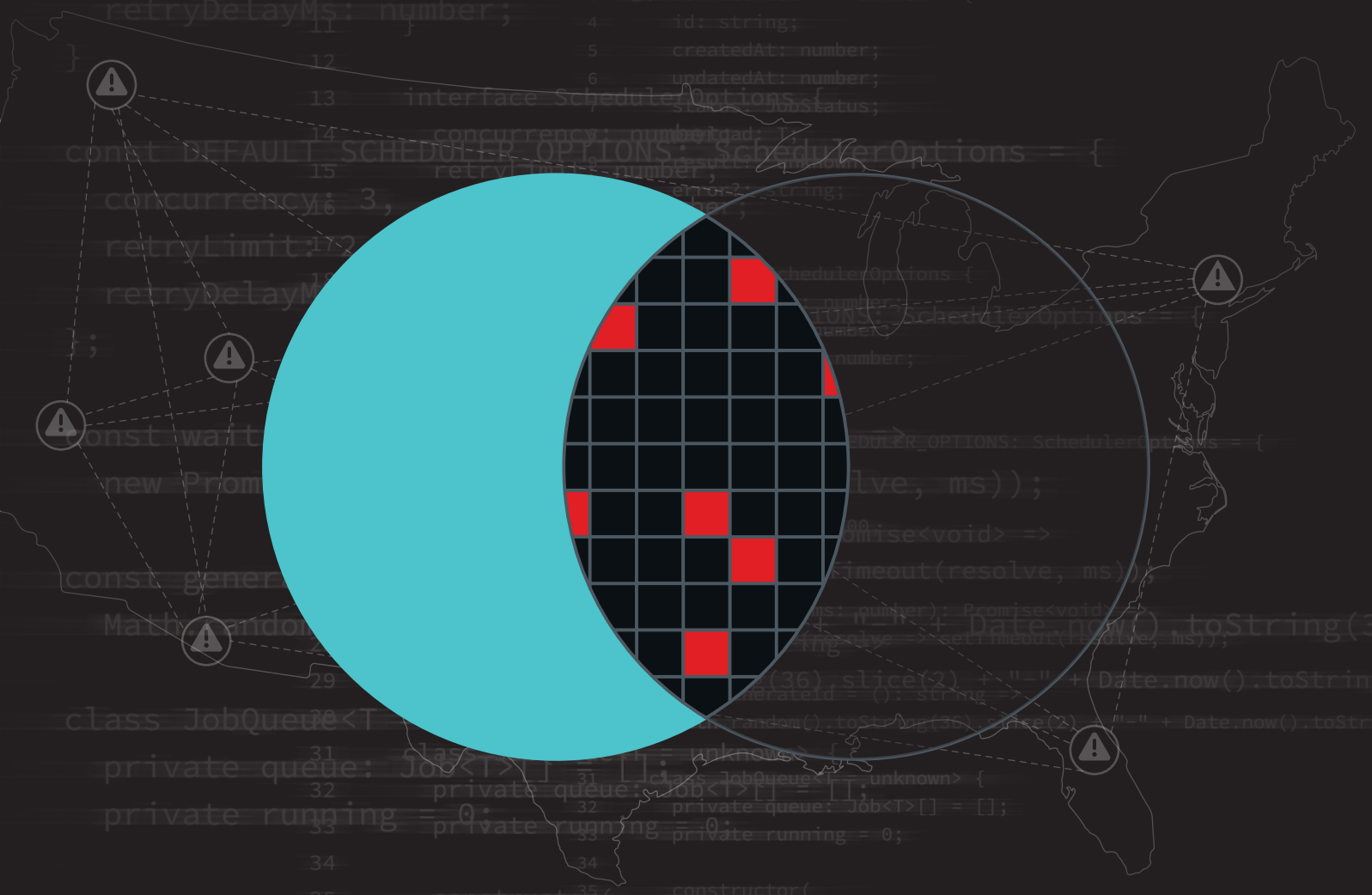
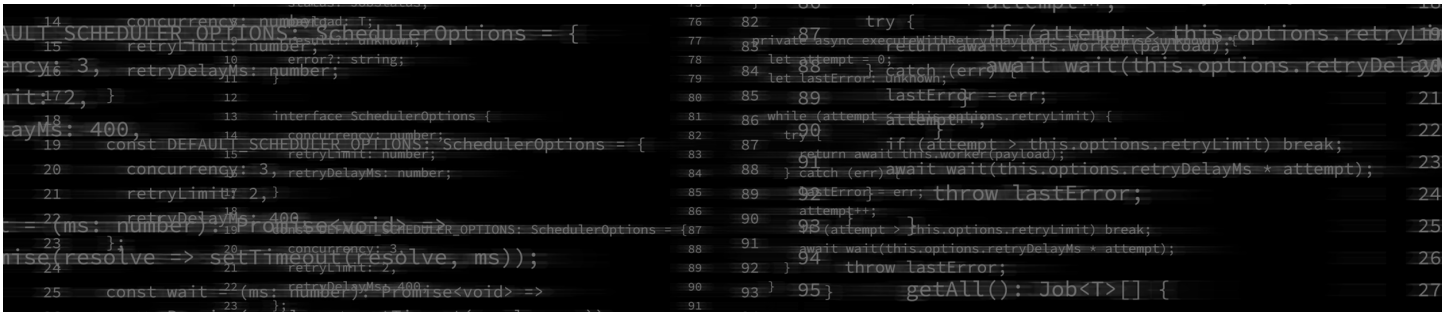
Booz Allen



