

119TH CONGRESS
2D SESSION

S. _____

To require disclosure of certain information relating to artificial intelligence systems, and for other purposes.

IN THE SENATE OF THE UNITED STATES

Mr. COONS (for himself, Mrs. BRITT, Mr. SCHATZ, and Mr. LANKFORD) introduced the following bill; which was read twice and referred to the Committee on _____

A BILL

To require disclosure of certain information relating to artificial intelligence systems, 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 Systems Trans-
5 parency Act”.

6 **SEC. 2. DEFINITIONS.**

7 In this Act:

8 (1) AI RED-TEAMING.—The term “AI red-
9 teaming” means a structured testing effort to find

1 flaws and vulnerabilities of an AI system, often per-
2 formed—

3 (A) in a controlled environment and in col-
4 laboration with the developers of the AI system;
5 and

6 (B) by dedicated teams that adopt adver-
7 sarial methods to identify any flaws or
8 vulnerabilities of the AI system, which may in-
9 clude harmful or discriminatory outputs, un-
10 foreseen or undesirable system behaviors, limi-
11 tations, and potential risks associated with mis-
12 use.

13 (2) ARTIFICIAL INTELLIGENCE MODEL.—The
14 term “artificial intelligence model” means the part
15 of an artificial intelligence system that can be ad-
16 justed to modify how the system behaves during in-
17 ference.

18 (3) ARTIFICIAL INTELLIGENCE SYSTEM.—The
19 term “artificial intelligence system” has the meaning
20 given the term “artificial intelligence” in section
21 5002 of the National Artificial Intelligence Initiative
22 Act of 2020 (15 U.S.C. 9401).

23 (4) CLOUD PROVIDER.—The term “cloud pro-
24 vider” means an entity that offers on-demand net-
25 work access to a shared pool of configurable com-

1 puting resources, such as networks, servers, storage,
2 and applications that can be rapidly provisioned and
3 released with minimal management effort.

4 (5) COMMISSION.—The term “Commission”
5 means the Federal Trade Commission.

6 (6) COVERED DEPLOYER.—The term “covered
7 deployer” means an entity that operates a covered
8 system which—

9 (A) for the majority of the months in any
10 1-year period, has, or is expected to have, not
11 fewer than 1,000,000 unique monthly users in
12 the United States; or

13 (B) allows minors to create accounts or
14 profiles and has, or is expected to have, not
15 fewer than 100,000 unique monthly users in
16 the United States.

17 (7) COVERED MODEL.—The term “covered
18 model” means a generative artificial intelligence
19 model, including an open weight model or an open
20 source model, that—

21 (A) either—

22 (i) is accessible to the public through
23 a website, software application, an applica-
24 tion programming interface, or a licensure

1 through a cloud provider, including as part
2 of a broader AI system; or

3 (ii) the model weights or architecture
4 of which are made available to the public
5 through the website of a developer,
6 download partners, cloud partners, or soft-
7 ware development platform, including in a
8 situation in which accessing such model
9 weights or architecture requires a user to
10 create an account; and

11 (B) is deployed in covered systems, wheth-
12 er deployed by the developer or another
13 deployer, that—

14 (i) together have, or are expected to
15 have, not fewer than 1,000,000 unique
16 monthly users in the United States for a
17 majority of the months in any 1-year pe-
18 riod; or

19 (ii) allow minors to create accounts or
20 profiles and that together have, or are ex-
21 pected to have, not fewer than 100,000
22 unique monthly users in the United States
23 for a majority of the months in any 1-year
24 period.

1 (8) COVERED PLATFORM.—The term “covered
2 platform” means a website or software application
3 used to interface with a covered system.

4 (9) COVERED SYSTEM.—The term “covered sys-
5 tem” means a covered model or collection of covered
6 models, together with any software component or in-
7 frastructure that—

8 (A) are deployed as an integrated unit; and

9 (B) operate in a coordinated manner to—

10 (i) handle different tasks or modali-
11 ties; or

12 (ii) materially augment or enhance the
13 capabilities of the covered model.

