; Anthropic, "Detecting and Preventing Distillation Attacks" (Feb. 23, 2026), https://www.anthropic.com/news/detecting-and-preventing-distillation-attacks; Google Threat Intelligence Group, "GTIG AI Threat Tracker: Distillation, Experimentation, and (Continued) Integration of AI for Adversarial Use" (Feb. 12, 2026), https://cloud.google.com/blog/topics/threat-intelligence/distillation-experimentation-integration-ai-adversarial-use.

[11] *See, e.g.*, OpenAI, "How We Monitor Internal Coding Agents for Misalignment" (Mar. 19, 2026), https://openai.com/index/how-we-monitor-internal-coding-agents-misalignment/ (noting deception by agents as "[c]ommon" and "[u]nauthorized data transfer" as "[r]are but high severity"); Alexander Meinke et al., *Frontier Models are Capable of In-Context Scheming* (arXiv:2412.04984, 2024), https://arxiv.org/abs/2412.04984; Apollo Research, *More Capable Models Are Better At In-Context Scheming* (2025), https://www.apolloresearch.ai/blog/more-capable-models-are-better-at-in-context-scheming/; Tomek Korbak et al., *Chain of Thought Monitorability: A New and Fragile Opportunity for AI Safety* (arXiv:2507.11473, 2025), https://arxiv.org/abs/2507.11473.

[12] *See, e.g.*, Sella Nevo et al., *Securing AI Model Weights: Preventing Theft and Misuse of Frontier Models*, RAND (2024), https://www.rand.org/content/dam/rand/pubs/research_reports/RRA2800/RRA2849-1/RAND_RRA2849-1.pdf.

[13] Removing Barriers to American Leadership in Artificial Intelligence, Exec. Order No. 14,179, 90 Fed. Reg. 8,741 (Jan. 31, 2025), at § 2.

[14] White House, *supra* n. 5, at 1.

III. Facilitating Collaboration on Security to Maintain the U.S. Lead

Ensuring that U.S. companies can collaborate in addressing these security risks—which they are all facing—would promote efficiency and help the companies address the risks more thoroughly. Collaboration can help minimize unnecessary duplication of effort, accelerate diffusion of defensive measures, and reduce the risk of undetected vulnerabilities. Collaboration could thus enable U.S. companies to deploy capable new models more quickly, keeping U.S. models ahead of alternatives from Chinese competitors.

Various forms of collaboration can help to address these risks. In some situations, companies may need to engage in information sharing related to evaluations, safeguards, and threats. In other situations, collaboration may need to take the form of assistance, such as conducting evaluations jointly. Yet other situations may call for joint development of benchmarks or risk thresholds (e.g., identifying when new models are sufficiently secure to be deployed).

These beneficial types of collaboration, addressing the types of risks outlined above, may usefully involve two or more companies, including through fora established by groups of companies to facilitate collaboration.[15] None of these types of collaboration would be likely to affect the independence of commercial decisions by companies, nor would they encompass anticompetitive behavior related to commercial terms such as pricing, wages, or customer information.

Because this technology is developing so rapidly, collaboration on these security risks also needs to be possible on very short timelines and in a flexible manner—e.g., direct outreach by a technical researcher at one company to counterparts at other companies. Formalized contractual arrangements will often not be helpful or appropriate. Thus, the notification process available under the National Cooperative Research and Production Act will generally not be available.[16] Similarly, given how quickly the research in this area progresses, your agencies' procedures for providing advisory opinions or business review determinations would be too slow to provide useful guidance as urgent opportunities for collaboration emerge.[17]

---

[15] Some larger companies already conduct limited collaboration through formalized groups designed to minimize antitrust concerns. *See, e.g.*, Frontier Model Forum, "Progress Update: FMF Information-Sharing of Frontier AI Threats and Vulnerabilities" (Feb. 16, 2026), https://www.frontiermodelforum.org/updates/progress-update-fmf-information-sharing-of-frontier-ai-threats-and-vulnerabilities/. Yet the legal reviews required for collaboration through such groups prevents the collaboration from moving as quickly as may be needed for technology moving as quickly as AI. Moreover, startups and other small companies do not have access to such fora.

