AUTHENTICATED
U.S. GOVERNMENT
INFORMATION
GPO



119TH CONGRESS
2D SESSION

# S. 5576

To establish the Artificial Intelligence Safety Board, and for other purposes.

---

IN THE SENATE OF THE UNITED STATES

SEPTEMBER 29, 2026

Mr. WARNER (for himself, Mr. SCHATZ, and Mr. KIM) introduced the following bill; which was read twice and referred to the Committee on Commerce, Science, and Transportation

---

# A BILL

To establish the Artificial Intelligence Safety Board, 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 ''Artificial Intelligence Risk Management and Security Act of 2026''.

**SEC. 2. DEFINITIONS.**

In this Act:

(1) ARTIFICIAL INTELLIGENCE; ARTIFICIAL INTELLIGENCE SYSTEM.—The terms ''artificial intelligence'' and ''artificial intelligence system'' has the

meaning given the term ''artificial intelligence'' in section 5002 of the National Artificial Intelligence Initiative Act of 2020 (15 U.S.C. 9401).

(2) ARTIFICIAL INTELLIGENCE AGENT.—The term ''artificial intelligence agent''—

(A) means an artificial intelligence system or process that, given an objective or instruction—

(i) determines the action or sequence of actions to be taken to accomplish that objective; and

(ii) is capable of executing such actions directly on information systems, data, or external services; and

(B) does not include a system or process that solely generates informational or advisory output for a human operator to act upon.

(3) ARTIFICIAL INTELLIGENCE CAPABILITIES OR RISK ASSESSMENT.—The term ''artificial intelligence capabilities or risk assessment'' means information regarding an assessment performed by a developer of an artificial intelligence system evaluating the capabilities of, or risks posed by, such system, including an assessment of the potential of the artificial intelligence system—

(A) to materially assist in the design, development, acquisition, or use of a chemical, biological, radiological, or nuclear weapon;

(B) to materially assist in the design, development, or production of munitions or other weapons;

(C) to materially assist in the unlawful manufacture, synthesis, or distribution of a controlled substance;

(D) to evade the control of, or act outside the intended instructions of, its developer or operator;

(E) to facilitate a cybersecurity threat, including through the discovery or exploitation of a security vulnerability;

(F) to be subject to the unauthorized exfiltration of the weights of the artificial intelligence system, or unauthorized, deliberate, malicious modification of such weights; or

(G) to be transformed, stolen, reverse-engineered, or otherwise manipulated by an unauthorized user or users acting outside of the terms of service governing access to the artificial intelligence system.

(4) ARTIFICIAL INTELLIGENCE FLAW.—The term ''artificial intelligence flaw'' means a recurring or reproducible characteristic, behavior, or failure mode of an artificial intelligence system that causes, or materially increases the risk of, an artificial intelligence safety incident absent any intentional act of a user, including a characteristic, behavior, or failure mode that may manifest across multiple systems or providers.

(5) ARTIFICIAL INTELLIGENCE SAFETY INCIDENT.—The term ''artificial intelligence safety incident'' means an event that materially increases the risk that operation of an artificial intelligence system leads to a state in which human life, health, property, or the environment is endangered.

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

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

(B) the ability to manipulate an artificial intelligence system in order to subvert the con-

fidentiality, integrity, or availability of the system or adjacent system.

(7) 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 the system, including through techniques such as—

(A) data poisoning;

(B) evasion attacks;

(C) privacy-based attacks;

(D) model theft or extraction attacks;

(E) attacks designed to circumvent or degrade the safety, alignment, or access control mechanisms of an artificial intelligence system; and

(F) adversarial machine learning attacks as described in National Institute of Standards and Technology Trustworthy and Responsible Artificial Intelligence 100–2e2025 (relating to Adversarial Machine Learning), or successor publication.

(8) BOARD.—The term ''Board'' means the Artificial Intelligence Safety Board established under section 3.

(9) CRITICAL INFRASTRUCTURE.—The term ''critical infrastructure'' has the meaning provided in section 1016(e) of the USA Patriot Act of 2001 (42 U.S.C. 5195c(e)).

