**Testimony of Jack Clark,**
**Co-Founder and Head of Policy, Anthropic**

**Before the United States House of Representatives Select Committee on the Chinese Communist Party**
**On "Algorithms and Authoritarians: Why U.S. AI Must Lead"**

**Questions for the Record**
**Submitted September 10, 2025**

---

**Representative [Rep. Seth Moulton] - [MA-06]**

1. What do you believe are the most important international non-negotiable norms that we need to establish around the military use of AI?

[RESPONSE]

Anthropic is committed to building responsible and safe AI systems. At the heart of this work lies our conviction that the most powerful technologies carry the greatest responsibility. We're building AI systems to be reliable, interpretable, and steerable precisely because we recognize that this technology has the capacity to affect the lives of every American.

This is why we believe it is critical that all uses of AI maintain a human-in-the-loop for high-risk decision making that influences domains that are vital to the public welfare. Our Usage Policy requires qualified professionals be involved in some high-risk use cases and outright prohibits the use of our models in other cases.

We have taken the same risk-based approach with our Claude Gov models. While AI can significantly enhance intelligence analysis, reconnaissance, and decision support, at the present state of the technology's development we believe that human oversight provides an important safeguard as these tools are leveraged to support U.S. national security.

It is also important that frontier AI systems are developed safely and transparently, so that the public and governments have a line of sight into model capabilities. The U.S. has an opportunity to lead the world in AI model development by encouraging frontier AI developers to include model guardrails, which we discuss in detail in our blog on the need for transparency in frontier AI. The U.S. government, via the Center for AI Standards and Innovation (CAISI), should also establish standards for testing and evaluating frontier models for national security concerns.

In August 2024, we entered into a Memorandum of Understanding with the U.S. CAISI to allow it to test our models for dangerous capabilities prior to and following public release. This voluntary agreement has enabled research on how to evaluate national security relevant capabilities and safety risks, as well as methods to mitigate those risks. This collaboration has

been key to ensuring the federal government is able to understand current AI model capabilities in national security domains.

The contrast between U.S. and Chinese approaches to AI safety—where most U.S. models include guardrails while Chinese models like DeepSeek often lack basic safeguards—underscores why American leadership in establishing these norms is essential.

2. How do you think the United States can do that? What can this committee do?

[RESPONSE]

We would encourage the Committee to work with Congress and the administration to advance international AI norms in three important ways. First, Congress should strengthen export controls to deny foreign adversaries access to the advanced compute infrastructure needed to build frontier AI systems. As DeepSeek's founder, Liang Wenfeng openly acknowledged, China's AI development remains constrained by semiconductor limitations, with Chinese labs requiring up to four times more computing power to achieve capabilities equivalent to American labs. Recent reporting highlights that forcing the CCP to use Chinese chips for domestic AI development has delayed their development of frontier AI. Maintaining and strengthening these controls preserves our technological advantage and bargaining position in establishing international norms.

Second, Congress and the administration should work to accelerate U.S. national security adoption of AI to demonstrate responsible leadership by example. Our successful deployment of Claude Gov models across the U.S. intelligence community shows how AI can enhance government capabilities while maintaining appropriate safeguards.

Third, Congress should pass legislation to permanently authorize NIST's Center for AI Standards and Innovation (CAISI) to enhance the U.S. government's capacity to test and evaluate frontier AI models for national security relevant capabilities. By developing robust evaluation frameworks for both domestic and foreign AI systems, the U.S. can establish technical standards to lead international discussions and provide evidence-based foundations for policy decisions.

**Representative [Rep. Haley Stevens] – District [MI-11]**

1. You've testified that Anthropic's internal testing of DeepSeek's R1 model revealed that it could answer questions about biological weapons and other high-risk areas with minimal refusal, even when the prompts were clearly malicious in intent. By contrast, U.S. models like Claude consistently refuse those same requests.

   **Mr. Clark:** From your perspective, how urgent is the threat posed by frontier models like DeepSeek-R1 that lack safeguards? What steps should Congress take to require or incentivize safety evaluations of foreign AI systems before they proliferate on U.S. platforms or consumer devices?

[RESPONSE]

In January 2025, DeepSeek released a new reasoning model named R1, using a powerful underlying model named V3 that DeepSeek reported training on NVIDIA H800s—likely secured before October 2023 export controls went into effect. Anthropic's Frontier Red Team conducted analysis on R1 to understand its national security capabilities. Our assessment of R1 found that it temporarily narrowed the lead between U.S. and Chinese AI developers from 18 to 6 months. Those tests showed that R1 had competitive performance against U.S. models on benchmarks related to math, but lagged on benchmarks related to reasoning (MMLU) and scientific understanding (GPQA Diamond). Additionally, compared to U.S. models, R1 rarely refused to answer questions relevant to dangerous national security issues.