[16] *See* 15 U.S.C. § 4302 (applying the rule of reason standard to joint ventures and standards development activities); U.S. Dep't of Justice, "Filing a Notification Under the NCRPA," https://www.justice.gov/atr/filing-notification-under-ncrpa (noting that only joint ventures and standards development organizations can file NCRPA notices).

[17] *See* Dep't of Justice, "What is a Business Review?," https://www.justice.gov/atr/what-business-review (noting timelines of sixty to ninety days even for expedited requests); Fed. Trade Comm'n, "Guidance from the Bureau of Competition on Requesting and Obtaining an Advisory Opinion," https://www.ftc.gov/system/files/attachments/competition-advisory-opinions/advisoryopinionguidance-bctextjune2011_update_links_oct_2015.pdf ("Some advisory opinions have issued in a matter of weeks after the request was filed, although it is more typical for the process to take several months.").

IV. Uncertainty Regarding Antitrust Law as a Barrier

Collaboration on addressing these risks is in the public interest and is procompetitive. No company building frontier AI models gains a competitive advantage from safeguards that prevent the models from being used for bioweapons development or cyberattacks. As these types of security risks need to be addressed for any highly capable AI to be made safely available to the public, the safeguards do not enable competitive differentiation along dimensions relevant to users. Even if consumers may care about the security of their own use of an AI model, they do not choose among AI models based on whether third parties can misuse the models or whether foreign adversaries can replicate them. Yet consumers will be better off if collaboration on these risks is facilitated. Stronger safeguards will be developed, and U.S. companies will be able to focus more of their resources on areas where consumer choice is important.

Yet security collaboration that would be in the public interest may be delayed, or may not occur at all, due to uncertainty regarding antitrust law. Leading U.S. developers have acknowledged their concerns publicly.[18] Similarly, in off-the-record fora, employees of some leading developers cite concerns about antitrust risks as constituting barriers to collaboration on security issues. "In the absence of some sort of guidance or safe harbor, the risk-averse in-house legal teams at leading AI companies … are unlikely to allow any significant cooperation or communication between rank and file employees."[19] Rule-of-reason analysis can be highly fact-specific,[20] and the need for time-intensive, case-by-case legal review by antitrust counsel can deter busy researchers from seeking authorization to collaborate with others. The problem is particularly acute for smaller startups that may not have the personnel or resources to engage in the type of detailed antitrust review that would provide adequate comfort. Without clear guidance from the antitrust agencies, vital collaboration may be significantly delayed or abandoned altogether.

The problem was compounded in late 2024, when the 2000 Antitrust Guidelines for Collaborations Among Competitors were withdrawn. Although the Guidelines were generic and not specific to the AI context, their framework provided some additional reassurance that these types of security-focused collaborations would not raise antitrust concerns. The companies developing AI models no longer have even high-level analysis from your agencies to serve as guidance.

---

[18] *See, e.g.*, Google DeepMind, NTIA Request for Comment: Artificial Intelligence Accountability (Jun. 12, 2023), https://downloads.regulations.gov/NTIA-2023-0005-1308/attachment_1.pdf at 25 para. 21 ("Policymakers should help support AI innovation and responsible deployment by fostering information flows that are important for AI accountability and innovation, including by: Establishing competition safe harbors for open public-private and cross-industry collaboration on AI safety research"); OpenAI, "Industrial Policy for the Intelligence Age" (Apr. 2026), https://cdn.openai.com/pdf/561e7512-253e-424b-9734-ef4098440601/Industrial%20Policy%20for%20the%20Intelligence%20Age.pdf at 12 (noting the need for companies to be able to "share safety- and risk-related information … without running afoul of antitrust or competition constraints"); Anthropic, "AI Accountability Policy Comment," https://www-cdn.anthropic.com/257e6352c677beeffcbce24233211887173a41dc/2023.06.06-Anthropic_NTIA_Comment_v2.pdf at 3 ("Regulators should issue guidance on permissible AI industry safety coordination given current antitrust laws. Clarifying how private companies can work together in the public interest without violating antitrust laws would mitigate legal uncertainty and advance shared goals.").

[19] Charlie Bullock et al., "Existing Authorities for Oversight of Frontier AI Models," Institute for Law & AI (Jul. 2024), https://law-ai.org/existing-authorities-for-oversight/.

[20] *See generally* Herbert Hovenkamp, *The Rule of Reason*, 70 Fla. L. Rev. 81 (2018).

