AMENDMENT IN THE NATURE OF A SUBSTITUTE TO H.R. 9333 OFFERED BY MS. ROSS OF NORTH CAROLINA

Strike all after the enacting clause and insert the following:

SECTION 1. SHORT TITLE.

This Act may be cited as the ''AI Flaw Reporting and Security Enhancement Act''.

SEC. 2. SUPPORTING VOLUNTARY REPORTING OF ARTIFICIAL INTELLIGENCE FLAWS.

(a) IN GENERAL.—The Director of the National Institute of Standards and Technology (NIST), in consultation with the Director of the Cybersecurity and Infrastructure Security Agency of the Department of Homeland Security, shall carry out a program to support the voluntary reporting, collection, and tracking of artificial intelligence flaws (in this section referred to as the ''program'').

(b) ACTIVITIES.—In carrying out the program, the Director of the NIST shall seek to convene appropriate representatives of industry, academia, nonprofit organizations, standards development organizations, civil society groups, and appropriate Federal departments and agencies to carry out the following:

(1) Establish common definitions and characterizations for relevant aspects relating to artificial intelligence flaws, including consideration of the following:

(A) Definitions of the following terms, as such terms relate to artificial intelligence:

(i) Vulnerabilities.

(ii) Failure modes.

(iii) Accidents.

(iv) Failures.

(v) Hazards.

(vi) Catastrophes.

(vii) Misuse.

(viii) Incidents.

(ix) Adverse events.

(B) Taxonomies to classify such artificial intelligence flaws based on relevant characteristics, impacts, or other appropriate criteria to enable the management and prioritization of such flaws, including the following:

(i) Artificial intelligence security-related flaws.

(ii) Artificial intelligence safety-related flaws.

(iii) Criteria for reporting flaws referred to in clauses (i) and (ii), or associated aspects relating to such flaws under subparagraph (A), to be considered for inclusion in the national database described in subsection (c)(1).

(2) Support the development of technical standards and guidance related to artificial intelligence flaws and processes for managing such flaws.

(3) Support the development of methods, which may include measures of severity or risk associated with artificial intelligence flaws, to enable prioritization of remediation activities of such flaws.

(4) Support the development of technical approaches which accelerate detection and monitoring of artificial intelligence flaws.

(5) Identify and provide guidelines, best practices, methodologies, procedures, and processes for reporting, collecting, and tracking artificial intelligence flaws across different sectors and use cases, including processes for reporting such flaws to be considered for inclusion in the national database described in subsection (c)(1).

(6) Support the development of standardized reporting and documentation mechanisms, including

automated mechanisms, that would help provide information, including public information, regarding artificial intelligence flaws.

(7) Support the development of norms for appropriate disclosure and reporting of artificial intelligence flaws, including when it is appropriate to publicly disclose such flaws.

(c) DEVELOPMENT OF INFRASTRUCTURE FOR THE MEASUREMENT AND MONITORING OF ARTIFICIAL INTELLIGENCE FLAWS.—

(1) IN GENERAL.—In carrying out the program, the Director of NIST shall, in consultation with representatives of industry, academia, nonprofit organizations, standards development organizations, civil society groups, appropriate public sector entities, and appropriate Federal departments and agencies, develop, or enter into cooperative agreements with one or more eligible entity designated by the Director to develop, infrastructure for the voluntary reporting, collection, and tracking of artificial intelligence flaws. Such infrastructure shall include a national database of artificial intelligence flaws or the modification of an existing national database to account for such flaws, as determined appropriate by the Director. Such database may be maintained by

NIST or one or more eligible entities designated by the Director.

(2) CONSIDERATIONS.—In carrying out this subsection, the Director shall consider the following:

(A) Technical standards and best practices regarding machine-readability.

(B) Interoperability of the infrastructure described in paragraph (1) with relevant existing standards, best practices, and systems.

(C) Future updates to the infrastructure described in paragraph (1) that may include additional types of information and taxonomies relevant to new stakeholders and coordination mechanisms.