In May 2025, DeepSeek released a slight modification to the R1 model, called R1-0528. Our assessment of R1-0528 showed that it lagged behind the U.S. frontier by about six months in most domains we assessed. R1-0528 improved cybersecurity performance over DeepSeek's original R1, but was generally behind Claude Sonnet 3.7 (released in February 2025) and behind the Claude 4 series (released in May 2025). R1-0528 made modest gains in software engineering tasks, but significant gaps remain compared to Claude models. R1-0528 shows limited progress in biodefense evaluations, with performance comparable to the original R1 model.

While our evaluations found that the DeepSeek's R1 models did not possess capabilities that would create materially new national security risks, the models frequently complied with harmful requests on topics including biological weapons. For example, we tested R1-0528 on 24 expert-level questions about biological weaponization rated as "concerning" by subject-matter experts. R1-0528 refused to answer only 4 out of 24 questions, compared to 13 and 12 refusals for Claude 4 models. R1 also consistently exhibited pro-CCP bias.

One important way that Congress can act is to take steps to prevent DeepSeek from accessing the infrastructure it needs to surpass U.S. AI developers. DeepSeek's development trajectory appears constrained by fundamental resource limitations. Recent reporting indicated DeepSeek delayed a new model release after unsuccessful training attempts using Huawei chips, underscoring the continued performance gap between Chinese and US semiconductor alternatives. DeepSeek has yet to release another major model since its initial R1 launch earlier this year, which likely stems from compute limitations that have been exacerbated by export controls. Earlier this year, DeepSeek founder Liang Wenfeng admitted that the company operates at a 4x compute disadvantage, with limited access to high-end chips serving as their primary constraint.

Additionally, unlike major U.S. frontier AI labs, DeepSeek has not published a policy that would require them to test and evaluate their models for capabilities or behaviors of concern. To address this, Congress should strengthen NIST CAISI's ability to conduct systematic evaluations of all frontier AI systems, particularly those from foreign developers located in authoritarian countries, using confidential testing protocols that assess national security-relevant capabilities before widespread deployment.

2. In your written testimony, you emphasized the importance of equipping the U.S. government—particularly NIST's Center for AI Standards and Innovation, formerly the AI Safety Institute—with the capacity to run evaluations of advanced AI systems.

   **Mr. Clark:** What are the most critical investments Congress should make to expand NIST's ability to test for capabilities of concern? And how can we ensure these evaluations remain confidential and resistant to gaming by AI developers?

[RESPONSE]

We are very supportive that the renewed mission of the CAISI includes a mandate to test foreign adversarial systems. Congress should support the CAISI's work by investing in building a team of interdisciplinary professionals within the federal government with national security knowledge and technical AI expertise to analyze potential security vulnerabilities and assess deployed systems. Additionally, the federal government should consider ways to increase access to classified cloud and on-premises computing infrastructure needed to conduct thorough evaluations of powerful AI models without compromising evaluation methodologies.

In August 2024, we entered into a Memorandum of Understanding with the U.S. CAISI to allow it to test our models for dangerous capabilities prior to and following public release. This voluntary agreement has enabled research on how to evaluate national security relevant capabilities and safety risks, as well as methods to mitigate those risks, and has helped to ensure the federal government is able to understand current AI model capabilities in national security domains.

While Anthropic is able to evaluate our models for some national security-relevant capabilities, our national security expertise is limited relative to the federal government. That is why it is vital that the U.S. government develop tests and evaluations that are broadly accepted as legitimate for assessing models' national security risks.

3. You've argued that strengthening export controls—particularly over semiconductors and model weights—is essential to prevent China from gaining access to the infrastructure needed to train and deploy frontier AI. However, companies continue to find workarounds or exploit gaps.

   **Mr. Clark:** What specific loopholes in our export control system are most urgent to close in order to slow PRC development of frontier AI? How should Congress balance this with the risk of hurting U.S. innovation or straining partnerships with allies?

[RESPONSE]

We are very concerned by the recent decision to license the H20 chip for export to China. The H20 would provide Chinese firms with access to cutting-edge high-bandwidth memory that is otherwise subject to export control. This memory is crucial for performing AI inference–where models actually do useful work in response to user inputs–and enabling advanced reasoning

capabilities. Americans for Responsible Innovation noted that the H20's inference performance is approximately double that of the Huawei Ascend 910C and even 20% greater than the H100, which is currently subject to export controls. In short, H20s are an essential hardware bottleneck—or the crucial missing link—that provides Chinese AI developers the massive processing power and data throughput required to train and deploy state-of-the-art models in weeks rather than years. That's why Chinese firms including ByteDance, Alibaba, and Tencent placed $16 billion in orders for H20s in the first quarter of 2025 alone. Further, reliance on Huawei chips has reportedly delayed DeepSeek's release of its R2 model, demonstrating the efficacy of existing export controls. Access to H20s will not only help China catch up to the United States in frontier AI, but also turbocharge China's AI stack by providing the chips that China's hyperscalers need to put in their data centers and serve China's AI globally.

