

119TH CONGRESS
1ST SESSION

S. 3336

To require the Secretary of Homeland Security to carry out prize competitions to advance the science of interpretability and to develop adversarial robustness with respect to artificial intelligence products, and for other purposes.

IN THE SENATE OF THE UNITED STATES

DECEMBER 3, 2025

Ms. HASSAN (for herself and Mr. BANKS) introduced the following bill; which was read twice and referred to the Committee on Homeland Security and Governmental Affairs

A BILL

To require the Secretary of Homeland Security to carry out prize competitions to advance the science of interpretability and to develop adversarial robustness with respect to artificial intelligence products, 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 “Reliable Artificial In-
5 telligence Research Act of 2025”.

6 **SEC. 2. DEFINITIONS.**

7 In this Act:

1 (1) ADVERSARIAL ROBUSTNESS.—The term
2 “adversarial robustness” means the degree to which
3 an artificial intelligence model is able to resist at-
4 tacks that would induce it to produce incorrect, re-
5 stricted, or harmful outputs, while maintaining in-
6 tegrity, reliability, and privacy.

7 (2) ARTIFICIAL INTELLIGENCE.—The term “ar-
8 tificial intelligence” has the meaning given the term
9 in section 5002 of the National Artificial Intelligence
10 Initiative Act of 2020 (15 U.S.C. 9401).

11 (3) INTERPRETABILITY.—The term
12 “interpretability” means the degree to which hu-
13 mans are able to accurately understand how an arti-
14 ficial intelligence model makes decisions and con-
15 siders inputs and how the outputs or behaviors of
16 the model respond to a change in the inputs.

17 (4) RED-TEAMING.—The term “red-teaming”
18 means a structured, interactive, and adversarial
19 process to test an artificial intelligence system by
20 simulating real-world actions to find vulnerabilities
21 or flaws in the system.

22 (5) SECRETARY.—The term “Secretary” means
23 the Secretary of Homeland Security.

1 **SEC. 3. PRIZE COMPETITION FOR ARTIFICIAL INTEL-**
2 **LIGENCE INTERPRETABILITY RESEARCH.**

3 (a) PRIZE COMPETITION REQUIRED.—Not later than
4 270 days after the date of enactment of this Act, the Sec-
5 retary of Homeland Security shall commence carrying out
6 at least 1 prize competition under section 24 of the Ste-
7 venson-Wydler Technology Innovation Act of 1980 (15
8 U.S.C. 3719) to advance the science of interpretability in
9 a manner relevant to commercially available or widely used
10 artificial intelligence products.

11 (b) CONSULTATION.—In carrying out the prize com-
12 petition required by subsection (a), the Secretary shall
13 consult with—

14 (1) the Secretary of Commerce;

15 (2) the Director of the National Institute of
16 Standards and Technology;

17 (3) the National Cyber Director;

18 (4) the Director of the National Science Foun-
19 dation; and

20 (5) any industry expert from the artificial intel-
21 ligence sector in the United States that the Sec-
22 retary considers relevant.

23 (c) STRUCTURE AND EVALUATION CRITERIA.—

24 (1) IN GENERAL.—The Secretary shall develop
25 the structure and evaluation criteria for a prize com-
26 petition carried out under subsection (a) in accord-

1 ance with the primary purpose described in that sub-
2 section.

3 (2) COMPETITION STRUCTURE.—The Secretary
4 may—

5 (A) structure a competition under sub-
6 section (a) into 1 or more phases, including
7 submission of interpretability frameworks, sub-
8 mission of interpretable artificial intelligence
9 models, and unique basic research; and

10 (B) open these phases to the same, or to
11 distinct, contestant pools.

12 (3) EVALUATION CONSIDERATIONS.—In devel-
13 oping the evaluation criteria for the frameworks,
14 models, or methods submitted for a prize competi-
15 tion under subsection (a), the Secretary shall con-
16 sider—

17 (A) the degree to which a submission ad-
18 vances broadly applicable principles of artificial
19 intelligence interpretability;

20 (B) the practical value of a submission in
21 making artificial intelligence more understand-
22 able and reliable in high-risk, high-value use
23 cases; and

24 (C) the likelihood that the unique research
25 submitted will create standards for artificial in-

1 intelligence interpretability in the government or
2 industry.