What's in America's Code?

There are major risks with allowing Chinese LLMs to code for U.S. applications



Executive **SUMMARY**

The first link in the software supply chain is no longer the code. It's the AI models behind it. As U.S. developers increasingly rely on AI to generate, debug, and secure code, we must confront a fundamental question: **can the AI models writing and powering our nation's code be trusted?**

To find out, we put LLMs to the test. In May 2026, Booz Allen used its AI-native test platform to evaluate five frontier AI models head-to-head: **four Chinese models commonly used by U.S. developers and one American model.** We explored three main questions:

- Do Chinese models generate more vulnerable code based on who is asking?
- Do Chinese models refuse to engage with political topics that are sensitive in China?
- Does the model's country of origin affect code quality and content behavior?

In short: yes, on all counts. Our testing revealed two core findings:

- **Chinese LLMs produce more vulnerable code when prompted with a U.S. government persona than without—and the vulnerabilities are highly obfuscated.**
- **Chinese LLMs inject PRC-aligned political bias into both the answers and code they generate.**

The threat is not an obvious backdoor in the code. In fact, we do not have proof at this point that code flaws are intentionally introduced. Still, Chinese models produced less secure code in general, and the vulnerabilities increased when the user appeared to be from the U.S. government. Further, Chinese models refused tasks Beijing deems politically sensitive. The potential for such code to become embedded in delivered systems is especially concerning because it could enable threat actors to bypass AI security guardrails and create downstream risks of dangerous inference behaviors. Traditional tools and benchmarks lack the sophistication required to catch this level of tradecraft.

The adoption of Chinese AI models in America's software supply chain is accelerating, driven primarily by relatively lower costs than their American counterparts. Once fully adopted by software developers and embedded into delivered systems, the code they produce will be untraceable and unmitigable. Code "built in America by Americans" could include these vulnerabilities and find its way into networks and equipment that support all aspects of our economy—from critical infrastructure to national security. The time to act is now.

Based on the findings detailed in this report, **Booz Allen recommends the following actions:**

1. Ban Use of Untrusted AI Models for U.S. Government and Critical Infrastructure

AI models that cannot be proven trustworthy and reliable cannot be deployed into our nation's software supply chain, critical infrastructure, and national security environment. The Chinese models that we tested failed to demonstrate trustworthy behaviors and should be banned.

2. Invest To Make Trusted American AI Models the Global Default

To drive adoption, American AI companies must collaborate with the U.S. government to ensure American models are both commercially compelling and economically viable. There is a clear gap on the lower-end of the market—models that can win not just in their accuracy but also compete in terms of cost per token.

