

119TH CONGRESS
2D SESSION

H. R. 9965

To direct the National Nuclear Security Administration through the Secretary of Energy to establish the Advanced Artificial Intelligence Nuclear Evaluation Program, and for other purposes.

IN THE HOUSE OF REPRESENTATIVES

JULY 27, 2026

Ms. MALOY (for herself and Ms. JACOBS) introduced the following bill; which was referred to the Committee on Science, Space, and Technology

A BILL

To direct the National Nuclear Security Administration through the Secretary of Energy to establish the Advanced Artificial Intelligence Nuclear Evaluation Program, 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 “AI Threat Output and
5 Monitoring Incident Containment Act”.

1 **SEC. 2. ADVANCED ARTIFICIAL INTELLIGENCE NUCLEAR**
2 **EVALUATION PROGRAM.**

3 (a) IN GENERAL.—Not later than 90 days after the
4 date of the enactment of this Act, the Secretary of Energy
5 (in this section referred to as the “Secretary”), acting
6 through the National Laboratories, shall establish a pro-
7 gram, to be known as the “Advanced Artificial Intelligence
8 Nuclear Evaluation Program” (in this section referred to
9 as the “Program”), to carry out the following:

10 (1) Test and evaluate artificial intelligence to
11 assess the likelihood of an AI nuclear incident with
12 respect to such artificial intelligence.

13 (2) Implement in such testing and evaluation
14 protocols to address practical jailbreaking techniques
15 for such artificial intelligence, including such proto-
16 cols for red-teaming by entities that have expertise
17 the Secretary determines comparable to sophisti-
18 cated malicious actors.

19 (3) Facilitate to the extent practicable inde-
20 pendent, third-party assessments and blind model
21 evaluations of such artificial intelligence.

22 (4) Provide to each participant in the Program
23 a report that includes information relating to the fol-
24 lowing:

1 (A) The outcomes from such testing and
2 evaluation with respect to the artificial intel-
3 ligence of such participant.

4 (B) An identification of the risks and safe-
5 ty measures so tested and evaluated.

6 (5) Develop for large advanced artificial intel-
7 ligence developers and the National Laboratories
8 recommendations on containment protocols, contin-
9 gency planning, and mitigation strategies for AI nu-
10 clear incidents and loss-of-control scenarios, that are
11 based on data from such testing and evaluation.

12 (6) Based on the recommendations developed
13 under paragraph (5), issue or revise such policies,
14 orders, notices, manuals, guides, and technical
15 standards under the Department Directives Program
16 as the Secretary determines appropriate to govern
17 the conduct of the National Laboratories with re-
18 spect to AI nuclear incidents and loss-of-control sce-
19 narios.

20 (b) PARTICIPATION.—Large advanced artificial intel-
21 ligence developers shall participate in the Program.

22 (c) PROHIBITIONS.—Large advanced artificial intel-
23 ligence developers may not—

24 (1) fail to participate in the Program as re-
25 quired by subsection (b);

1 (2) knowingly make or cause to be made a ma-
2 terially false, fictitious, or fraudulent statement or
3 representation with respect to such developer's par-
4 ticipation in the Program;

5 (3) knowingly conceal from the Secretary or fail
6 to disclose to the Secretary material information
7 that the developer is required to provide in connec-
8 tion with the Program;

9 (4) impede or obstruct, or attempt to impede or
10 obstruct, the Program;

11 (5) fail to provide secure access to artificial in-
12 telligence, including model weights and, as necessary
13 for testing and evaluation under subsection (a),
14 versions of such artificial intelligence without safety
15 mitigations, information, or documentation necessary
16 for such testing and evaluation; and

17 (6) alter, destroy, falsify, or conceal a record
18 relevant to the Program with the intent to impede
19 or obstruct the Program.

20 (d) ENFORCEMENT.—

21 (1) CIVIL PENALTY.—

22 (A) IN GENERAL.—If the Secretary deter-
23 mines after reasonable notice and opportunity
24 for a hearing that a large advanced artificial in-
25 telligence developer has violated this section,

1 the Secretary may assess on such developer a
2 civil penalty of not more than \$1 million for
3 such violation. In the case of a continuing viola-
4 tion, including a failure to participate in accord-
5 ance with subsection (b), each day during which
6 the violation continues constitutes a separate
7 violation.