(D) Relevant policies, procedures, and norms regarding dissemination of reported artificial intelligence flaws and public disclosures.

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

(1) IN GENERAL.—The Director of NIST, in consultation with the Director of the Cybersecurity and Infrastructure Security Agency of the Department of Homeland Security, shall seek to convene stakeholders to consider the development of a proc-

ess relating to the voluntary collection, reporting, and tracking of artificial intelligence security incidents and artificial intelligence safety incidents.

(2) ACTIVITIES.—In carrying out paragraph (1), the Director of NIST shall seek to convene appropriate representatives of industry, academia, nonprofit organizations, standards development organizations, civil society groups, Sector Risk Management Agencies, and appropriate Federal departments and agencies to carry out the following:

(A) Establish common definitions and characterizations for relevant aspects of artificial intelligence security incidents and artificial intelligence safety incidents, which may include the following:

(i) Classifications that sufficiently differentiate between the following:

(I) Artificial intelligence security incidents.

(II) Artificial intelligence safety incidents.

(ii) Taxonomies to classify incidents referred to in clause (i) based on relevant characteristics, impacts, or other appropriate criteria.

(iii) Criteria for incidents referred to in subsection (A) to be considered for inclusion in the reporting and tracking under subparagraphs (B) through (E).

(B) Assess the usefulness and cost-effectiveness of an effort to voluntarily track artificial intelligence security incidents and artificial intelligence safety incidents.

(C) Identify and provide guidelines, best practices, methodologies, procedures, and processes for tracking and reporting artificial intelligence security incidents and artificial intelligence safety incidents across different sectors and use cases.

(D) Support the development of standardized reporting and documentation mechanisms, including automated mechanisms, that would help provide information, including public information, regarding artificial intelligence security incidents and artificial intelligence safety incidents.

(E) Support the development of norms for reporting of artificial intelligence security incidents and artificial intelligence safety incidents,

taking into account when it is appropriate to publicly disclose such incidents.

(e) REPORT.—Not later than three years after the date of the enactment of this Act, the Director of NIST shall submit to Congress a report on the implementation of this section. Such report shall include the following:

(1) Findings from the multi-stakeholder activities under subsections (b) through (d).

(2) A description of the infrastructure developed pursuant to subsection (c), including a description of the national database referred to in such subsection.

(3) An assessment of and recommendations for establishing reporting and collection mechanisms by which industry, academia, nonprofit organizations, standards development organizations, civil society groups, and appropriate public sector entities may voluntarily share standardized information regarding artificial intelligence flaws and artificial intelligence incidents.

(f) DEFINITIONS.—In this section:

(1) ARTIFICIAL INTELLIGENCE.—The term ''artificial intelligence'' has the meaning given such term in section 5002 of the National Artificial Intelligence Initiative Act of 2020 (15 U.S.C. 9401).

(2) ARTIFICIAL INTELLIGENCE FLAW.—The term ''artificial intelligence flaw'' means a set of conditions or behaviors that allow the violation of an explicit or implicit policy related to the safety, security, or other undesirable effects from use of an artificial intelligence system, and which is not dependent on the presence of malicious intent or related harm.

(3) ARTIFICIAL INTELLIGENCE SYSTEM.—The term ''artificial intelligence system'' has the meaning given such term in section 7223 of the Advancing American AI Act (40 U.S.C. 11301 note; as enacted as part of title LXXII of division G of the James M. Inhofe National Defense Authorization Act for Fiscal Year 2023; Public Law 117–263).

(4) ELIGIBLE ENTITY.—The term ''eligible entity'' means an institution of higher education (as such term is defined in section 101(a) of the Higher Education Act of 1965 (20 U.S.C. 1001)), a research institution (as such term is defined in section 9 of the Small Business Act (15 U.S.C. 638(e)(8)), or consortia thereof.

(5) SECTOR RISK MANAGEMENT AGENCY.—The term ''Sector Risk Management Agency'' has the

meaning given such term in section 2200 of the Homeland Security Act of 2002 (6 U.S.C. 650).

☒