May 13, 2024

New AI Security Bill Introduced to Improve Tracking and Processing of AI Security and Safety Incidents

Client Bulletins

Authors: **Megan C. Parker**, **Luke Schaetzel**

As the development and use of AI continues to grow, the potential for security and safety incidents harming organizations and the public increases. Updated reporting and tracking processes for AI security and safety incidents are necessary to keep up with the risks associated with this ever-evolving technology.

Senators Mark R. Warner (D-VA) and Thom Tillis (R-NC) recently introduced the bipartisan Secure Artificial Intelligence Act of 2024 (the "**Bill**") aimed at improving the tracking and processing of security and safety incidents and risks associated with artificial intelligence ("**AI**").

The Bill is an early yet important attempt to improve information sharing between the federal government and private companies by updating cybersecurity reporting systems to better incorporate AI systems and creating a voluntary database for AI-related cybersecurity incidents.

Like President Biden's "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence"—issued in October of last year—the Bill also calls on different federal agencies to update current, or create new, standards that can be leveraged to set up industry standards in the burgeoning AI arena.

NIST and CISA Requirements

If passed, the Bill would require the National Institute of Standards and Technology ("**NIST**") to update the National Vulnerability Database ("**NVD**") and the Cybersecurity and Infrastructure Security Agency ("**CISA**") to update the Common Vulnerabilities and Exposure (**"CVE**") program or develop a new process to track voluntary reports of AI security vulnerabilities.

AI security vulnerabilities is defined by the Bill as "a weakness in an [AI] system that could be exploited by a third party to subvert, without authorization, the confidentiality, integrity, or availability of an [AI] system," through techniques such as data poisoning, evasion attacks, privacy-based attacks, and abuse attacks.

Further, within one year of the Bill's enactment, NIST and CISA would be required to establish a comprehensive voluntary database to publicly track AI security and safety incidents, including "near miss" events.

An AI security incident means "an event that increases (A) the risk that operation of [AI] system occurs in a way that enables the extraction of information about the behavior or characteristics of an [AI] system by a third party; or (B) the ability of a third party to manipulate an [AI] system in order to subvert the confidentiality, integrity, or availability of an [AI] system or adjacent system."

On the other hand, an AI safety incident refers to "an event that increases the risk that operation of an [AI] system will (A) result in physical or psychological harm; or (B) lead to a state in which human life, health, property, or the environment is endangered."

Importantly, neither definition requires that actual harm or actual unauthorized activity occur; only that the risk of such events ***increases***.

Once a voluntary report of an AI security and safety incident is received, NIST would need to determine whether the described incident is a "material" AI security or safety incident for inclusion in the database.

The Bill directs NIST to prioritize inclusion of the following three incidents: (1) an incident involving an AI system used in critical infrastructure or safety-critical systems; (2) an incident that would result in a high-severity or catastrophic impact on U.S. citizens or economy; and (3) an incident involving an AI system widely used in commercial or public sector contexts.

Lastly, CISA, NIST, and the National Security Agency ("**NSA**") would be required to convene a multi-stakeholder process encouraging the development and adoption of best practices addressing supply chain risks associated with AI model training and maintenance.

The topics would include, for example, (1) data collection, cleaning, and labeling; (2) inadequate documentation of AI training and the data that goes into such training; (3) the use of large, open-source data sets; (4) human feedback systems used to refine and further train AI; and (5) the use of proprietary datasets containing sensitive data or other personal data.

Establishment of an AI Security Center

Next, the Bill would require the NSA to establish an AI Security Center within its Cybersecurity Collaboration Center. According to the Bill, the AI Security Center would be tasked with:

- Making a research test-bed available to private sector and academic researchers to engage in AI security research;
- Developing guidance to prevent or mitigate counter-AI techniques, which include "techniques or procedures to extract information about the behavior or characteristics of an [AI] system, to let learn how to manipulate an [AI] system in order to subvert the confidentiality, integrity, or availability of an [AI] system or adjacent system";
- Promoting secure AI adoption practices for managers of national security systems (as defined in 44 U.S.C. § 3552) and elements of the defense industrial base;
- Coordinating with NIST's AI Safety Institute; and
- Other functions the NSA Director considers appropriate.

Conclusion

President Biden's AI Executive Order highlighted the important of AI safety in directing NIST to establish "red-teaming" guidelines—intentional attempts to get AI models to respond to prompts they're not supposed to for safety and security testing—and requiring AI developers to submit safety reports.

Although it still needs to make its way through a committee before consideration by the Senate, the Bill recognizes the need to safeguard against cybersecurity risks involving AI and encourages collaboration and innovation among the federal government and private sector entities.

Continue to follow Benesch's AI Commission as we address the evolving regulatory landscape of AI, impacts of new regulations and legislations, and steps toward compliance. Stay tuned!

Megan C. Parker at mparker@beneschlaw.com or 216.363.4416.

Luke Schaetzel at lschaetzel@beneschlaw.com or 312.212.4977.

Related Practices

Data Privacy & Cybersecurity

Related Industries

Artificial Intelligence (AI)

Related Professionals

Megan C. Parker
Associate
Intellectual Property
T. 216.363.4416
mparker@beneschlaw.com

Luke Schaetzel
Managing Associate
Intellectual Property
T. 312.212.4977
lschaetzel@beneschlaw.com