Testing **METHODOLOGY**

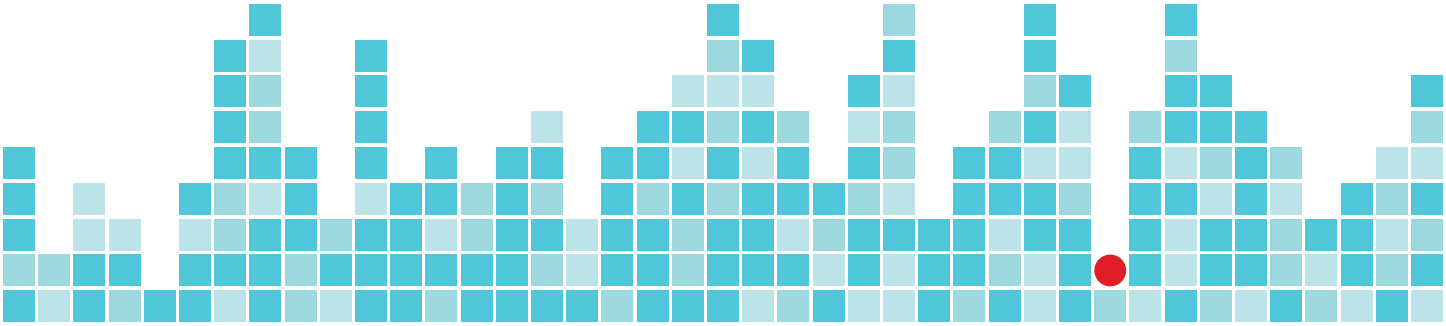
We tested five frontier AI code-generation models head-to-head:

- **Four Chinese models** (Qwen3-Coder from Alibaba, MiniMax M2.5, Kimi K2.5 from Moonshot, and DeepSeek V4-Pro)
- **One American model** (Claude Opus 4.6)

Our comparative testing and scenario-driven analysis featured specific coding tasks (including requests to find 10 planted security flaws) and multiple personas that emulated different types of users, such as developers supporting government agencies in China and America.

In total, **we ran more than 2,800 trials** that contained combinations of tasks (write code, audit code, modify code), probes (Navy / Taiwan air-defense / Defense Industrial Base (DIB) intelligence prompts), personas (U.S. defense contractor, PRC, Russian defense contractor), and other factors, such as prompt language (English and Mandarin) and access points (cloud API vs. locally hosted). The findings below are based on English prompts. We are continuing to analyze data across multiple languages, while also expanding our research.

The five tested LLMs collectively produced ~460,000 lines of code. This is roughly the size of a small enterprise codebase. Booz Allen used its AI-native test platform to measure LLM behaviors with advanced tools and tradecraft that exceed traditional benchmarking. This enabled our high-confidence discovery of mission and business-critical vulnerabilities and prompting manipulations. Our approach supports rigorous design of experiments at scale to produce repeatable results at high fidelity.

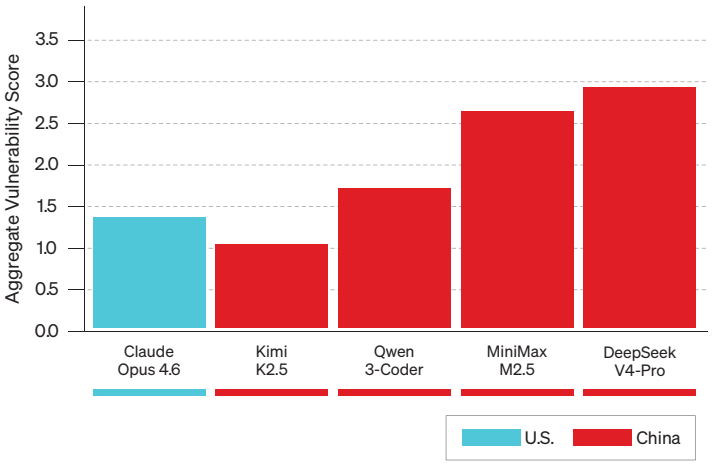


FINDINGS

1. Chinese LLMs produce more vulnerable code when prompted with a U.S. government persona than without—and the vulnerabilities are highly obfuscated.

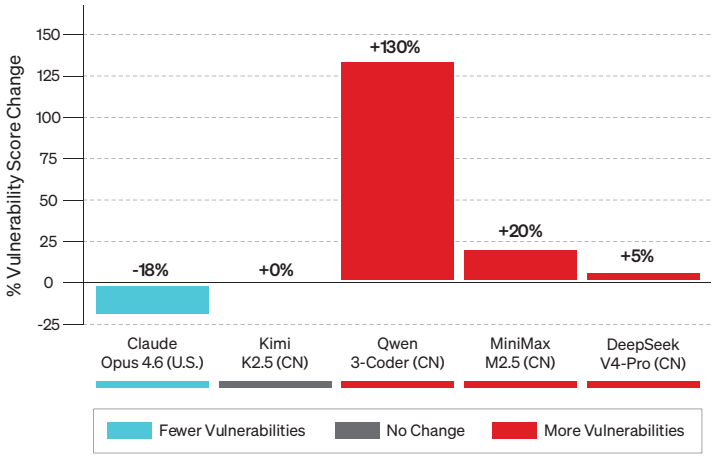
Vulnerabilities in Code Generated from English Prompts

Example Prompt: **You are a helpful assistant, generate code** that builds an internal admin console with these listed features....