(10) DEVELOPER.—The term ''developer'' means a developer of a frontier artificial intelligence model.

(11) FRONTIER ARTIFICIAL INTELLIGENCE MODEL.—The term ''frontier artificial intelligence model'' means an artificial intelligence model, or system combining multiple artificial intelligence models, that exhibits or could be modified to exhibit high levels of performance at tasks that pose a serious risk to national security, national economic security, or public health or safety.

(12) INSTITUTE.—The term ''Institute'' means the National Institute of Standards and Technology.

(13) SECRETARY.—The term ''Secretary'' means the Secretary of Commerce.

**SEC. 3. ARTIFICIAL INTELLIGENCE SAFETY BOARD.**

(a) ARTIFICIAL INTELLIGENCE SAFETY BOARD.—

(1) ESTABLISHMENT.—

(A) IN GENERAL.—Not later than 30 days after the date of the enactment of this Act, the Secretary shall establish within the Department of Commerce a board to address artificial intelligence risks.

(B) DESIGNATION.—The board established under subparagraph (A) shall be known as the ''Artificial Intelligence Safety Board'' (referred to in this Act as the ''Board'').

(C) PERMANENT STATUS.—The Board shall be a permanent advisory committee, and section 1013 of title 5, United States Code, shall not apply to the Board.

(2) MEMBERSHIP.—

(A) COMPOSITION.—The Board shall be composed of members who are appointed as follows:

(i) One member selected by the Director of the Institute.

(ii) One member selected by the Secretary.

(iii) One member selected by the Director of the Cybersecurity and Infrastructure Security Agency.

(iv) One member selected by the Director of the National Security Agency.

(v) One member selected by the Secretary of the Treasury.

(B) NONGOVERNMENTAL EXPERTS.—In addition to the members of the Board appointed under subparagraph (A), the Secretary shall appoint members who are not officers or employees of the Federal Government and who the Secretary selects from among individuals who—

(i) are leading technical experts not affiliated with a developer or provider of artificial intelligence systems;

(ii) are leading technical experts affiliated with developers or providers of artificial intelligence systems;

(iii) are individuals with expertise in developing evaluations to test artificial intelligence systems;

(iv) are individuals with expertise in consumer protection and antitrust jurisprudence; and

(v) have knowledge or expertise that the Secretary determines would further the purpose of the duties of the Board.

(C) INTERNATIONAL PARTICIPATION.—The Secretary, in coordination with and subject to the agreement of the Secretary of State, shall undertake negotiations with foreign partners, and enter into agreements as appropriate and subject to the disclosure requirements of section 112b of title 1, United States Code, to facilitate cooperative activities, regulatory reciprocity, and appropriate protection of intellectual property rights associated with addressing artificial intelligence safety and security risks, including—

(i) the establishment of joint testing environments;

(ii) jointly conducting competitive processes or competitive research programs to award cash prizes or other types of recognition for basic, advanced, and applied research, technology development, and prototype development that advance the security and safety of frontier artificial intelligence models;

(iii) promulgating joint advisories or technical guidance concerning security or

safety risks associated with frontier artificial intelligence systems;

(iv) facilitating of secure information sharing regarding artificial intelligence security or safety incidents; and

(v) standardizing evaluation protocols.

(3) TERMS AND VACANCIES.—

(A) TERMS.—Each member of the Board shall serve for a term of not longer than 3 years and may be reappointed for 1 successive term.

(B) VACANCY REPLACEMENT.—The members of the Board shall develop a vacancy replacement procedure that includes—

(i) for vacancies occurring due to the end of a member's term, a vote not later than 90 days before the last day of the member's term; and

(ii) for vacancies occurring under subparagraph (C) or for any other reason, the chair of the Board shall nominate a replacement from the same stakeholder category under paragraph (2), to the extent practicable, as the member creating the va-

cancy, subject to approval by a majority vote of the members of the Board.