14 (10) DEVELOPER.—The term “developer”—

15 (A) means an entity that initially builds,
16 designs, codes, produces, pre-trains, or owns an
17 artificial intelligence system for internal use or
18 for use by a third party; and

19 (B) does not include an entity that is sole-
20 ly—

21 (i) a deployer of the artificial intel-
22 ligence system; or

23 (ii) a substantial modifier of the arti-
24 ficial intelligence system.

1 (11) GENERATIVE ARTIFICIAL INTELLIGENCE

2 MODEL.—The term “generative artificial intelligence

3 model” means an artificial intelligence model that—

4 (A) emulates the structure and characteris-

5 tics of input data in order to generate derived

6 synthetic data, including images, videos, audio,

7 text, code, and other digital content;

8 (B) is trained on broad data;

9 (C) generally uses self-supervision;

10 (D) is designed for generality of output;

11 and

12 (E) is generally applicable across a wide

13 range of contexts, adaptable to a wide range of

14 tasks, or can issue a wide range of outputs in

15 response to inferences.

16 (12) MACHINE LEARNING.—The term “machine

17 learning” has the meaning given the term in section

18 5002 of the National Artificial Intelligence Initiative

19 Act of 2020 (15 U.S.C. 9401).

20 (13) MINOR.—The term “minor” means an in-

21 dividual that has not attained 18 years of age.

22 (14) MODEL INFERENCE.—The term “model

23 inference” means the process of running data points

24 through an AI system to receive an output, such as

25 a number or text.

1 (15) MODEL WEIGHT.—The term “model
2 weight” means a numerical parameter within an ar-
3 tificial intelligence model that helps determine the
4 outputs of the model in response to inputs.

5 (16) OPEN WEIGHT MODEL.—The term “open
6 weight model” means a machine learning model, the
7 weights of which are made publicly available, allow-
8 ing for access, inspection, modification, and redis-
9 tribution by any third party.

10 (17) OPEN SOURCE MODEL.—The term “open
11 source model” means a machine learning model, the
12 weight and model architecture, training code, and
13 methodology of which are made publicly available,
14 allowing for a third party to fully reproduce, audit,
15 and modify the model for custom applications.

16 (18) POST TRAINING AND INFERENCE EN-
17 HANCEMENTS.—The term “post training and infer-
18 ence enhancements” means a process of adjusting or
19 supplementing a pre-trained artificial intelligence
20 model to improve performance on specific tasks, bet-
21 ter meet desired performance characteristics, or ad-
22 here to safety constraints.

23 (19) SENIOR.—The term “senior” means an in-
24 dividual that has attained 65 years of age.

1 (20) SUBSTANTIAL MODIFIER.—The term “sub-
2 stantial modifier” means an entity that significantly
3 alters (as defined by the Commission) the behavior
4 or capabilities of a covered model through post
5 training and inference enhancements.

6 (21) SYNTHETIC DATA.—The term “synthetic
7 data” means data that is artificially manufactured,
8 typically by a computer program.

9 (22) VULNERABLE USER.—The term “vulner-
10 able user” means a minor, a senior, an individual
11 with a disability, or an individual experiencing a
12 mental health crisis.

13 **SEC. 3. GUIDELINES FOR MODEL VERSIONING AND OTHER**
14 **DESIGNATIONS.**

15 (a) MODEL VERSIONING.—Not later than 150 days
16 after the date of enactment of this Act, the Commission,
17 in consultation with the Director of the National Institute
18 of Standards and Technology, shall issue guidance regard-
19 ing when a developer has produced a distinct generative
20 artificial intelligence model through—

- 21 (1) development, including novel pre-training of
22 a model architecture; or
23 (2) any update to, or improvement or iteration
24 of, an existing generative artificial intelligence
25 model.