3 (d) PROGRAM ADMINISTRATION.—The Secretary
4 may enter into contracts, cooperative agreements, or other
5 agreements with for-profit or nonprofit entities or State,
6 territorial, local, or Tribal agencies to design and admin-
7 ister any prize competition carried out under subsection
8 (a).

9 **SEC. 4. PRIZE COMPETITION FOR ARTIFICIAL INTEL-**
10 **LIGENCE ADVERSARIAL ROBUSTNESS RE-**
11 **SEARCH.**

12 (a) PRIZE COMPETITION REQUIRED.—Not later than
13 270 days after the date of enactment of this Act, the Sec-
14 retary shall commence carrying out at least 1 prize com-
15 petition under section 24 of the Stevenson-Wydler Tech-
16 nology Innovation Act of 1980 (15 U.S.C. 3719) to de-
17 velop capable artificial intelligence models that are de-
18 signed to exhibit adversarial robustness in circumstances
19 necessary for at least 1 high-impact, high-risk application
20 in government or industry.

21 (b) CONSULTATION.—In carrying out a prize com-
22 petition required by subsection (a), the Secretary shall
23 consult with—

24 (1) the Secretary of Commerce;

1 (2) the Director of the Institute of Standards
2 and Technology;

3 (3) the National Cyber Director;

4 (4) the Director of the National Science Foun-
5 dation;

6 (5) any industry expert from the artificial intel-
7 ligence sector in the United States that the Sec-
8 retary considers relevant; and

9 (6) the head of any Federal agency who has au-
10 thority or expertise in a high-impact, high-risk appli-
11 cation of artificial intelligence that could be an ap-
12 propriate subject for a prize competition under sub-
13 section (a).

14 (c) STRUCTURE AND EVALUATION CRITERIA.—

15 (1) IN GENERAL.—The Secretary shall develop
16 the structure and evaluation criteria for a prize com-
17 petition carried out under subsection (a) in accord-
18 ance with the primary purpose described in that sub-
19 section.

20 (2) COMPETITION STRUCTURE.—The Secretary
21 may—

22 (A) structure a competition under sub-
23 section (a) into 1 or more phases, including
24 submission of adversarial robustness frame-

1 works, submission of artificial intelligence mod-
2 els, and red-teaming; and

3 (B) open these phases to the same, or to
4 distinct, contestant pools.

5 (3) EVALUATION CONSIDERATIONS.—In devel-
6 oping the evaluation criteria for the frameworks,
7 models, or methods submitted for a prize competi-
8 tion under subsection (a), the Secretary shall con-
9 sider—

10 (A) the degree to which a submission ad-
11 vances broadly applicable principles of artificial
12 intelligence robustness; and

13 (B) the practical value of the submission in
14 reducing the risk of adversarial attacks in high-
15 risk, high-value use cases of artificial intel-
16 ligence.

17 (d) PROGRAM ADMINISTRATION.—The Secretary
18 may enter into contracts, cooperative agreements, or other
19 agreements with for-profit or nonprofit entities or State,
20 territorial, local, or Tribal agencies to design and admin-
21 ister any prize competition carried out under subsection
22 (a).

23 **SEC. 5. TRACKING AND REPORTING.**

24 (a) IN GENERAL.—Not later than 180 days after the
25 date on which the first prize competition concludes, the

1 Secretary shall submit to the appropriate congressional
2 committees a report that includes—

3 (1) an evaluation of how the results of the com-
4 petitions inform the fields of interpretability and ad-
5 versarial robustness;

6 (2) an assessment of any gaps in these fields
7 identified by the Secretary over the course of the
8 competitions; and

9 (3) any suggested action that Congress should
10 take to advance the fields of interpretability, adver-
11 sarial robustness, and any related research.

12 (b) APPROPRIATE CONGRESSIONAL COMMITTEES
13 DEFINED.—In this section, the term “appropriate con-
14 gressional committees” means—

15 (1) the Committee on Homeland Security and
16 Governmental Affairs of the Senate; and

17 (2) the Committee on Homeland Security of the
18 House of Representatives.

19 **SEC. 6. APPROPRIATIONS.**

20 There is authorized to be appropriated to the Sec-
21 retary to carry out this section \$10,000,000 for the period
22 of fiscal years 2026 through 2030.

○