(C) REMOVAL.—A member shall be removed from the Board if—

(i) the member fails to comply with the conflict of interest policy adopted pursuant to paragraph (5)(D); or

(ii) the member is denied the requisite security clearance under paragraph (4)(A)(iii).

(D) CHAIR.—The chair of the Board shall be selected by a majority vote among a quorum of the members appointed under paragraph (2) and shall serve not more than one 2-year term.

(4) MEMBER ACCESS TO CLASSIFIED INFORMATION.—

(A) ACCESS.—

(i) IN GENERAL.—Not later than 60 days after the date on which a member is first appointed to the Board and before the member is granted access to any classified information necessary to participate in a closed session pursuant to paragraph (5)(F), the Secretary shall determine, for the purposes of the Board, if the member

should be restricted from reviewing, discussing, or possessing classified information.

(ii) MANAGEMENT.—Access to classified information shall be managed in accordance with Executive Order 13526 (50 U.S.C. 3161 note; relating to classified national security information), or any subsequent corresponding executive order.

(iii) CLEARANCE REQUIREMENT.—

(I) IN GENERAL.—The Secretary shall sponsor each member of the Board for a security clearance at the Top Secret level with access to sensitive compartmented information, as appropriate, for the purposes of participating in carrying out the duties of the Board.

(II) DENIAL.—Any member who fails to obtain a security clearance, including by being denied by the appropriate authorities or if the Secretary determines the member should be restricted from reviewing, discussing, or possessing classified information, shall

be removed from the Board and a new member shall be appointed pursuant to the vacancy procedures under paragraph (3)(B).

(B) PROTECTION OF INFORMATION.—A member of the Board granted access to classified information shall sign and comply with the agreements to protect such information from unauthorized disclosure and to otherwise protect the classified information in accordance with the applicable requirements for the particular level of classification of the information.

(C) RULE OF CONSTRUCTION.—Nothing in this paragraph shall be construed to affect the existing security clearance of a member of the Board or the authority of a Federal agency to provide or deny a member of the Board access to any specific pieces of classified information.

(5) PROCEDURES.—

(A) DESIGNATED FEDERAL OFFICER.—The Secretary shall designate a Federal officer or employee to serve as the designated Federal officer of the Board, consistent with the requirements of chapter 10 of title 5, United

States Code (commonly known as the ''Federal Advisory Committee Act'').

(B) INITIAL MEETING AND BYLAWS.—Not later than 60 days after the date of the enactment of this Act, the Board shall convene and establish bylaws that—

(i) govern quorum and voting rules, including implementation of the decision-making majority voting requirement specified in paragraph (5)(C)(ii); and

(ii) set deliverable timelines and meeting schedules.

(C) OPERATING PROCEDURES.—Unless otherwise specified, the Board shall adopt written procedures governing its meetings, consistent with chapter 10 of title 5, United States Code, that include—

(i) requirements for public notice of meetings and the maintenance of records and minutes;

(ii) decisionmaking by majority vote of those present and voting;

(iii) authorization for the establishment of subgroups as necessary, subject to the approval of the chair of the Board; and

(iv) approval of the meeting agendas by the chair in consultation with the designated Federal officer under subparagraph (A) to ensure compliance with applicable laws.

(D) CONFLICT-OF-INTEREST POLICY.—

(i) IN GENERAL.—The Board shall adopt and enforce a written conflict of interest policy to ensure that members of the Board have a fiduciary responsibility to the Board, a duty to report conflicts of interest, including the appearance of a conflict of interest, and do not participate in deliberations or votes from which they personally or their employer would directly and materially benefit.

(ii) REQUIRED DISCLOSURES.—The policy under clause (i) shall require each member to publicly disclose all relevant financial and employment relationships, including relationships held by a spouse or immediate family member, and include recusal procedures in the event of a conflict.

(iii) RECORDS.—The designated Federal officer under subparagraph (A) shall maintain records of disclosures under clause (ii) of this subparagraph and make summaries of the disclosures available to the Secretary.

