

118TH CONGRESS  
2D SESSION

# S. 4230

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

---

IN THE SENATE OF THE UNITED STATES

MAY 1, 2024

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

---

## A BILL

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

1 *Be it enacted by the Senate and House of Representa-*  
2 *tives of the United States of America in Congress assembled,*

3 **SECTION 1. SHORT TITLE.**

4 This Act may be cited as the “Secure Artificial Intel-  
5 ligence Act of 2024” or the “Secure A.I. Act of 2024”.

6 **SEC. 2. DEFINITIONS.**

7 In this Act:

8 (1) **ARTIFICIAL INTELLIGENCE SAFETY INCI-**  
9 **DENT.**—The term “artificial intelligence safety inci-

1 dent” means an event that increases the risk that  
2 operation of an artificial intelligence system will—

3 (A) result in physical or psychological  
4 harm; or

5 (B) lead to a state in which human life,  
6 health, property, or the environment is endan-  
7 gered.

8 (2) ARTIFICIAL INTELLIGENCE SECURITY INCI-  
9 DENT.—The term “artificial intelligence security in-  
10 cident” means an event that increases—

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

16 (B) the ability of a third party to manipu-  
17 late an artificial intelligence system in order to  
18 subvert the confidentiality, integrity, or avail-  
19 ability of an artificial intelligence system or ad-  
20 jacent system.

21 (3) ARTIFICIAL INTELLIGENCE SECURITY VUL-  
22 NERABILITY.—The term “artificial intelligence secu-  
23 rity vulnerability” means a weakness in an artificial  
24 intelligence system that could be exploited by a third  
25 party to subvert, without authorization, the con-

1        confidentiality, integrity, or availability of an artificial  
2        intelligence system, including through techniques  
3        such as—

- 4                    (A) data poisoning;
- 5                    (B) evasion attacks;
- 6                    (C) privacy-based attacks; and
- 7                    (D) abuse attacks.

8                    (4) COUNTER-ARTIFICIAL INTELLIGENCE.—The  
9        term “counter-artificial intelligence” means tech-  
10       niques or procedures to extract information about  
11       the behavior or characteristics of an artificial intel-  
12       ligence system, or to learn how to manipulate an ar-  
13       tificial intelligence system, in order to subvert the  
14       confidentiality, integrity, or availability of an artifi-  
15       cial intelligence system or adjacent system.

16 **SEC. 3. VOLUNTARY TRACKING AND PROCESSING OF SECU-**  
17 **RITY AND SAFETY INCIDENTS AND RISKS AS-**  
18 **SOCIATED WITH ARTIFICIAL INTELLIGENCE.**

19        (a) PROCESSES AND PROCEDURES FOR VULNER-  
20       ABILITY MANAGEMENT.—Not later than 180 days after  
21       the date of the enactment of this Act, the Director of the  
22       National Institute of Standards and Technology shall—  
23                    (1) initiate a process to update processes and  
24       procedures associated with the National Vulner-  
25       ability Database of the Institute to ensure that the

1 database and associated vulnerability management  
2 processes incorporate artificial intelligence security  
3 vulnerabilities to the greatest extent practicable; and

4 (2) identify any characteristics of artificial in-  
5 telligence security vulnerabilities that make utiliza-  
6 tion of the National Vulnerability Database inappro-  
7 priate for their management and develop processes  
8 and procedures for vulnerability management for  
9 those vulnerabilities.

10 (b) VOLUNTARY TRACKING OF ARTIFICIAL INTEL-  
11 LIGENCE SECURITY AND ARTIFICIAL INTELLIGENCE  
12 SAFETY INCIDENTS.—

13 (1) VOLUNTARY DATABASE REQUIRED.—Not  
14 later than 1 year after the date of the enactment of  
15 this Act, the Director of the Institute, in coordina-  
16 tion with the Director of the Cybersecurity and In-  
17 frastructure Security Agency, shall—

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

22 (B) in establishing the database under sub-  
23 paragraph (A)—

24 (i) establish mechanisms by which pri-  
25 vate sector entities, public sector organiza-

1 tions, civil society groups, and academic re-  
2 searchers may voluntarily share informa-  
3 tion with the Institute on confirmed or  
4 suspected artificial intelligence security or  
5 artificial intelligence safety incidents, in a  
6 manner that preserves confidentiality of  
7 any affected party;

8 (ii) leverage, to the greatest extent  
9 possible, standardized disclosure and inci-  
10 dent description formats;

11 (iii) develop processes to associate re-  
12 ports pertaining to the same incident with  
13 a single incident identifier;

14 (iv) establish classification, informa-  
15 tion retrieval, and reporting mechanisms  
16 that sufficiently differentiate between arti-  
17 ficial intelligence security incidents and ar-  
18 tificial intelligence safety incidents; and

19 (v) create appropriate taxonomies to  
20 classify incidents based on relevant charac-  
21 teristics, impact, or other relevant criteria.

