SENTINEL BRIEF

ASP | AI IMPERATIVE 2030

DECEMBER 2025

The AI Threat Transparency Gap:
Risk Reporting Practices in the U.S. and China

BLUF

Industry-led voluntary reporting of novel AI-assisted risks, threats, and harms is critical to the global development of ethical technologies and norms. A review of thirty AI companies in the U.S. and China finds that American firms maintain far more transparent, accountable, and robust threat reporting practices than their Chinese counterparts. At time of publication, 60% of evaluated U.S. firms and no Chinese firms publish publicly available, detailed information on novel risks involving their products, services, or consumers.

BACKGROUND

Artificial intelligence's revolutionary potential and rapid pace of advancement often precede governments' ability to enact timely policy interventions. To empower consumers and policymakers to make informed decisions as AI development accelerates, responsible AI firms publish voluntary reports on emerging and unprecedented misuse of their products and services. Transparent, proactive threat reporting mitigates risk to consumers while benefitting the global AI ecosystem by setting norms and guardrails for ethical development. This report finds that U.S. AI firms far surpass Chinese firms in their voluntary reporting and transparency practices, underscoring how Chinese firms' ideological perspectives are shaped by state capitalism and civil-military fusion.

White House National Cyber Director Sean Cairncross speaks during the 2025 Billington Cybersecurity Summit in Washington, D.C.

Image credit: National Security Authority Slovakia

1.1 The Imperative for AI Threat Reporting

Failing to mitigate known or foreseeable risks from emerging technologies can result in societal, political, or economic consequences that slow innovation, harm consumers, and jeapordize national security. Documenting cases of AI misuse that cross the barrier from theoretical to actual is beneficial for the global AI ecosystem in three primary ways.

Consumers accept some risk when they adopt emerging technologies, invest in AI companies, or share personal data that becomes part of a company's data corpus. Transparency about novel and emerging risks enables informed consumer decision-making and builds public confidence in AI.

AI developers around the world benefit from an information-sharing ecosystem that proactively identifies and mitigates risk. Failing to percieve or mitigate harms from emerging technologies increases regulatory, legal, and economic liability, increasing costs and slowing the pace of innovation.

Policymakers, law enforcement officials, and oversight institutions require timely, accurate threat data to craft AI strategies and policies that safeguard national security and reduce consumer risk without constraining innovation. Detailed, multi-source reporting anchors these efforts around documented threat patterns rather than hypotheticals.

1.2 Metrics for Responsible AI Incident Tracking

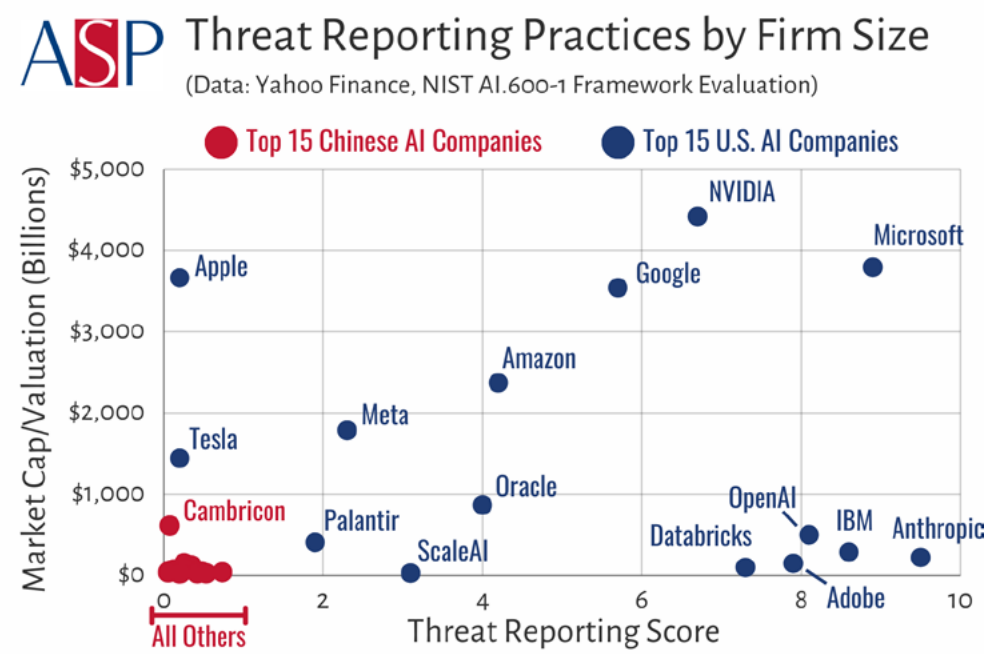
Timeliness: Are users and governments alerted to novel AI threats at sufficiently advanced timelines to help mitigate cyber-attacks or other malign impacts? Are third-party vendors given adequate warning to enact monitoring or mitigation strategies for future incidents?

