118TH CONGRESS
2D SESSION

H. R. 9737

To improve the tracking and processing of security and safety incidents and risks associated with artificial intelligence, and for other purposes.

IN THE HOUSE OF REPRESENTATIVES

SEPTEMBER 20, 2024

Ms. ROSS (for herself and Mr. BEYER) introduced the following bill; which was referred to the Committee on Science, Space, and Technology, and in addition to the Committees on Homeland Security, Intelligence (Permanent Select), and Education and the Workforce, for a period to be subsequently determined by the Speaker, in each case for consideration of such provisions as fall within the jurisdiction of the committee concerned

A BILL

To improve the tracking and processing of security and safety incidents and risks associated with artificial intelligence, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE.

This Act may be cited as the ‘‘Secure Artificial Intelligence Act of 2024’’ or the ‘‘Secure A.I. Act of 2024’’.

SEC. 2. DEFINITIONS.

In this Act:

(1) ARTIFICIAL INTELLIGENCE SAFETY INCIDENT.—The term ''artificial intelligence safety incident'' means an event that increases the risk that operation of an artificial intelligence system will—

(A) result in physical or psychological harm; or

(B) lead to a state in which human life, health, or property is endangered.

(2) ARTIFICIAL INTELLIGENCE SECURITY INCIDENT.—The term ''artificial intelligence security incident'' means an event that increases—

(A) the risk that operation of an artificial intelligence system occurs in a way that enables the extraction of information about the behavior or characteristics of an artificial intelligence system by a third party; or

(B) the ability of a third party to manipulate an artificial intelligence system in order to subvert the confidentiality, integrity, or availability of an artificial intelligence system or adjacent system.

(3) ARTIFICIAL INTELLIGENCE SECURITY VULNERABILITY.—The term ''artificial intelligence security vulnerability'' means a weakness in an artificial intelligence system that could be exploited by a third

party to subvert, without authorization, the confidentiality, integrity, or availability of an artificial intelligence system, including through techniques such as—

(A) data poisoning;

(B) evasion attacks;

(C) privacy-based attacks; and

(D) abuse attacks.

(4) COUNTER-ARTIFICIAL INTELLIGENCE.—The term ''counter-artificial intelligence'' means techniques or procedures to extract information about the behavior or characteristics of an artificial intelligence system, or to learn how to manipulate an artificial intelligence system, in order to subvert the confidentiality, integrity, or availability of an artificial intelligence system or adjacent system.

SEC. 3. VOLUNTARY TRACKING AND PROCESSING OF SECURITY AND SAFETY INCIDENTS AND RISKS ASSOCIATED WITH ARTIFICIAL INTELLIGENCE.

(a) PROCESSES AND PROCEDURES FOR VULNERABILITY MANAGEMENT.—Not later than 180 days after the date of the enactment of this Act, the Director of the National Institute of Standards and Technology shall—

(1) initiate a process to update processes and procedures associated with the National Vulner-

ability Database of the Institute to ensure that the database and associated vulnerability management processes incorporate artificial intelligence security vulnerabilities to the greatest extent practicable; and

(2) identify any characteristics of artificial intelligence security vulnerabilities that make utilization of the National Vulnerability Database inappropriate for their management and develop processes and procedures for vulnerability management for those vulnerabilities.

(b) VOLUNTARY TRACKING OF ARTIFICIAL INTELLIGENCE SECURITY AND ARTIFICIAL INTELLIGENCE SAFETY INCIDENTS.—

(1) VOLUNTARY DATABASE REQUIRED.—Not later than 1 year after the date of the enactment of this Act, the Director of the Institute, in coordination with the Director of the Cybersecurity and Infrastructure Security Agency, shall—

(A) develop and establish a comprehensive, voluntary database to publicly track artificial intelligence security and artificial intelligence safety incidents; and

(B) in establishing the database under subparagraph (A)—

(i) establish mechanisms by which private sector entities, public sector organizations, civil society groups, and academic researchers, (including current and former employees and contractors of such entities, organizations, and groups), may voluntarily share information with the Institute on confirmed or suspected artificial intelligence security or artificial intelligence safety incidents, in a manner that preserves confidentiality of any affected party and of the reporting party;

(ii) leverage, to the greatest extent possible, standardized disclosure and incident description formats;

(iii) develop processes to associate reports pertaining to the same incident with a single incident identifier;

(iv) establish classification, information retrieval, and reporting mechanisms that sufficiently differentiate between artificial intelligence security incidents and artificial intelligence safety incidents; and

(v) create appropriate taxonomies to classify incidents based on relevant characteristics, impact, or other relevant criteria.