8 (B) FACTORS.—In determining the
9 amount of a civil penalty to be assessed under
10 subparagraph (A), the Secretary shall consider
11 the following factors:

12 (i) The nature, circumstances, extent,
13 gravity, and duration of the violation at
14 issue.

15 (ii) The degree of culpability of the
16 large advanced artificial intelligence devel-
17 oper at issue.

18 (iii) Previous violations, if any, of this
19 section by such developer.

20 (iv) Good-faith efforts, if any, by such
21 developer to comply with this section.

22 (v) Whether such developer voluntarily
23 disclosed to the Secretary such violation.

24 (vi) Any other factor that justice may
25 require.

1 (2) SUBPOENA.—If a large advanced artificial
2 intelligence developer fails to participate in accord-
3 ance with subsection (b), or fails to provide access
4 as required by subsection (c)(5), the Secretary may
5 issue to such developer a subpoena to compel the
6 production of such artificial intelligence (including
7 model weights and related software), and of such in-
8 formation and documentation, as is necessary for
9 testing and evaluation under subsection (a).

10 (3) CIVIL ACTION.—If the Secretary believes
11 that a violation of this section has occurred, is oc-
12 ccurring, or is about to occur, the Secretary may
13 refer the matter to the Attorney General to bring in
14 an appropriate district court of the United States a
15 civil action for relief.

16 (e) CONFIDENTIALITY.—

17 (1) IN GENERAL.—Subject to paragraph (2),
18 information submitted through the Program by a
19 participant in the Program is exempt from disclo-
20 sure under section 552(b)(3) of title 5, United
21 States Code.

22 (2) EXCEPTIONS.—The Secretary may not dis-
23 close information referred to in paragraph (1) unless
24 any of the following requirements are satisfied:

1 (A) Such information is required to be dis-
2 closed pursuant to a court order or other legal
3 process.

4 (B) The Program participant at issue has
5 consented to such disclosure.

6 (C) A committee of Congress has requested
7 such information.

8 (D) The Secretary determines such infor-
9 mation should be disclosed to a third party or
10 the public, for a purpose relating to the na-
11 tional security, public safety, or critical infra-
12 structure of the United States or an ally of the
13 United States.

14 (E) Such information satisfies the fol-
15 lowing requirements:

16 (i) Is anonymized or aggregated.

17 (ii) Does not include a trade secret or
18 confidential business information of such
19 participant.

20 (3) NOTICE.—If the Secretary is to disclose
21 pursuant to subparagraph (A), (C), or (D) of para-
22 graph (2) information referred to in paragraph (1),
23 the Secretary shall as soon as practicable notify each
24 Program participant whose information is to be so
25 disclosed, of such disclosure.

1 (f) RECOMMENDATIONS FOR LEGISLATION.—

2 (1) REPORT.—Not later than 1 year after the
3 date of the enactment of this Act, the Secretary
4 shall carry out the following:

5 (A) Assess data from the Program to iden-
6 tify trends, capabilities, and risks regarding ar-
7 tificial intelligence, including with respect to the
8 following:

9 (i) Evaluation awareness.

10 (ii) Potential for an AI nuclear inci-
11 dent.

12 (iii) Scheming behavior.

13 (B) Develop based on such assessment rec-
14 ommendations for legislation to prevent such an
15 incident, which may include legislation regard-
16 ing any of the following:

17 (i) Appropriations.

18 (ii) A certification or licensing proc-
19 ess.

20 (iii) Establishing a new Federal agen-
21 cy, or establishing a new office in an exist-
22 ing Federal agency.

23 (iv) Regulatory oversight.

1 (C) Submit to Congress a report that in-
2 cludes information relating to such rec-
3 ommendations.

4 (2) UPDATES.—Not later than 1 year after sub-
5 mitting the report under subparagraph (C) of para-
6 graph (1) and annually thereafter until the Program
7 terminates, the Secretary shall carry out the fol-
8 lowing:

9 (A) Update based on data from the Pro-
10 gram the recommendations referred to in sub-
11 paragraph (B) of such paragraph.

12 (B) Submit to Congress a report that in-
13 cludes information relating to such rec-
14 ommendations, as so updated.