Detail: Are threat reports comprehensive enough to provide actionable insights? Could a fellow developer understand the situation with enough detail to avoid or mitigate similar threats?

Transparency: Are incident reports easily accessible by governments, AI researchers, impacted individuals, and the public? Is there a mechanism for consumers or policymakers to reach out to the company for additional information?

Accountability: Did the disclosure contain information about company efforts to identify or mitigate the threat? If the incident involved human error, a lapse in protocol, or a cyber vulnerability, has the relevant action been taken to prevent similar incidents from occuring in the future?

President Donald Trump delivers remarks at the White House AI Summit at the Andrew W. Mellon Auditorium in Washington, DC in July 2025. (Photo by Joyce N. Boghosian/White House)



Methodolody

ASP researchers evaluated the fifteen largest firms offering AI products and services by either market capitalization or recent valuation in the United States and China, then graded each firm's AI threat reporting practices on a ten-point scale in 1) timeliness, 2) detail, 3) transparency, 4) accountability, and 5) stated internal policy as defined in the NIST AI.600-1 framework. Researchers then used a weighted mean rank formula to determine a final score.

1.3 American AI Ecosystem Research Findings

According to NIST AI.600-1 guidelines, greater awareness and industry standardization of AI threat reporting practices would improve risk management across the AI ecosystem and empower AI actors to better respond to AI incidents. Currently, neither the United States of America nor the People's Republic of China maintain federal policies requiring companies to publish public threat reports. However, the European Union and states like California and Colorado require disclosure of AI incidents over a specific magnitude.

Even in the absense of federal requirements, 60% of leading U.S. AI firms voluntarily publish either regular or ad-hoc public reports on AI security incidents, often in sophisticated detail and with actionable insights for policymakers, users, and fellow developers. 80% solicit information on novel vulnerabilities from users, with 30% offering monetary rewards for doing so. Open-source projects like the AI Incident Database allow researchers to search and classify security reports, furthering discourse on AI safety and ethics.

1.4 Chinese AI Ecosystem Research Findings

Out of the fifteen Chinese companies evaluated by ASP, none published public information that could be construed as AI threat or incident reports by any definition. Threat reporting scores above zero in the dataset were based on occasional public aknowledgement of security errors in media interviews (2 firms), "security reports" that did not evaluate AI risks or threats (1 firm), publication of AI safety research in academic journals (4 firms), and blog posts or web pages detailing firms' internal security policies (5 firms). References to AI risk (应用安全风险) and threat reporting (威胁报告) most often described potential risks of AI "violating core socialist values."

This lack of transparency by Chinese firms underscores a radically different industrial ecosystem shaped by state capitalism and civil-military fusion, in which authoritarian institutions both fund AI model development and provide monetary incentives for bad actors to deploy AI for cyber intrusions. Chinese regulations such as the "Provisions on the Management of Security Vulnerabilities" effectively bar companies and researchers from publicly disclosing security vulnerabilities, while severe fines, business suspensions, and executive liability regimes for security lapses drives a culture of secrecy. These practices compromise AI security and safety.