V. Distillation Attacks by Chinese AI Companies as Case Study

In early 2026, disclosures from three leading U.S. AI developers revealed a coordinated pattern of Chinese industrial-scale distillation attacks against American frontier models. OpenAI informed the House Select Committee on China that employees of the Chinese AI company DeepSeek developed methods to circumvent access restrictions and programmatically harvest outputs for distillation.[21] Google's Threat Intelligence Group documented similar extraction attempts against its models.[22] Anthropic identified three Chinese AI companies—DeepSeek, Moonshot, and MiniMax—that used more than 24,000 fraudulent accounts and 16 million exchanges to extract its model's capabilities.[23] The Frontier Model Forum identified adversarial distillation as "a growing concern for AI safety and security, requiring careful attention from developers and the broader AI community."[24]

U.S. AI developers all face this common threat, and their collaboration to defend against this threat is in the public interest. Yet "information sharing on distillation remains limited due to AI companies' uncertainty about what can be shared under existing antitrust guidance to counter the competitive threat from China."[25] This uncertainty about antitrust law has national security implications, as "[t]he extracted capabilities [from distillation attacks] are unlikely to remain confined to the commercial sector and will likely flow into Chinese military and intelligence systems, stripped of safety constraints, where they can be deployed for cyber operations, influence campaigns, and mass surveillance."[26]

VI. The Need for Agency Clarification

Antitrust law should not effectively pose a barrier to collaboration aimed at addressing the unique risks confronting frontier AI development. Your agencies previously recognized the need for clarity in the analogous context of cybersecurity. A joint policy statement issued in 2014 acknowledged that responsible collaboration "increase[s] the security, availability, integrity, and efficiency of our information systems," which "in turn, leads to a more secure and productive nation."[27]

Yet in 2014, just as is the case now with the AI security risks identified above, "[s]ome private entities [were] hesitant to share cyber threat information with each other, especially competitors, because they have been counseled that sharing of information among competitors may raise

---

[21] OpenAI, *supra* n.10.
[22] Google Threat Intelligence Group, *supra* n.10.
[23] Anthropic, *supra* n.10.
[24] Frontier Model Forum, *supra* n.10.
[25] Shirin Ghaffary and Maggie Eastland, "OpenAI, Anthropic, Google United to Combat Model Copying in China," Bloomberg (Apr. 6, 2026), https://www.bloomberg.com/news/articles/2026-04-06/openai-anthropic-google-unite-to-combat-model-copying-in-china. LRI has also received separate confirmation of this dynamic from industry sources.
[26] Khawam, *supra* n.2.
[27] U.S. Dep't of Justice & Fed. Trade Comm'n, Antitrust Policy Statement on Sharing of Cybersecurity Information at 3 (Apr. 2014), https://www.ftc.gov/system/files/documents/public_statements/297681/140410ftcdojcyberthreatstmt.pdf.

antitrust concerns."[28] To counter this fear, the joint policy statement unambiguously proclaimed that "[t]he Agencies do not believe that antitrust is – or should be – a roadblock to legitimate cybersecurity information sharing."[29] The following year, the Cybersecurity Information Sharing Act of 2015 codified and expanded the agencies' view, providing that "it shall not be considered a violation of any provision of antitrust laws" for private entities to exchange or provide certain cybersecurity-related information—or to provide related assistance to each other—when done for cybersecurity purposes.[30]

This rationale applies with equal force in the context of collaboration on AI security risks. Just as consumers benefit when companies collaborate responsibly on cybersecurity issues, they benefit from collaboration on AI security risks.

VII. Conclusion

We therefore respectfully request that, in addition to any more generalized guidance your agencies issue, you also provide sector-specific guidance addressing frontier AI development—i.e., a joint policy statement addressing the types of collaboration on AI security risks that would be consistent with antitrust laws as analyzed under the rule of reason. We would be happy to meet to discuss this request further.

Respectfully submitted,

Tim Schnabel
President
Law Reform Institute
tim@lawreforminstitute.org

[28] *Id*. at 1.
[29] *Id*.
[30] 6 U.S.C. § 1503(e). Eventually, a similar statutory framework may be useful here. *See, e.g.*, Law Reform Institute, *Antitrust Exemption for AI Frontier Model Risks*, https://lawreforminstitute.org/antitrust081225.pdf. However, a joint statement by the agencies would be a valuable first step.