(2) IDENTIFICATION AND TREATMENT OF MATERIAL ARTIFICIAL INTELLIGENCE SECURITY OR ARTIFICIAL INTELLIGENCE SAFETY RISKS.—

(A) IN GENERAL.—Upon receipt of relevant information on an artificial intelligence security or artificial intelligence safety incident, the Director of the Institute shall determine whether the described incident presents a material artificial intelligence security or artificial intelligence safety risk sufficient for inclusion in the database developed and established under paragraph (1).

(B) PRIORITIES.—In evaluating a reported incident pursuant to paragraph (1), the Director shall prioritize inclusion in the database cases in which a described incident—

(i) describes an artificial intelligence system used in critical infrastructure or safety-critical systems;

(ii) would result in a high-severity or catastrophic impact to the people or economy of the United States; or

(iii) includes an artificial intelligence system widely used in commercial or public sector contexts.

(3) REPORTS AND ANONYMITY.—The Director shall populate the voluntary database developed and established under paragraph (1) with incidents based on public reports and information shared using the mechanism established pursuant to subparagraph (B)(i) of such paragraph, ensuring that any incident description sufficiently anonymizes those affected, unless those who are affected have consented to their names being included in the database.

(4) PROTECTION OF REPORTING PARTIES.—

(A) PROHIBITION AGAINST RETALIATION.—No employer may, directly or indirectly, discharge, demote, suspend, threaten, blacklist, harass, or in any other manner discriminate against any current or former employee or contractor in the terms and conditions of employment or postemployment because of any act done by such employee or contractor—

(i) in reporting incidents in accordance with the mechanisms established in this section;

(ii) in reporting incidents to any Member of Congress or any committee of Congress; or

(iii) in initiating, testifying in, or assisting in any investigation or judicial or administrative action based upon or related to the incidents described in clause (i) or (ii).

In addition, no employer may require their employees or contractors to obtain prior consent from such employer to report incidents using the reporting mechanism established in this section or to any Member of Congress or any committee of Congress, or to obtain prior consent to participate in investigations, judicial, or administrative actions based upon or related to such incidents.

(B) ENFORCEMENT.—Any individual who alleges discharge or other discrimination, or is otherwise aggrieved by an employer or former employer, in violation of subparagraph (A), may seek relief by—

(i) filing a complaint with the Secretary of Labor in accordance with the requirements of this subsection; or

(ii) if the Secretary of Labor has not issued a final decision within 180 days of the filing of a complaint under clause (i), and there is no showing that such a delay is due to the bad faith of the claimant, bringing an action against the employer at law or in equity in the appropriate district court of the United States, which shall have jurisdiction over such an action without regard to the amount in controversy.

(C) CONFIDENTIALITY.—The Director, and any officer or employee of the National Institute of Standards and Technology or the Cybersecurity and Infrastructure Security Agency, shall not disclose any information, including information provided by a whistleblower to either such official, which could reasonably be expected to reveal the identity of a whistleblower, except in accordance with the provisions of section 552a of title 5, United States Code, unless and until required to be disclosed to a defendant or respondent in connection with a public proceeding instituted by the appropriate such official.

(D) RIGHTS RETAINED.—Nothing in this section shall be deemed to diminish the rights, privileges, or remedies of any whistleblower under any Federal or State law or under any collective bargaining agreement.

SEC. 4. UPDATING PROCESSES AND PROCEDURES RELATING TO COMMON VULNERABILITIES AND EXPOSURES PROGRAM AND EVALUATION OF CONSENSUS STANDARDS RELATING TO ARTIFICIAL INTELLIGENCE SECURITY VULNERABILITY REPORTING.

(a) DEFINITIONS.—In this section:

(1) COMMON VULNERABILITIES AND EXPOSURES PROGRAM.—The term ''Common Vulnerabilities and Exposures Program'' means the reference guide and classification system for publicly known information security vulnerabilities sponsored by the Cybersecurity and Infrastructure Security Agency.

(2) RELEVANT CONGRESSIONAL COMMITTEES.—The term ''relevant congressional committees'' means—

(A) the Committee on Homeland Security and Governmental Affairs, the Committee on Commerce, Science, and Transportation, the

Select Committee on Intelligence, and the Committee on the Judiciary of the Senate; and

(B) the Committee on Oversight and Accountability, the Committee on Energy and Commerce, the Permanent Select Committee on Intelligence, and the Committee on the Judiciary of the House of Representatives.

(b) IN GENERAL.—Not later than 180 days after the date of enactment of this Act, the Director of the Cybersecurity and Infrastructure Security Agency shall—

(1) initiate a process to update processes and procedures associated with the Common Vulnerabilities and Exposures Program to ensure that the program and associated processes identify and enumerate artificial intelligence security vulnerabilities to the greatest extent practicable; and

(2) identify any characteristic of artificial intelligence security vulnerabilities that make utilization of the Common Vulnerabilities and Exposures Program inappropriate for their management and develop processes and procedures for vulnerability identification and enumeration for those artificial intelligence security vulnerabilities.