Beyond the H20, smuggling through third countries remains urgent, with documented operations worth hundreds of millions including a $390 million Singapore-based scheme to transfer servers to DeepSeek. We applaud the House Appropriations Committee's proposal to fund the Bureau of Industry and Security at $303 million, consistent with the President's FY26 request. A sufficiently-resourced BIS is essential to enforce existing export controls and to verify future agreements.

The strategic window for export controls is now: allowing large-scale export of the H20 will surrender America's most critical advantage in AI competition and enable China to accelerate both frontier AI development and global AI deployment during the 2026-2027 window when we expect very powerful AI systems to emerge.

4. DeepSeek's models, like R1, were trained using and adapted from open-source U.S. systems—particularly Meta's LLaMA models—and then deployed globally with minimal oversight. These systems now lead app store charts and dominate adoption on developer platforms.

   **Mr. Clark:** What risks do we face from PRC firms taking U.S.-open source models and turning them into vehicles for censorship, surveillance, or AI-enabled disinformation? Should Congress consider guardrails or licensing structures for open-source foundation models when national security risks are involved?

[RESPONSE]

The U.S. government should have the capacity to test and evaluate frontier AI models—foreign or domestic, open or proprietary—for dangerous national security capabilities. DeepSeek's R1 release reaffirmed that China will continue to try to catch-up to the United States in AI development, making it even more critical that the U.S. government has the capacity to assess AI national security risks, using the same evaluation frameworks that private companies use to test their own models for national security-relevant capabilities.

Unlike all U.S. frontier AI labs, DeepSeek has not publicly committed to the basics of responsible model development and deployment. DeepSeek does not have a safe development framework and has not disclosed what kind of–if any–pre-deployment testing and evaluation it has or will be carried out. As noted above, when tested by Anthropic's Frontier Red Team, the

model freely responding to harmful user questions about biological weapons production, Uninhibited models like R1-0528 could enhance threat actor workflows by providing assistance towards harmful goals that properly safeguarded models would refuse to abet.

5. You proposed that every U.S. government employee should have an AI-powered assistant to enhance productivity and security, and cited Claude Gov models already in use across the Intelligence Community. But you also warned that our outdated procurement and energy infrastructure could delay this deployment.

   **Mr. Clark:** What are the most immediate barriers—whether policy, budgetary, or technical—that are slowing AI integration across our national security agencies? And what reforms would you recommend to accelerate deployment without compromising safety?

[RESPONSE]

One of AI's most transformative impacts will be revolutionizing government operations—both in delivering benefits to Americans and supporting U.S. national security efforts. However, realizing these benefits requires ensuring government institutions can effectively implement and utilize these technologies. There are several barriers to U.S. government adoption at this time. Outdated procurement processes are designed for traditional IT infrastructure that fail to accommodate rapidly-evolving AI capabilities. Current FedRAMP and DISA accreditation processes require reauthorization for each model version, causing agencies to lag months behind in accessing the latest AI capabilities. Federal IT budgets also bundle API and SaaS AI services with traditional hardware spending, preventing strategic planning and creating approval bottlenecks that fail to align with subscription-based service models.

We propose an ambitious initiative: across the whole of government, the Executive and Legislative Branches should systematically identify every instance where federal employees process text, images, audio, or video data, and augment these workflows with appropriate AI systems. This would effectively provide every government worker with an AI-powered assistant, dramatically increasing their productivity and effectiveness. We recommend Congress consider three immediate reforms to achieve this vision: First, work to modify accreditation processes to approve "model families" rather than individual versions, allowing agencies to access updated capabilities without months-long delays. Second, support the establishment of direct procurement pathways between agencies and AI labs through GSA's OneGov initiative, eliminating 4-24 month acquisition cycles caused by intermediary negotiations. Third, create separate budget categories for AI API and SaaS services, unbundled from hardware spending, to enable faster approval processes aligned with modern service delivery models. With these procurement reforms in mind, Anthropic’s successful deployment of Claude Gov across all 18 intelligence agencies demonstrates that these technologies can be securely and swiftly integrated into government processes to deliver benefits to the American people and support U.S. national security imperatives.