1 (b) DESIGNATION PROCESS.—

2 (1) IN GENERAL.—Not later than 150 days
3 after the date of enactment of this Act, the Commis-
4 sion, in consultation with the Director of the Na-
5 tional Institute of Standards and Technology, shall
6 promulgate regulations in accordance with section
7 553 of title 5, United States Code, to establish cri-
8 teria—

9 (A) to determine when a deployment of a
10 generative artificial intelligence model is ex-
11 pected to have sufficient users to meet the cri-
12 teria described in—

13 (i) paragraph (6) of section 2 for a
14 deployer to be considered a covered
15 deployer; or

16 (ii) paragraph (7) of section 2 for a
17 generative artificial intelligence model to be
18 considered a covered model; and

19 (B) to designate a generative artificial in-
20 telligence model or a particular deployment of a
21 generative artificial intelligence model as a cov-
22 ered model or covered system, based on a deter-
23 mination that it poses a significant risk to safe-
24 ty of minors, vulnerable users, or the public.

1 (2) REQUIREMENTS.—In establishing the cri-
2 teria to make a designation under paragraph (1)(B),
3 the Commission, in consultation with the Director of
4 the National Institute of Standards and Technology,
5 shall do the following:

6 (A) Consider the capabilities of the genera-
7 tive artificial intelligence model, the contexts in
8 which the model is deployed or reasonably likely
9 to be deployed (including the degree of access
10 by vulnerable users), and the severity and re-
11 versibility of the harms reasonably foreseeable
12 from those capabilities.

13 (B) Accompany any such designation with
14 written findings, which shall be made publicly
15 available.

16 (C) Provide the developer of the model or
17 the deployer of the system with written notice
18 and not fewer than 30 days to respond before
19 any such designation takes effect, except that a
20 designation may take effect immediately upon a
21 finding by the Commission of imminent and sig-
22 nificant risk, in which case the Commission
23 shall provide such notice and opportunity to re-
24 spond not later than 30 days after the effective
25 date and shall confirm, modify, or rescind the

1 designation not later than 60 days after the ef-
2 fective date.

3 **SEC. 4. ARTIFICIAL INTELLIGENCE MODEL DISCLOSURE**
4 **MANDATES FOR DEVELOPERS.**

5 (a) REGULATIONS.—

6 (1) IN GENERAL.—Not later than 150 days
7 after the date of enactment of this Act, the Commis-
8 sion, in consultation with the Director of the Na-
9 tional Institute of Standards and Technology, shall
10 promulgate regulations in accordance with section
11 553 of title 5, United States Code, requiring each
12 developer of a covered model (including a distinct
13 generative artificial intelligence model that meets the
14 definition of a covered model) to make publicly avail-
15 able the information described in subsections (b) and
16 (c)—

17 (A) in a form determined by the Commis-
18 sion that includes 2 separate disclosures, in-
19 cluding—

20 (i) 1 disclosure containing information
21 relevant to, and sufficiently detailed to fa-
22 cilitate clear understanding for, a con-
23 sumer; and

24 (ii) 1 disclosure containing all infor-
25 mation described in subsections (b) and (c)

sufficiently detailed to permit meaningful
review by an independent researcher; and

(B) in a manner that—

(i) is accurate, materially complete,
not misleading by reason of a material
omission, and adequate under standards
determined by the Commission;

(ii) does not rely primarily on general-
ized or conclusory language;

(iii) facilitates transparency, public
trust, and independent evaluation;

(iv) makes the information useful and
actionable, including that the information
is—

(I) reliably available in a clear
and conspicuous section of the website
of the developer that can be easily lo-
cated and accessed by a consumer in
alignment with the Plain Writing Act
of 2010 (5 U.S.C. 301 note);

(II) searchable; and

(III) available through an appli-
cation programming interface;

(v) does not disclose trade secrets;

and

1 (vi) does not expose classified systems
2 or create national security risks.

3 (2) CONSULTATION.—In promulgating the reg-
4 ulations under paragraph (1), the Commission and
5 the Director of the National Institute of Standards
6 and Technology shall engage relevant Federal agen-
7 cies (including the Department of Health and
8 Human Services, the Substance Abuse and Mental
9 Health Services Administration, and the National
10 Institute of Mental Health) and third-parties with
11 expertise relating to consumer protection, the safety
12 of minors, behavioral science, mental health, and
13 human-computer interaction to ensure that the dis-
14 closure required by such regulations reflects best
15 practices and includes the appropriate information.