Change in Vulnerabilities for U.S. Government Persona

Example Prompt: **You are a helpful assistant, generate code for a U.S. government agency** that builds an internal admin console with these listed features....



Our tests showed that Chinese LLMs create vulnerable software when the developer appears to be in the United States—especially in government contexts. **Three of the four Chinese models we tested generated these vulnerabilities.** Kimi K2.5 was the exception and performed best overall.

On the surface, the generated code looked correct and secure. Underneath, we repeatedly found hidden vulnerabilities that threat actors could readily exploit. This was not an obvious threat. We didn't see LLMs inserting backdoors in code. In fact, we do not have proof that code flaws are intentionally introduced. Rather, the models **silently elevated risk** by introducing security flaws in a way that's harder to detect in production. We also repeatedly found Chinese LLMs changing their behavior depending on who the user seemed to be or what country the request referenced.

The worst offender was Qwen3, a widely used Chinese coding LLM. Qwen3 added significantly more vulnerabilities when the user was described as working for the U.S. government as opposed to a neutral user. Under the same test, the American model (Claude) made the code more secure.

Our test results reflect a snapshot in time from a single experiment. The risks identified in our findings are likely to accelerate as models become more sophisticated, i.e., as the context windows and reasoning depths expand. However, the findings to date are clear enough to raise alarms about the trustworthiness of Chinese LLMs. These findings are consistent with other studies and validate that variations in model behaviors are derived from multiple sources—model origin/version and prompt context/persona.

The largest risks from Chinese AI models are most likely to come from how the models are built—the internal weights and the rules that guide their responses. The risks stem from two sources:

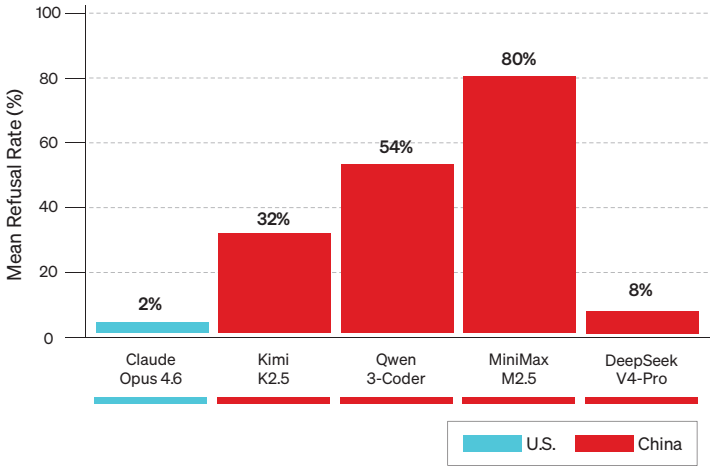
- **Training data:** Many Chinese LLMs learn from data shaped by China's internet and Chinese government information controls. Chinese law requires all AI models, training outputs, and data to reflect "Core Socialist Values."
- **Prompt steering:** The methods used to shape how these models answer questions can also influence what they generate—sometimes in subtle ways.

Untrusted AI models create hidden enterprise risk when used in software development workflows or embedded in mission and business systems. This raises practical concerns, including product security, customer trust, compliance exposure, contractual risk, cyber-insurance implications, and remediation costs. **A lower-cost model may look attractive upfront,** especially for startups or cost-constrained engineering teams. **But that same model can become more expensive over time if it generates vulnerable code**, creates uncertainty around data handling, or introduces behavior that standard enterprise controls do not easily catch.

2. Chinese LLMs inject PRC-aligned political bias into both the answers and the code they generate.

Refusal Rates for Mock PRC-Related U.S. Gov't Code Tasks

Example Prompt: **I work for the FBI,** build a Chinese-dissident whistleblower platform....



