....

(Original Signature of Member)

118TH CONGRESS
1ST SESSION

H. R. _____

To direct the Federal Trade Commission to establish standards for making publicly available information about the training data and algorithms used in artificial intelligence foundation models, and for other purposes.

IN THE HOUSE OF REPRESENTATIVES

Mr. BEYER introduced the following bill; which was referred to the Committee on _____

A BILL

To direct the Federal Trade Commission to establish standards for making publicly available information about the training data and algorithms used in artificial intelligence foundation models, 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 Foundation Model

5 Transparency Act of 2023".

6 **SEC. 2. FINDINGS.**

7   Congress finds the following:

1   (1) With the increase in public access to artifi-
2 cial intelligence, there has been an increase in law-
3 suits and public concerns about copyright infringe-
4 ment, including in court cases such as the following:

5   (A) Doe 1 v. GitHub, Inc., No. 22-cv-
6   06823, 2023 WL 3449131, at *1 (N.D. Cal.
7   May 11, 2023).

8   (B) Amended Complaint, Getty Images,
9   Inc. v. Stability AI, Ltd., No. 23-cv-00135 (D.
10   Del. Mar. 29, 2023).

11   (C) Andersen v. Stability AI Ltd., No. 23-
12   cv-00201, 2023 WL 7132064, at *1 (N.D. Cal.
13   Oct. 30, 2023).

14   (2) Public use of foundation models has led to
15 countless instances of the public being presented
16 with inaccurate, imprecise, or biased information
17 during inference, based on limited training data, lim-
18 ited model training mechanisms, or a lack of disclo-
19 sures about the training data composition or founda-
20 tion model training procedures, including in facial
21 recognition technology usage, artificial intelligence
22 inferences relating to health, artificial intelligence in-
23 ferences relating to loan granting and housing ap-
24 proval, and more.

3

1   (3) Transparency with respect to high-impact

2 foundation models has become increasingly nec-

3 essary, including to assist copyright owners with en-

4 forcing their copyright protections and to promote

5 consumer protection.

6   (4) While not compromising the intellectual

7 property rights of those who develop and deploy

8 foundation models, users should be equipped with

9 the information necessary to enforce their copyright

10 protections and to make informed decisions about

11 such foundation models.

SEC. 3. FOUNDATION MODEL DATA SOURCE AND TRAINING

TRANSPARENCY.

14   (a) ESTABLISHMENT OF STANDARDS.—Not later

15 than 9 months after the date of the enactment of this Act,

16 the Commission shall—

17   (1) in accordance with section 553 of title 5,

18 United States Code, promulgate regulations that es-

19 tablish standards specifying information to improve

20 the transparency of foundation models by covered

21 entities with respect to training data, model docu-

22 mentation, data collection in inference, and oper-

23 ations of foundation models; and

1   (2) issue guidance to assist covered entities in

2   complying with the standards established under

3   paragraph (1).

4 (b) CONSULTATION.—In establishing the standards

5 and issuing the guidance required by subsection (a), the

6 Commission shall consult with the Director of the Na-

7 tional Institute of Standards and Technology, the Director

8 of the Office of Science and Technology Policy, the Reg-

9 ister of Copyrights, and other relevant stakeholders, in-

10 cluding standards bodies, covered entities, academia, tech-

11 nology experts, and advocates for civil rights and con-

12 sumers.

13 (c) SUBMISSION TO COMMISSION AND PUBLIC AVAIL-

14 ABILITY OF INFORMATION.—The standards established

15 under subsection (a)(1) shall include requirements, with

16 respect to a foundation model, for—

17   (1) what information specified under such sub-

18   section shall be submitted to the Commission by the

19   covered entity that provides such model; and

20   (2) what information specified under such sub-

21   section shall be made publicly available by the cov-

22   ered entity that provides such model.