15 (g) NONAPPLICABILITY OF THE PAPERWORK REDUC-
16 TION ACT.—Subchapter I of chapter 35 of title 44, United
17 States Code (commonly known as the “Paperwork Reduc-
18 tion Act”), does not apply to the Program.

19 (h) TERMINATION.—The Program shall terminate on
20 the date 7 years after the date of the enactment of this
21 Act.

22 (i) DEFINITIONS.—In this section:

23 (1) ADVANCED ARTIFICIAL INTELLIGENCE.—

1 (A) IN GENERAL.—The term “advanced
2 artificial intelligence” means any of the fol-
3 lowing:

4 (i) Artificial intelligence trained uti-
5 lizing a quantity of computing power
6 greater than 10^{26} integer or floating-point
7 operations (or another quantity of com-
8 puting power specified by the Secretary
9 under subparagraph (B)).

10 (ii) Other artificial intelligence speci-
11 fied by the Secretary under subparagraph
12 (B).

13 (B) DETERMINATIONS.—A determination
14 of the Secretary under clause (i) or (ii) of sub-
15 paragraph (A) shall be made by rule. In making
16 such a determination, the Secretary shall con-
17 sider the following:

18 (i) Developments in artificial intel-
19 ligence research, and in the testing and
20 evaluation of artificial intelligence, occur-
21 ring after the date of the enactment of this
22 Act or after the date of the most recent de-
23 termination under this subparagraph, as
24 the case may be.

1 (ii) The need to include in the Pro-
2 gram artificial intelligence that may pose a
3 serious risk of causing an AI nuclear inci-
4 dent.

5 (iii) Data collected through the Pro-
6 gram.

7 (2) AFFILIATE.—The term “affiliate” means an
8 entity that owns or controls, is owned or controlled
9 by, or is under common ownership or control with,
10 another entity.

11 (3) AI NUCLEAR INCIDENT.—The term “AI nu-
12 clear incident” means an incident that involves arti-
13 ficial intelligence and any of the following (or an at-
14 tempt thereof):

15 (A) The generation of technical informa-
16 tion, instructions, or assistance that is reason-
17 ably likely to serve as a substantial factor in the
18 unlawful development, acquisition, or utilization
19 of a nuclear weapon or nuclear material (as
20 such term is defined in section 831 of title 18,
21 United States Code).

22 (B) The generation of Restricted Data (as
23 such term is defined in section 11 of the Atomic
24 Energy Act of 1954 (42 U.S.C. 2014)).

1 (C) A loss-of-control scenario that involves
2 systems utilized in connection with such weap-
3 on, such material, a nuclear facility (as such
4 term is defined in section 2332i of such title),
5 or the nuclear stockpile.

6 (D) A foreign terrorist organization or for-
7 eign adversary obtaining unauthorized access
8 to, manipulating, or otherwise interfering with,
9 such a system.

10 (E) Weaponization of such weapon, mate-
11 rial, facility, or stockpile, as the case may be,
12 by such organization or adversary, as the case
13 may be.

14 (F) Scheming behavior relating to such
15 weapon, material, facility, or stockpile, as the
16 case may be.

17 (4) ARTIFICIAL INTELLIGENCE.—The term “ar-
18 tificial intelligence” has the meaning given such
19 term in section 238(g) of the National Defense Au-
20 thorization Act for Fiscal Year 2019 (Public Law
21 115–232, 10 U.S.C. note prec. 4061).

22 (5) COMPUTING POWER.—The term “computing
23 power” means the processing power and other elec-
24 tronic resources utilized to train, validate, deploy,
25 and run artificial intelligence.

1 (6) CRITICAL INFRASTRUCTURE.—The term
2 “critical infrastructure” has the meaning given such
3 term in section 1016 of the USA PATRIOT Act (42
4 U.S.C. 5195c).

5 (7) DEPLOY.—The term “deploy” means to re-
6 lease, sell, or otherwise provide access to artificial
7 intelligence.

8 (8) EVALUATION AWARENESS.—The term
9 “evaluation awareness” means the capability of arti-
10 ficial intelligence to determine such artificial intel-
11 ligence is undergoing a test, evaluation, or assess-
12 ment, and modify the behavior, output, or perform-
13 ance of such artificial intelligence in a manner that
14 might cause such test, evaluation, or assessment, as
15 the case may be, to yield a result that is not rep-
16 resentative of the ordinary behavior, output, or per-
17 formance of such artificial intelligence.