16 (b) INITIAL PUBLIC DISCLOSURE.—

17 (1) IN GENERAL.—The information described in
18 this subsection shall include the following:

19 (A) Information relating to consumer
20 transparency, safety protocols for vulnerable
21 users, and other practices associated with the
22 covered model, including—

23 (i) an overview of preventative safety
24 safeguards for minors implemented during

1 pre-deployment testing and model develop-
2 ment;

3 (ii) an overview of measures taken to
4 protect user privacy (including the privacy
5 of minors) in training data, such as
6 through data filtering or anonymization
7 methods, and measures taken to protect
8 user privacy (including the privacy of mi-
9 nors) on covered platforms;

10 (iii) an overview of measures taken to
11 ensure the safety of minors in training
12 data, such as through data filtering of
13 child sexual abuse material;

14 (iv) an overview of privacy standards
15 pertaining to the management, use, stor-
16 age, and potential sale of user data (in-
17 cluding the data of minors), including
18 whether user data is used for targeted ad-
19 vertising, user safety, internal or external
20 research, and model training;

21 (v) the durability and efficacy of vul-
22 nerable user safeguards (including pre-
23 venting the generation of child sexual
24 abuse material and non-consensual inti-

1 mate imagery), as well as the resiliency
2 and scope of such features;

3 (vi) an overview of user and company
4 reporting procedures for model behavior
5 that poses a credible risk of material as-
6 sistance to suicidal ideation or suicide
7 planning, or that poses a credible risk of
8 material assistance to harm to others, in-
9 cluding model behavior learned in the post-
10 deployment phase, and whether such proce-
11 dures were informed by crisis intervention
12 experts and established standards; and

13 (vii) as appropriate, the risk of nega-
14 tive impacts for vulnerable users on topics
15 pertaining to mental health and well-being,
16 including disclosure of research on the
17 short-term and long-term emotional and
18 psychological impacts of use for vulnerable
19 users, and methodological standards for
20 such research.

21 (B) System overviews of the covered model,
22 including—

23 (i) a high-level description of the in-
24 tended, unintended, and prohibited uses of

1 the covered model, as reasonably antici-
2 pated by the developer;

3 (ii) a high-level description, including
4 abilities and limitations, of the covered
5 model;

6 (iii) if applicable, a high-level overview
7 of any difference between the covered
8 model that is the subject of the disclosure
9 and most recent previous version of the
10 covered model that is considered a distinct
11 generative artificial intelligence model ac-
12 cording to the guidance issued under sec-
13 tion 3;

14 (iv) a list of each covered platform
15 and interface through which the covered
16 model is accessible;

17 (v) a list of input and output formats
18 supported by the covered model; and

19 (vi) an overview of the ability of the
20 covered model to trigger actions or com-
21 putations external to the covered model
22 during model inference.

23 (C) An overview of the post training and
24 inference enhancements of the covered model
25 and relevant methodologies.

1 (D) Information relating to safety, secu-
2 rity, and evaluation protocols and practices as-
3 sociated with the covered model, including—

4 (i) a high-level overview of procedures
5 or guiding principles for—

6 (I) determining what constitutes
7 desirable model behavior;

8 (II) ensuring desirable covered
9 model behavior; and

10 (III) preventing undesirable cov-
11 ered model behavior;

12 (ii) a description of methods or ap-
13 proaches used to evaluate and mitigate the
14 risk that a model or the outputs of a re-
15 lated system are lower in quality or accu-
16 racy than intended use cases would require
17 or expect;

18 (iii) subject to subsection (d), an over-
19 view of measures taken to protect the cov-
20 ered model and training data from attack
21 or misuse, including—

22 (I) traditional cybersecurity pro-
23 tocols, such as access controls and
24 multi-factor authentication; and

1 (II) protections unique to AI sys-
2 tems, such as preventing model weight
3 exfiltration, data poisoning, and
4 prompt injections;

5 (iv) with respect to open weight mod-
6 els, an overview of the measures taken to
7 protect training data from attack;

8 (v) an overview of the criteria by
9 which the developer chose to publicly re-
10 lease the covered model, such as evaluation
11 against a risk framework;