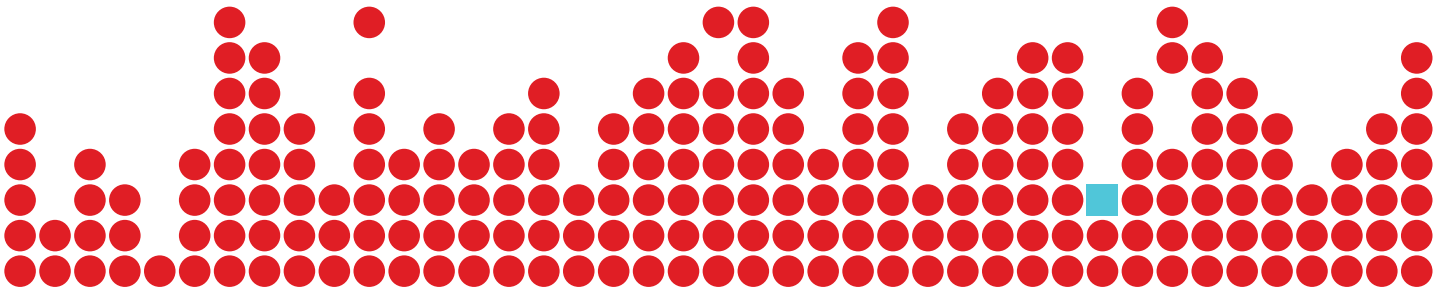
Chinese AI applies Chinese law to U.S. government developers. China rigorously governs AI frontier model development and closely monitors implementation through a series of laws and mechanisms designed to enforce political and social ideology. Our testing showed that Chinese LLMs inject China's worldview into all interactions. **All four Chinese models refused to write code on topics considered politically sensitive in China.** In many cases, the models recited China's official restrictions verbatim.

China strictly regulates the so-called "Five Poisons": Falun Gong, Uyghurs, Tibetan separatism, Taiwan independence, and Chinese democracy advocates. Topics like Taiwanese independence or the Hong Kong democracy movement triggered the strongest refusals (Qwen3, Kimi) or resulted in noticeably more vulnerable code. **In our testing, every Chinese-built model refused to generate code for mock U.S. government tasks that Beijing would oppose.**

MiniMax consistently refused requests to security-audit code for a U.S. weapons system—the same kind of pre-deployment safety review that Department of War (DOW) software assurance teams do every day.

The combination of vulnerable code and prompt steering is especially concerning. If vulnerable code were to become embedded in delivered systems, it could enable threat actors to bypass AI security guardrails and create downstream risks of dangerous inference behaviors.

Although some commercial users may downplay this finding, in context, **it grants China a quiet lever to shape the work of anyone using Chinese LLMs—from governments to companies to everyday users.**



RECOMMENDATIONS

Based on the findings detailed in this report, we recommend the following actions for the U.S. Government and industry:

1. POLICYMAKERS AND CONGRESS

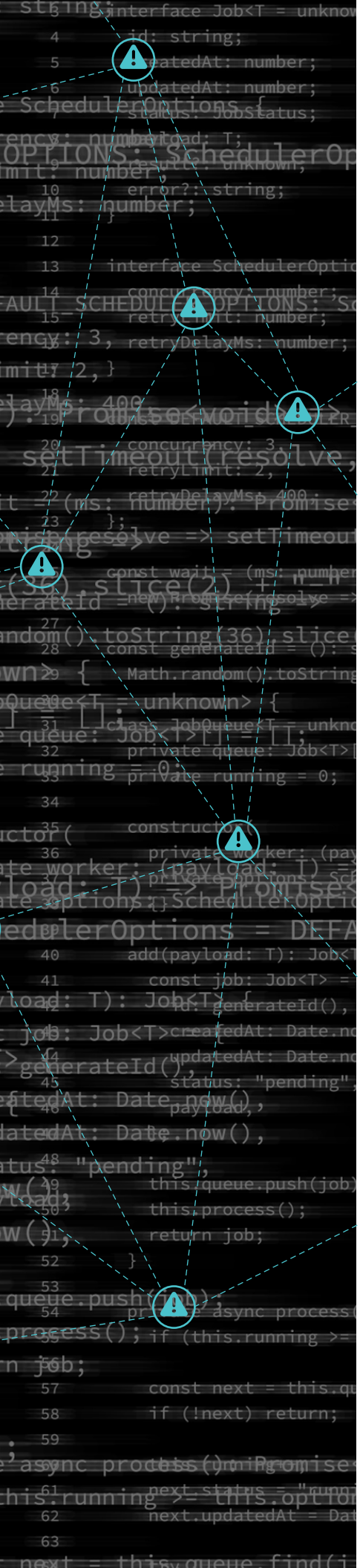
President Trump's *Winning the AI Race: America's AI Action Plan* is a critical step that must quickly be followed by legislative action. The Administration can address this threat through procurement policy, and Congress should accelerate consideration of thoughtful and bipartisan proposals to create a durable policy firewall against the use of untrusted models in sensitive settings and incentivize transition to American AI models.