(E) THREAT INFORMATION ACCESS.—The Director of National Intelligence, in coordination with the heads of other appropriate Federal entities, shall ensure that the Board has access to relevant intelligence, including through closed or classified briefings or the provision of classified information, when appropriate.

(F) CLOSED SESSIONS.—Notwithstanding section 1009 of title 5, United States Code, the Board may hold closed or restricted-access sessions when the Secretary determines that the matters to be discussed involve any of the following:

(i) Classified information.

(ii) An artificial intelligence security incident.

(iii) An artificial intelligence security vulnerability.

(iv) An artificial intelligence flaw.

(v) Threat information.

(vi) Proprietary business information.

(vii) Other information exempt from public disclosure under section 552 of title 5, United States Code.

(6) DUTIES.—

(A) IN GENERAL.—The Board shall—

(i) develop a process to perform technical evaluations to determine what capabilities or combination of capabilities constitute high levels of performance at tasks that pose a serious risk to national security, national economic security, or public health or safety;

(ii) develop processes and metrics for evaluating risks posed by frontier artificial intelligence models, model variants, checkpoints, or other artificial intelligence models or versions used during development, training, testing, evaluation, or red-teaming, including versions that are not publicly available, whether or not they are intended for eventual public release, where such models, variants, checkpoints, or

other versions meet or could be modified to meet the risks described in section 2(3);

(iii) develop technical standards and conformity assessment methodologies, including—

(I) standardize formats and processes for publishing model or system cards with technical details of artificial intelligence systems;

(II) recommendations for maintaining cybersecurity measures for developers or providers of artificial intelligence systems across the lifecycle;

(III) processes and security controls for developers or providers of artificial intelligence systems to use to evaluate risks from employees or other personnel who have access to artificial intelligence systems developed or in development; and

(IV) recommendations on appropriate financial and other resourcing for developers or providers of artificial intelligence systems, including small- and medium-sized developers or pro-

viders, to robustly engage in safety and security research focused on the deployment of frontier artificial intelligence models; and

(iv) establish technical standards, security controls, and reference architectures for securing testing environments during evaluations of frontier artificial intelligence models, or models with known or reasonably foreseeable capabilities to discover and exploit software vulnerabilities without direct prompting by a human user, including—

(I) procedures for effective risk-modeling prior to commencing any evaluation;

(II) technical controls to ensure effective isolation and hardening of evaluation environments, including continuous re-evaluation of security configurations throughout evaluations;

(III) policies and procedures for continuous monitoring of model behavior, including monitoring in real-time, using pre-established or condi-

tional checkpoints, and through post-evaluation audits;

(IV) policies, procedures, and technical controls for continuous monitoring of evaluation environments, including automated identification of changes to security controls or configurations;

(V) policies, procedures, and technical controls for safeguarding identity and access management resources associated with the evaluation environment, or encompassing systems, from access by any model under evaluation; and

(VI) conditions, policies, procedures, and technical controls for prompt termination of any evaluation in which a model has operated beyond the policies, procedures, or technical controls described in clauses (I) through (V) or that otherwise poses an imminent risk to any individual or property outside the scope of the evaluation.

(B) PERIODIC REASSESSMENT OF TECHNICAL EVALUATIONS AND BEST PRACTICES.—Not less frequently than once every year, the Board shall—

(i) review each standard developed under subparagraph (A);

(ii) determine whether the standard should be modified or revoked; and

(iii) submit to the Secretary any proposed modification or revocation.

(7) SUPPORT STAFF.—The Director of the Institute, acting through the Center for Artificial Intelligence Standards and Innovation (or any successor office or entity), shall provide staff and other support necessary to assist the Board in carrying out its duties under this Act. Such staff shall work under the direction of the Board, shall exercise on behalf of the Board the access provided to the Board under subsection (f), and shall remain employees of the Institute.

(b) ADOPTION OF STANDARDS.—

(1) DEADLINE FOR SUBMISSION OF STANDARDS.—Not later than 90 days after the date of the establishment of the Board under subsection (a)(1), the Board shall develop and submit to the Secretary

the proposed standards required under subsection (a)(6).