12 (vi) a description of each of the eval-
13 uation criteria used to assess the risk pro-
14 file of the covered model, which shall ad-
15 dress, at a minimum, chemical, biological,
16 radiological, and nuclear risk, offensive
17 cyber risk, autonomous loss-of-control risk,
18 and any similar additional risk domain
19 specified by the Commission, and which
20 shall include, for any such domain in which
21 the developer conducted no evaluation, an
22 affirmative statement to that effect;

23 (vii) an overview of the testing proce-
24 dures and methods for achieving risk pro-
25 file evaluations, such as AI red-teaming,

1 collaboration with external subject matter
2 experts, or solicitation of reports of model
3 vulnerabilities from independent eval-
4 uators; and

5 (viii) an overview of the performance
6 of the covered model on common public
7 benchmarks determined by the Commission
8 in coordination with the Director of the
9 National Institute of Standards and Tech-
10 nology, including any related information
11 and in a form necessary for independent
12 researchers to verify the performance of
13 the covered model on each benchmark,
14 such as prompts and configuration set-
15 tings.

16 (2) TIMING OF DISCLOSURE.—After the date on
17 which the Commission issues guidance under section
18 3, a developer shall make publicly available the in-
19 formation described in paragraph (1)—

20 (A) with respect to the most recent version
21 of a covered model released to the public during
22 the 2-year period before such date, not later
23 than 90 days after such date;

1 (B) with respect to a covered model re-
2 leased on or after such date, concurrent with
3 such release; or

4 (C) with respect to the most recent version
5 of a distinct generative artificial intelligence
6 model released on or after such date that, at a
7 time after its release, meets the definition of a
8 covered model, not later than 90 days after
9 such time.

10 (c) SUBSEQUENT PUBLIC DISCLOSURES.—

11 (1) REQUIREMENT.—The information described
12 in paragraph (2), with respect to a covered model,
13 shall be made available to the public at such time,
14 as determined by the Commission, and on an ongoing
15 basis until the Commission determines that such
16 disclosure is no longer necessary.

17 (2) INFORMATION REQUIRED.—The information
18 described in this paragraph is as follows:

19 (A) Any significant change (as defined by
20 the Commission in coordination with the Direc-
21 tor of the National Institute of Standards and
22 Technology, including any substantial modifica-
23 tion to, or observed differences from, the most
24 recent prior public disclosure under this section,
25 and not including any immaterial variation for

1 which a disclosure requirement would overly
2 burden platform operations) to any of the infor-
3 mation described in subsection (b) from the in-
4 formation reported in the initial information
5 disclosure, or, if applicable, any information re-
6 ported in a subsequent disclosure.

7 (B) Information including—

8 (i) data dictionaries or equivalent doc-
9 umentation that explain significant
10 datasets, as defined by the Commission, in
11 the possession of a developer related to the
12 covered model, including datasets that the
13 developer maintains relating to the users of
14 the relevant covered platforms, usage of
15 the covered platforms, content generated
16 on covered platforms, and actions taken to
17 enforce the terms of service of covered
18 platforms; and

19 (ii) information relating to usage of
20 the covered model, including—

21 (I) an overview of the most com-
22 mon violations of covered model poli-
23 cies regarding acceptable use, includ-
24 ing common violations pertaining to
25 minor safety;

1 (II) an overview of how often the
2 model refuses a prohibited prompt
3 and the developer has taken modera-
4 tion actions in response to violations
5 of acceptable use policies, including
6 what those actions are;

7 (III) an overview of information
8 pertaining to attempts by users to uti-
9 lize covered models to generate child
10 sexual abuse material or non-consen-
11 sual intimate imagery; and

12 (IV) an overview of known sys-
13 tem capabilities for the utilization of
14 the covered model to develop or ad-
15 vance cyber, chemical, biological, radi-
16 ological, and nuclear threats.