22 (2) IDENTIFICATION AND TREATMENT OF MA-  
23 TERIAL ARTIFICIAL INTELLIGENCE SECURITY OR AR-  
24 TIFICIAL INTELLIGENCE SAFETY RISKS.—

1 (A) IN GENERAL.—Upon receipt of rel-  
2 evant information on an artificial intelligence  
3 security or artificial intelligence safety incident,  
4 the Director of the Institute shall determine  
5 whether the described incident presents a mate-  
6 rial artificial intelligence security or artificial  
7 intelligence safety risk sufficient for inclusion in  
8 the database developed and established under  
9 paragraph (1).

10 (B) PRIORITIES.—In evaluating a reported  
11 incident pursuant to paragraph (1), the Direc-  
12 tor shall prioritize inclusion in the database  
13 cases in which a described incident—

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

17 (ii) would result in a high-severity or  
18 catastrophic impact to the people or econ-  
19 omy of the United States; or

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

23 (3) REPORTS AND ANONYMITY.—The Director  
24 shall populate the voluntary database developed and  
25 established under paragraph (1) with incidents

1 based on public reports and information shared  
 2 using the mechanism established pursuant to sub-  
 3 paragraph (B)(i) of such paragraph, ensuring that  
 4 any incident description sufficiently anonymizes  
 5 those affected, unless those who are affected have  
 6 consented to their names being included in the data-  
 7 base.

8 **SEC. 4. UPDATING PROCESSES AND PROCEDURES RELAT-**  
 9 **ING TO COMMON VULNERABILITIES AND EX-**  
 10 **POSURES PROGRAM AND EVALUATION OF**  
 11 **CONSENSUS STANDARDS RELATING TO ARTI-**  
 12 **FICIAL INTELLIGENCE SECURITY VULNER-**  
 13 **ABILITY REPORTING.**

14 (a) DEFINITIONS.—In this section:

15 (1) COMMON VULNERABILITIES AND EXPO-  
 16 SURES PROGRAM.—The term “Common  
 17 Vulnerabilities and Exposures Program” means the  
 18 reference guide and classification system for publicly  
 19 known information security vulnerabilities sponsored  
 20 by the Cybersecurity and Infrastructure Security  
 21 Agency.

22 (2) RELEVANT CONGRESSIONAL COMMIT-  
 23 TEES.—The term “relevant congressional commit-  
 24 tees” means—

1 (A) the Committee on Homeland Security  
2 and Governmental Affairs, the Committee on  
3 Commerce, Science, and Transportation, the  
4 Select Committee on Intelligence, and the Com-  
5 mittee on the Judiciary of the Senate; and

6 (B) the Committee on Oversight and Ac-  
7 countability, the Committee on Energy and  
8 Commerce, the Permanent Select Committee on  
9 Intelligence, and the Committee on the Judici-  
10 ary of the House of Representatives.

11 (b) IN GENERAL.—Not later than 180 days after the  
12 date of enactment of this Act, the Director of the Cyberse-  
13 curity and Infrastructure Security Agency shall—

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

20 (2) identify any characteristic of artificial intel-  
21 ligence security vulnerabilities that make utilization  
22 of the Common Vulnerabilities and Exposures Pro-  
23 gram inappropriate for their management and de-  
24 velop processes and procedures for vulnerability



1 identification and enumeration for those artificial in-  
2 telligence security vulnerabilities.

3 (c) EVALUATION OF CONSENSUS STANDARDS.—

4 (1) IN GENERAL.—Not later than 30 days after  
5 the date of enactment of this Act, the Director of  
6 the National Institute of Standards and Technology  
7 shall initiate a multi-stakeholder process to evaluate  
8 whether existing voluntary consensus standards for  
9 vulnerability reporting effectively accommodate arti-  
10 ficial intelligence security vulnerabilities.

11 (2) REPORT.—

12 (A) SUBMISSION.—Not later than 180  
13 days after the date on which the evaluation  
14 under paragraph (1) is carried out, the Director  
15 shall submit a report to the relevant congress-  
16 sional committees on the sufficiency of existing  
17 vulnerability reporting processes and standards  
18 to accommodate artificial intelligence security  
19 vulnerabilities.