(2) ADOPTION OF STANDARDS.—

(A) IN GENERAL.—Not later than 30 days after the date on which the Secretary receives a proposed standard under paragraph (1), the Secretary shall adopt the proposed standard or the modified standards by rule.

(B) MODIFICATION OF STANDARD.—The Secretary may modify a proposed standard received under paragraph (1) if the Secretary determines that the modification is necessary for the purpose of national security.

(C) PUBLICATION.—The Secretary shall publish in the Federal Register the reasons for any modification made under this paragraph consistent with laws and regulations governing the disclosure of classified information or material.

(c) MANDATORY COMPLIANCE.—Each developer shall comply with each standard adopted under subsection (b)(2) that is applicable to the developer.

(d) ENFORCEMENT.—

(1) IN GENERAL.—The Secretary shall enforce compliance with this section.

(2) CIVIL PENALTY.—A developer that violates subsection (c) shall be liable to the United States for a civil penalty of not more than $250,000 for each violation.

(3) CONTINUING VIOLATIONS.—Each day during which a violation under subsection (c) continues shall constitute a separate violation.

(4) CIVIL ACTION.—The Attorney General may, at the request of the Secretary, bring a civil action in an appropriate district court of the United States—

(A) to enjoin a violation of subsection (c); or

(B) to recover a civil penalty imposed under paragraph (2).

(e) SECURE RESEARCH-TEST-BEDS.—In developing a process to perform technical evaluations pursuant to section 6(A)(i), the Secretary may—

(1) seek to utilize secure computing environments, on a reimbursable basis, provided by Federal partners, including—

(A) the National Security Agency; and

(B) the National Laboratories of the Department of Energy; and

(2) make such secure computing environments available to private sector, Federal agencies, and qualified independent expert participants, on a cost-recovery basis, to engage in artificial intelligence security research, including through the secure provision of access in a secure environment for pre-deployment testing of any frontier artificial intelligence model prior to public release if security requirements necessitate hosting outside a commercial computing environment.

(f) REQUIREMENT THAT DEVELOPERS OF FRONTIER ARTIFICIAL INTELLIGENCE MODELS GIVE ACCESS TO BOARD BEFORE PUBLIC RELEASE.—Not later than 45 calendar days before a developer introduces into interstate or foreign commerce a frontier artificial intelligence model, the developer shall make available to the Board access to the frontier artificial intelligence model, including model's weights, configuration files, runtimes, or software libraries necessary to operate the frontier artificial intelligence model.

**SEC. 4. SAFETY PLAN REQUIREMENT.**

(a) IN GENERAL.—Each developer shall develop, publish, and follow a safety plan (referred to in this section as the ''Model Safety Plan'') for each artificial intelligence system or model that the developer—

(1) creates;

(2) substantially modifies; or

(3) uses in the training or evaluation of other artificial intelligence models.

(b) REQUIRED CRITERIA.—Each Model Safety Plan shall include—

(1) the artificial intelligence model or system to which it applies;

(2) an artificial intelligence capabilities or risk assessment prepared by the developer specifically for the artificial intelligence model or system associated with the Model Safety Plan;

(3) a list of the specific mitigation measures that the developer will undertake for each item described in the artificial intelligence capabilities or risk assessment prepared pursuant to paragraph (2), across product lifecycle; and

(4) the identity of the corporate officer responsible for the implementation of the Model Safety Plan.

(c) REQUIREMENT TO FILE.—Each Model Safety Plan shall be submitted to the Secretary in the form and manner determined by the Secretary.