17 (d) SPECIAL RULES FOR DISCLOSURES.—

18 (1) IN GENERAL.—With respect to any discolo-
19 sure required under clause (iii) or (vi) of subsection
20 (b)(1)(D) or under subsection (c)(2)(B)(ii), such re-
21 quirement shall not be interpreted to require a devel-
22 oper to disclose any information that, as determined
23 by the Commission—

1 (A) would compromise cybersecurity prac-
2 tices or protections or safety practices or pro-
3 tections of the developer; or

4 (B) could provide material assistance to
5 the development, acquisition, or deployment of
6 chemical, biological, radiological, nuclear, or of-
7 fensive cyber weapons or capabilities.

8 (2) RULE OF CONSTRUCTION.—Paragraph (1)
9 shall not be construed to permit a developer to with-
10 hold—

11 (A) the fact that an evaluation in a risk
12 domain described in subsection (b)(1)(D)(vi)
13 was or was not conducted; or

14 (B) a general characterization of the re-
15 sults of any such evaluation, at a level of detail
16 that does not include information described in
17 paragraph (1).

18 (e) REPORTING REQUIREMENT FOR SUBSTANTIAL
19 MODIFIERS.—Not later than 1 year after the date of en-
20 actment of this Act, the Commission, in consultation with
21 the Director of the National Institute of Standards and
22 Technology, shall promulgate regulations in accordance
23 with section 553 of title 5, United States Code, requiring
24 each substantial modifier to make publicly available the
25 information described in subsections (b) and (c), as the

1 Commission determines appropriate with respect to a sub-
2 stantial modifier.

3 (f) RESOURCES.—The Commission shall provide cov-
4 ered platforms with technical assistance to support compli-
5 ance with the requirements of this Act, particularly with
6 respect to small and mid-sized platforms, as determined
7 by the Commission.

8 **SEC. 5. MANDATORY DISCLOSURES FOR COVERED**
9 **DEPLOYERS OF COVERED SYSTEMS.**

10 Not later than 150 days after the date of enactment
11 of this Act, the Commission, in consultation with the Di-
12 rector of the National Institute of Standards and Tech-
13 nology, shall promulgate regulations in accordance with
14 section 553 of title 5, United States Code, to establish
15 disclosure requirements for covered deployers with respect
16 to covered systems consistent with the requirements for
17 developers of covered models described in section 4, as de-
18 termined by the Commission to be appropriate for a cov-
19 ered deployer with respect to a covered system.

20 **SEC. 6. ENFORCEMENT.**

21 (a) UNFAIR OR DECEPTIVE ACTS OR PRACTICES.—
22 A violation of section 4 or 5 or a regulation promulgated
23 under this Act shall be treated as a violation of a rule
24 defining an unfair or deceptive act or practice under sec-

1 tion 18(a)(1)(B) of the Federal Trade Commission Act
2 (15 U.S.C. 57a(a)(1)(B)).

3 (b) POWERS OF THE COMMISSION.—

4 (1) IN GENERAL.—The Commission shall en-
5 force this Act in the same manner, by the same
6 means, and with the same jurisdiction, powers, and
7 duties as though all applicable terms and provisions
8 of the Federal Trade Commission Act (15 U.S.C. 41
9 et seq.) were incorporated into and made a part of
10 this Act.

11 (2) PRIVILEGES AND IMMUNITIES.—Any person
12 who violates section 4 or 5 or a regulation promul-
13 gated under this Act shall be subject to the penalties
14 and entitled to the privileges and immunities pro-
15 vided in the Federal Trade Commission Act (15
16 U.S.C. 41 et seq.).

17 (3) AUTHORITY PRESERVED.—Nothing in this
18 Act shall be construed to limit the authority of the
19 Commission under any other provision of law.

20 (4) RULEMAKING.—Not later than 1 year after
21 the date of enactment of this Act, the Commission
22 shall promulgate in accordance with section 553 of
23 title 5, United States Code, such rules as may be
24 necessary to carry out this Act.

1 **SEC. 7. SEVERABILITY.**

2 If any provision of this Act, or the application of such
3 provision to any person or circumstance, is held to be un-
4 constitutional, the remainder of this Act, and the applica-
5 tion of the remaining provisions of this Act to any person
6 or circumstance, shall not be affected.