20 (B) POST-REPORT ACTION.—If the Direc-  
21 tor concludes in the report submitted under  
22 subparagraph (A) that existing processes do not  
23 sufficiently accommodate reporting of artificial  
24 intelligence security vulnerabilities, the Director  
25 shall initiate a process, in consultation with the

1 Director of the National Institute of Standards  
2 and Technology and the Director of the Office  
3 of Management and Budget, to update relevant  
4 vulnerability reporting processes, including the  
5 Department of Homeland Security Binding  
6 Operational Directive 20–01, or any subsequent  
7 directive.

8 (d) BEST PRACTICES.—Not later than 90 days after  
9 the date of enactment of this Act, the Director of the Cy-  
10 bersecurity and Infrastructure Security Agency shall, in  
11 collaboration with the Director of the National Security  
12 Agency and the Director of the National Institute of  
13 Standards and Technology and by leveraging efforts of the  
14 Information Communications Technology Supply Chain  
15 Risk Management Task Force to the greatest extent prac-  
16 ticable, convene a multi-stakeholder process to encourage  
17 the development and adoption of best practices relating  
18 to addressing supply chain risks associated with training  
19 and maintaining artificial intelligence models, which shall  
20 ensure consideration of supply chain risks associated  
21 with—

22 (1) data collection, cleaning, and labeling, par-  
23 ticularly the supply chain risks of reliance on remote  
24 workforce and foreign labor for such tasks;

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

4           (3) human feedback systems used to refine arti-  
5           ficial intelligence systems, particularly the supply  
6           chain risks of reliance on remote workforce and for-  
7           eign labor for such tasks;

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

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

14          (e) **RULE OF CONSTRUCTION.**—To the extent prac-  
15          ticable, the Director shall examine the reporting require-  
16          ments pursuant to division Y of the Cyber Incident Re-  
17          porting for Critical Infrastructure Act of 2022 (Public  
18          Law 117–103) and the amendments made by that division  
19          and ensure that the requirements under this section are  
20          not duplicative of requirements set forth in that division  
21          and the amendments made by that division.

22          **SEC. 5. ESTABLISHMENT OF ARTIFICIAL INTELLIGENCE SE-**  
23          **CURITY CENTER.**

24          (a) **ESTABLISHMENT.**—Not later than 90 days after  
25          the date of the enactment of this Act, the Director of the

1 National Security Agency shall establish an Artificial In-  
2 telligence Security Center within the Cybersecurity Col-  
3 laboration Center of the National Security Agency.

4 (b) FUNCTIONS.—The functions of the Artificial In-  
5 telligence Security Center shall be as follows:

6 (1) Making available a research test-bed to pri-  
7 vate sector and academic researchers, on a sub-  
8 sidized basis, to engage in artificial intelligence secu-  
9 rity research, including through the secure provision  
10 of access in a secure environment to proprietary  
11 third-party models with the consent of the vendors  
12 of the models.

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

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

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

23 (5) Such other functions as the Director con-  
24 siders appropriate.

25 (c) TEST-BED REQUIREMENTS.—

1 (1) ACCESS AND TERMS OF USAGE.—

2 (A) RESEARCHER ACCESS.—The Director  
3 shall establish terms of usage governing re-  
4 searcher access to the test-bed made available  
5 under subsection (b)(1), with limitations on re-  
6 searcher publication only to the extent nec-  
7 essary to protect classified information or pro-  
8 prietary information concerning third-party  
9 models provided through the consent of model  
10 vendors.

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

16 (2) USE OF CERTAIN INFRASTRUCTURE AND  
17 OTHER RESOURCES.—In carrying out subsection  
18 (b)(1), the Director shall leverage, to the greatest  
19 extent practicable, infrastructure and other re-  
20 sources provided under section 5.2 of the Executive  
21 Order dated October 30, 2023 (relating to safe, se-  
22 cure, and trustworthy development and use of artifi-  
23 cial intelligence).

24 (d) ACCESS TO PROPRIETARY MODELS.—In carrying  
25 out this section, The Director shall establish such mecha-

1 nisms as the Director considers appropriate, including po-  
2 tential contractual incentives, to ensure the provision of  
3 access to proprietary models by qualified independent,  
4 third-party researchers, provided that commercial model  
5 vendors have voluntarily provided models and associated  
6 resources for such testing.

○