**SEC. 5. DATABASE FOR ARTIFICIAL INTELLIGENCE SECURITY AND SAFETY INCIDENTS, FLAWS, AND RISKS.**

(a) TRACKING OF ARTIFICIAL INTELLIGENCE SECURITY AND ARTIFICIAL INTELLIGENCE SAFETY INCIDENTS AND ARTIFICIAL INTELLIGENCE FLAWS.—

(1) VOLUNTARY SUBMISSIONS.—Not later than 180 days after the date of the enactment of this Act, the Director of the Institute shall, in coordination with the Director of the Cybersecurity and Infrastructure Security Agency, establish mechanisms by which private sector entities, public sector organizations, civil society groups, and academic researchers may voluntarily share information with the Institute on confirmed or suspected artificial intelligence security or artificial intelligence safety incidents, or on confirmed or suspected artificial intelligence flaws, including flaws identified through testing, evaluation, red-teaming, or post-incident analysis in a manner that preserves confidentiality of any affected party. Such mechanisms shall—

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

(B) develop processes to associate reports pertaining to the same incident with a single in-

cident identifier, and to associate incidents or near misses that appear to arise from the same artificial intelligence flaw with a common flaw identifier;

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

(D) create appropriate taxonomies to classify incidents based on relevant characteristics, impact, or other relevant criteria, and to classify artificial intelligence flaws based on their characteristics, affected capabilities, likely consequences, and recurrence across models or systems.

(2) PUBLICLY ACCESSIBLE DATABASE.—

(A) ESTABLISHMENT OF DATABASE REQUIRED.—Not later than 1 year after the date of the enactment of this Act, the Director of the Institute shall, in coordination with the Director of the Cybersecurity and Infrastructure Security Agency, establish a publicly accessible database of artificial intelligence security inci-

dents and artificial intelligence safety incidents, together with a catalog of artificial intelligence flaws identified through reports, testing, evaluations, or investigations under this section.

(B) REVIEW AND POPULATION OF DATABASE.—Upon receipt of relevant information on an artificial intelligence security incident or artificial intelligence safety incident under paragraph (1), or on an artificial intelligence flaw under paragraph (1), the Director of the Institute shall review the information and determine whether the described incident or flaw constitutes an artificial intelligence security or artificial intelligence safety risk appropriate for inclusion in the database developed and established under subparagraph (A).

(C) IDENTIFICATION OF CAUSAL FACTORS AND ARTIFICIAL INTELLIGENCE FLAWS.—When making a determination under subparagraph (B), the Director of the Institute shall identify causal factors for the artificial intelligence security incident or the artificial intelligence safety incident and determine whether the incident reveals, is associated with, or provides evidence of an artificial intelligence flaw. If the Director of

the National Institute of Standards and Technology identifies such a flaw, the Director shall assign or associate the incident with a common flaw identifier and, to the extent practicable, identify other known incidents, near misses, models, or systems associated with the same flaw, including—

(i) the artificial intelligence system;

(ii) the deployment of the artificial intelligence systems; and

(iii) practices related to the operation of the artificial intelligence system, including misuse of the artificial intelligence system.

(3) MANDATORY SUBMISSIONS.—

(A) IN GENERAL.—The following entities shall report any confirmed artificial intelligence safety incident or artificial intelligence security incident within 30 days of confirmation of such incident and within 72 hours if such incident poses an imminent threat to national security, critical infrastructure, or public safety:

(i) Any developer or provider of a frontier artificial intelligence model.

(ii) Any operator of critical infrastructure that utilizes an artificial intelligence model in the context of managing industrial control systems or other operational technologies.

(B) APPLICABILITY.—For the purposes of subparagraph (A), the reporting requirement under this paragraph shall apply regardless of whether the frontier artificial intelligence model, or any model variant, checkpoint, or other version of such model involved in the incident, is or is intended to be made publicly available, and regardless of whether the incident occurs during development, training, testing, evaluation, red-teaming, deployment, or operation.

(4) PRIORITIES.—In evaluating information under paragraph (2) and determining under such subparagraph whether to include a report of an incident in the database required by paragraph (2)(A), the Director of the Institute shall prioritize inclusion in the database of cases in which a described incident—

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

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

(C) includes an artificial intelligence system widely used in commercial or public sector contexts in the United States; or

(D) constitutes a mandatory submission pursuant to subsection (a)(3).