(c) EVALUATION OF CONSENSUS STANDARDS.—

(1) In general.—Not later than 30 days after the date of enactment of this Act, the Director of the National Institute of Standards and Technology shall initiate a multi-stakeholder process to evaluate whether existing voluntary consensus standards for vulnerability reporting effectively accommodate artificial intelligence security vulnerabilities.

(2) Report.—

(A) Submission.—Not later than 180 days after the date on which the evaluation under paragraph (1) is carried out, the Director shall submit a report to the relevant congressional committees on the sufficiency of existing vulnerability reporting processes and standards to accommodate artificial intelligence security vulnerabilities.

(B) Post-report action.—If the Director concludes in the report submitted under subparagraph (A) that existing processes do not sufficiently accommodate reporting of artificial intelligence security vulnerabilities, the Director shall initiate a process, in consultation with the Director of the National Institute of Standards and Technology and the Director of the Office of Management and Budget, to update relevant vulnerability reporting processes, including the Department of Homeland Security Binding Operational Directive 20–01, or any subsequent directive.

(d) BEST PRACTICES.—Not later than 90 days after the date of enactment of this Act, the Director of the Cybersecurity and Infrastructure Security Agency shall, in collaboration with the Director of the National Security Agency and the Director of the National Institute of Standards and Technology and by leveraging efforts of the Information Communications Technology Supply Chain Risk Management Task Force to the greatest extent practicable, convene a multi-stakeholder process to encourage the development and adoption of best practices relating to addressing supply chain risks associated with training and maintaining artificial intelligence models, which shall ensure consideration of supply chain risks associated with—

(1) data collection, cleaning, and labeling, particularly the supply chain risks of reliance on remote workforce and foreign labor for such tasks;

(2) inadequate documentation of training data and test data storage, as well as limited provenance of training data;

(3) human feedback systems used to refine artificial intelligence systems, particularly the supply chain risks of reliance on remote workforce and foreign labor for such tasks;

(4) the use of large-scale, open-source datasets, particularly the supply chain risks to repositories that host such datasets for use by public and private sector developers in the United States; and

(5) the use of proprietary datasets containing sensitive or personally identifiable information.

(e) RULE OF CONSTRUCTION.—To the extent practicable, the Director shall examine the reporting requirements pursuant to division Y of the Cyber Incident Reporting for Critical Infrastructure Act of 2022 (Public Law 117–103) and the amendments made by that division and ensure that the requirements under this section are not duplicative of requirements set forth in that division and the amendments made by that division.

SEC. 5. ESTABLISHMENT OF ARTIFICIAL INTELLIGENCE SECURITY CENTER.

(a) ESTABLISHMENT.—Not later than 90 days after the date of the enactment of this Act, the Director of the National Security Agency shall establish an Artificial Intelligence Security Center within the Cybersecurity Collaboration Center of the National Security Agency.

(b) FUNCTIONS.—The functions of the Artificial Intelligence Security Center shall be as follows:

(1) Making available a research test-bed to private sector and academic researchers, on a subsidized basis, to engage in artificial intelligence security research, including through the secure provision of access in a secure environment to proprietary third-party models with the consent of the vendors of the models.

(2) Developing guidance to prevent or mitigate counter-artificial intelligence techniques.

(3) Promoting secure artificial intelligence adoption practices for managers of national security systems (as defined in section 3552 of title 44, United States Code) and elements of the defense industrial base.

(4) Coordinating with the Artificial Intelligence Safety Institute within the National Institute of Standards and Technology.

(5) Such other functions as the Director considers appropriate.

(c) TEST-BED REQUIREMENTS.—

(1) ACCESS AND TERMS OF USAGE.—

(A) RESEARCHER ACCESS.—The Director shall establish terms of usage governing researcher access to the test-bed made available under subsection (b)(1), with limitations on researcher publication only to the extent necessary to protect classified information or proprietary information concerning third-party models provided through the consent of model vendors.

(B) AVAILABILITY TO FEDERAL AGENCIES.—The Director shall ensure that the test-bed made available under subsection (b)(1) is also made available to other Federal agencies on a cost-recovery basis.

(2) USE OF CERTAIN INFRASTRUCTURE AND OTHER RESOURCES.—In carrying out subsection (b)(1), the Director shall leverage, to the greatest extent practicable, infrastructure and other resources provided under section 5.2 of the Executive order dated October 30, 2023 (relating to safe, secure, and trustworthy development and use of artificial intelligence).

(d) ACCESS TO PROPRIETARY MODELS.—In carrying out this section, The Director shall establish such mechanisms as the Director considers appropriate, including potential contractual incentives, to ensure the provision of access to proprietary models by qualified independent, third-party researchers, provided that commercial model vendors have voluntarily provided models and associated resources for such testing.

○