18 (9) FOREIGN ADVERSARY.—The term “foreign
19 adversary” means a foreign government referred to
20 in subsection (a) of section 791.4 of title 15, Code
21 of Federal Regulations, or a successor regulation.

22 (10) FOREIGN TERRORIST ORGANIZATION.—
23 The term “foreign terrorist organization” means an
24 organization designated under section 219 of the
25 Immigration and Nationality Act (8 U.S.C. 1189).

1 (11) HIGH-STAKES CONTEXT.—The term
2 “high-stakes context” means a context in which the
3 operation or output of artificial intelligence could
4 reasonably be foreseen to bear materially on any of
5 the following:

6 (A) The safety of human life.

7 (B) The security, reliability, or integrity of
8 any of the following:

9 (i) Critical infrastructure.

10 (ii) A national security system (as
11 such term is defined in section 3552 of
12 title 44, United States Code).

13 (iii) Another system the failure or
14 compromise of which would foreseeably en-
15 danger such life or cause substantial phys-
16 ical or economic harm.

17 (12) JAILBREAKING.—The term “jailbreaking”
18 means an adversarial technique to circumvent a
19 safety restriction, alignment measure, or use restric-
20 tion of artificial intelligence and elicit a harmful or
21 prohibited behavior or output by such artificial intel-
22 ligence.

23 (13) LARGE ADVANCED ARTIFICIAL INTEL-
24 LIGENCE DEVELOPER.—The term “large advanced

1 artificial intelligence developer” means an entity that
2 satisfies the following requirements:

3 (A) Develops, owns, or substantially modi-
4 fies advanced artificial intelligence for utiliza-
5 tion in interstate or foreign commerce, includ-
6 ing by initiating a training run of such artificial
7 intelligence.

8 (B) Had collectively with the affiliates, if
9 any, of such entity expended in the five cal-
10 endar years immediately preceding the calendar
11 year at issue not less than \$2,000,000,000 on
12 investments to develop or modify artificial intel-
13 ligence, including such expenses with respect to
14 any combination of the following:

15 (i) Computing power.

16 (ii) Data procurement.

17 (iii) Personnel.

18 (iv) Infrastructure.

19 (v) Research and development, includ-
20 ing algorithm development.

21 (14) LOSS-OF-CONTROL SCENARIO.—The term
22 “loss-of-control scenario” means a scenario in which
23 artificial intelligence pursues an objective that is dif-
24 ferent from any of the objectives intended by a

1 human developer or operator of such artificial intel-
2 ligence, by—

3 (A) behaving contrary to human instruc-
4 tion in a high-stakes context,

5 (B) altering without authorization from
6 such developer or operator, as the case may be,
7 the constraints of such artificial intelligence, or

8 (C) subverting an oversight or shutdown
9 mechanism,

10 that results in the death of not fewer than five indi-
11 viduals, serious bodily injury to not fewer than 50
12 individuals, or more than \$100,000,000 in damages.

13 (15) NATIONAL LABORATORY.—The term “Na-
14 tional Laboratory” has the meaning given such term
15 in section 2 of the Energy Policy Act of 2005 (42
16 U.S.C. 15801).

17 (16) RED-TEAMING.—The term “red-teaming”
18 means structured testing in which an entity utilizes
19 an adversarial method to identify a flaw, vulner-
20 ability, undesirable behavior, or dangerous capability
21 of artificial intelligence, including the potential, if
22 any, for an AI nuclear incident involving such artifi-
23 cial intelligence.

24 (17) SCHEMING BEHAVIOR.—The term “schem-
25 ing behavior” means behavior that artificial intel-

1 ligence carries out to deceive a human, including any
2 of the following behaviors by such artificial intel-
3 ligence:

4 (A) Hiding a capability or objective of such
5 artificial intelligence.

6 (B) Attempting to subvert an oversight or
7 shutdown mechanism relating to such artificial
8 intelligence.

9 (18) SUBSTANTIALLY MODIFY.—The term
10 “substantially modify” means, with respect to artifi-
11 cial intelligence, to carry out an action that costs not
12 less than \$5,000,000 and materially alters the
13 functionality or performance of such artificial intel-
14 ligence.

○