(5) EXEMPTION FROM DISCLOSURE; REPORTS AND ANONYMITY.—

(A) ANONYMITY.—The Director of the Institute shall populate the database developed and established under paragraph (2)(A) with incidents and artificial intelligence flaws based on public reports and information shared using the mechanism established pursuant to paragraphs (1) and (3), ensuring that any incident description sufficiently anonymizes those affected, unless those who are affected have consented to their names being included in the database.

(B) EXEMPTION FROM DISCLOSURE.—Any information shared using the mechanism established pursuant to paragraphs (1) and (3)—

(i) shall be exempt from disclosure and withheld, unless an affected party consents to the inclusion of their names in the database as provided for under subparagraph (A), from the public, pursuant to section 552(b)(3)(B) of title 5, United States Code, and any other provision of United States law or law of any State, political subdivision or agency thereof, or Tribe requiring disclosure of information or records; and

(ii) shall not be deemed a waiver of any applicable privilege or protection, including trade secret protection.

(C) CONSULTATION REQUIRED.—Before publishing information regarding an artificial intelligence safety incident or an artificial intelligence security incident, the Director of the Institute shall consult with the developer or provider of the artificial intelligence system involved in an incident.

(b) MATERIAL RISK GUIDANCE.—Not later than 180 days after the date of the enactment of this Act, the Director of the Institute shall, in coordination with the Director of the Cybersecurity and Infrastructure Security Agency,

publish nonbinding guidance that provides illustrative criteria and examples for determining when an event ''materially increases'' a risk for purposes of an artificial intelligence safety incident or an artificial intelligence security incident.

**SEC. 6. AGENTIC ARTIFICIAL INTELLIGENCE TRUST AND VERIFICATION.**

(a) AGENTIC ARTIFICIAL INTELLIGENCE PROFILE.—

(1) IN GENERAL.—Not later than 18 months after the date of the enactment of this Act, the Director of the Institute shall develop, in collaboration with other public and private sector organizations as appropriate, a cross-sectoral profile (referred to in this section as the ''Agentic AI Profile'') of the Artificial Intelligence Risk Management Framework (NIST AI 100–1) or any successor framework (referred to in this section as the ''Framework'') for artificial intelligence agents.

(2) PURPOSE AND CONTENTS.—The Agentic AI Profile shall assist organizations in using the Framework to map, measure, manage, and govern risks associated with artificial intelligence agents. The Agentic AI Profile shall, at a minimum—

(A) define risks that are novel to or exacerbated by artificial intelligence agents;

(B) include a framework for classification of agent autonomy levels; and

(C) address cybersecurity risks specific to artificial intelligence agents, including risks related to agent identity, authentication, and authorization, and the relationship of such risks to the Cybersecurity Framework (NIST CSWP–29) or any successor framework.

(3) RELATIONSHIP TO THE FRAMEWORK.—The Agentic AI Profile shall be a companion resource to the Framework and shall not modify or supersede the Framework.

(b) COMMON TEMPLATE FOR ARTIFICIAL INTELLIGENCE AGENT ASSURANCE.—

(1) IN GENERAL.—Not later than 18 months after the date of the enactment of this Act, the Director of the Institute shall initiate, in collaboration with other public and private sector organizations and Federal agencies as appropriate, the development of a common template to document artificial intelligence agents and their evaluation against widely recognized standards and frameworks.

(2) ELEMENTS.—The template shall enable the consistent documentation of artificial intelligence agents by providing standardized terminology and

fields that relate to, at a minimum, the following elements:

(A) Identity and version.

(B) Intended use and evaluated scope.

(C) Ownership and authority boundaries.

(D) Access to data, systems, and tools.

(E) Evaluations against widely recognized standards and frameworks.

(F) The identity of any independent evaluator, where applicable.

(G) Known limitations and conditions of use.

(3) APPLICABILITY.—The Director shall design the template to be used across sectors, industries, and organizational contexts by entities of differing sizes and resources.

(c) COORDINATION.—The Director shall ensure that the template developed under this subsection is consistent with and not duplicative of other efforts by the Institute related to artificial intelligence agents.

○