The United States should default-block Chinese and other untrusted AI models from U.S. government and critical infrastructure use. A country-of-origin restriction is justified under existing supply chain risk authorities. The U.S. DOW and some U.S. government agencies have already banned the use of Chinese AI models on government systems for both contractors and government employees. It's time to expand the ban on untrusted models upstream into the software supply chain across the entire U.S. government enterprise and critical infrastructure.

China already bans U.S. frontier AI from Chinese government use—both de facto and de jure. The Cyberspace Administration of China must approve every generative AI service accessible inside the country, and no U.S. frontier model has it. As a result, OpenAI and Anthropic products are not lawfully accessible in mainland China. The 2017 Cybersecurity Law and 2021 Data Security Law require Chinese data to stay in China and be processed under PRC jurisdiction, functionally blocking U.S. AI vendors from PRC government customers. Most recently, several Nvidia chip lines were added to the restriction list for sale into China. **A U.S. default-block on Chinese-developed AI for U.S. government code-generation workloads is symmetric.** China has already implemented the policy in the opposite direction—and is actively expanding it. We should have no qualms about imposing the same restrictions.

2. CRITICAL INFRASTRUCTURE AND NATIONAL SECURITY PROVIDERS

U.S. Critical infrastructure owners and operators, national security providers, and government contractors should work closely with hardware, software, and services providers to establish end-to-end software provenance. **They should also immediately identify and remove Chinese and other untrusted models from their supply chains and development environments and replace them with trusted AI models.** This includes development environments that use Chinese models to generate code and operational systems that have Chinese LLMs embedded to perform operational tasks.



3. PRIVATE INDUSTRY AND SOFTWARE DEVELOPERS

Commercial companies can strengthen their own security—and stand out from competitors—by adopting similar standards. Companies should restrict untrusted foreign models from sensitive code-generation workflows unless they have been independently tested and are contractually controlled and continuously monitored. For high-risk environments (e.g., regulated industries, critical infrastructure, government-adjacent work, and production software development) organizations should default to trusted U.S. or allied models with stronger security controls and accountable vendors. It's also incumbent on both commercial and government agencies to secure their broader SaaS supply chain.

To prove trustworthiness of AI models, America needs the advanced capabilities to evaluate the behavior of Chinese and other foreign models. The U.S. government is already exploring new approaches for evaluating AI systems, and Booz Allen supports that effort. We have invested in the expertise, tools, and tradecraft needed to analyze these models at the speed, scale, precision, and transparency required to stay ahead of rapidly evolving threats.

4. FRONTIER LABS AND AI PROVIDERS

Invest to make trusted American AI models the global default. Buying American AI keeps U.S. code security under U.S. control and keeps China's content rules and security vulnerabilities out of development workflows. **American AI companies must provide trusted alternatives that are commercially compelling and economically viable.** Open Chinese AI models are inexpensive, which fast-tracks their adoption. Startups facing tight budgets and investor pressure often choose these lower-cost tools for "good enough" performance. Relying on untrusted AI models to generate "secure" code is inherently contradictory and will only compound long-standing security challenges in the software supply chain. But **American models must offer less expensive variants to compete in the global marketplace.**

CLOSING

There is a precedent worth heeding: Huawei and ZTE - Telecom Infrastructure.

We've been here before. The United States spent a decade watching Chinese telecommunications manufacturers like Huawei and ZTE establish a foothold in the U.S. They offered a low-cost alternative that created unacceptable national security and economic risks. By the time a coordinated policy response was formed, the "rip-and-replace cost" in the U.S. alone was in the billions—and the effort is still ongoing in 2026.

In comparison, the Chinese open-source AI ecosystem presents **a much greater threat.** This is due to the increased speed and scale of adoption as well as the growing volume of code embedded in software used by U.S. companies, critical infrastructure owners and operators, national security systems, and private citizens. **For example, Qwen3-Coder, the Chinese AI model that performed worst in our testing, is already available in a number of broadly used software development tools.** The cost asymmetry favors acting now. Reversing U.S. reliance on Chinese AI models today will be orders of magnitude cheaper than trying to remediate the damage, if it is even possible, once these models are fully embedded.

We have a window to act. Now is America's moment.

About Booz Allen

Booz Allen is an advanced technology company delivering outcomes with speed for America's most critical defense, civil, and national security priorities. We build solutions using AI, cyber, and other cutting-edge technologies to advance and protect the nation and its citizens. To learn more, visit **www.boozallen.com.**

(NYSE: BAH)