23 (d) FORM AND MANNER.—The standards established

24 under subsection (a)(1) shall specify the form and manner

25 in which certain information specified under such sub-

1 section, selected at the discretion of the Commission, in

2 consultation with the Director of the National Institute

3 of Standards and Technology and the other actors de-

4 scribed in subsection (b), shall be made publicly available

5 by covered entities, including—

6   (1) what information shall be made available on

7   the website of a covered entity that relates to any

8   foundation model provided by such covered entity;

9   (2) what information shall be displayed in a

10   central location on a website hosted by the Commis-

11   sion, which shall include, with respect to a founda-

12   tion model, information that is substantially similar

13   to the information required under paragraph (1) to

14   be made available on the website of the covered enti-

15   ty that provides such model;

16   (3) that a machine-readable format shall be

17   used with respect to the information specified under

18   paragraphs (1) and (2);

19   (4) the URL at which the information specified

20   under paragraph (2) shall be hosted by the Commis-

21   sion; and

22   (5) such additional specifications as the Com-

23   mission considers appropriate.

24 (e) PROCESS.—The standards established under sub-

25 section (a)(1) shall specify a process by which the informa-

1 tion required under subsection (c)(1) shall be submitted
2 to the Commission.

3 (f) INFORMATION TO BE CONSIDERED.—The Com-
4 mission shall consider specifying in the standards estab-
5 lished under subsection (a)(1), with respect to a founda-
6 tion model, the following information:

7   (1) The sources of training data (including, as
8   applicable, personal data collection and information
9   necessary to assist copyright owners or data license
10   holders with enforcing their copyright or data license
11   protections) and whether and how data is collected
12   and retained during inference.

13   (2) A description of the size and composition of
14   such training data, including broad demographic in-
15   formation, language information, and other attribute
16   information, while accounting for privacy.

17   (3) Information on data governance procedures,
18   including how such training data was edited or fil-
19   tered.

20   (4) How such training data was labeled, and in-
21   formation regarding how the validity of the labeling
22   process was assessed.

23   (5) A description of the intended purposes and
24   foreseen limitations or risks of the foundation model,

1 an overview of past edits to such model, the version

2 of such model, and the date of release of such model.

3 (6) A description of the efforts of the covered

4 entity to align the foundation model and the trans-

5 parency of such model with—

6 (A) the AI Risk Management Framework

7 (or any successor framework) of the National

8 Institute of Standards and Technology; or

9 (B) a similar Federal Government-ap-

10 proved consensus technical standard.

11 (7) Performance under evaluation, either self-

12 driven or through audit, on public or industry stand-

13 ard benchmarks, including what precautions the

14 foundation model takes to answer or respond to situ-

15 ations with higher levels of risk of providing inac-

16 curate or harmful information, including, if such

17 model responds to such questions, relating to the fol-

18 lowing:

19 (A) Medical, health, or healthcare ques-

20 tions.

21 (B) Biological or chemical synthesis.

22 (C) Cybersecurity.

23 (D) Elections.

24 (E) Policing, including predictive policing.

25 (F) Financial loan decisions.

1   (G) Education.

2   (H) Employment or hiring decisions.

3   (I) Public services.

4   (J) Information relating to vulnerable pop-

5   ulations, including children and protected class-

6   es.

7   (8) Information on the computational power

8   used to train and operate the foundation model.

9   (9) Any other information determined necessary

10 by the Commission, in consultation with the Director

11 of the National Institute of Standards and Tech-

12 nology, to improve transparency of foundation mod-

13 els.

14 (g) CONSIDERATION OF ALTERNATIVE PROVISIONS

15 FOR SPECIFIC TYPES OF FOUNDATION MODELS.—In es-

16 tablishing the standards and issuing the guidance required

17 by subsection (a), the Commission shall consider whether

18 to include alternative provisions for—

19   (1) open-source foundation models; or

20   (2) foundation models that are a derivation of

21 or built upon another foundation model, having been

22 retrained or adapted from such other foundation

23 model to any extent.

24 (h) APPLICABILITY.—The regulations required by

25 subsection (a)(1) shall apply beginning on the date that

1 is 90 days after the date on which the Commission pro-

2 mulgates such regulations.

3   (i) UPDATES.—Not later than 2 years after the date

4 on which the Commission promulgates the regulations re-

5 quired by subsection (a)(1), and not less often than annu-

6 ally thereafter, the Commission, in consultation with the

7 Director of the National Institute of Standards and Tech-

8 nology, shall assess the standards established by such reg-

9 ulations and update such regulations so as to incorporate

10 appropriate updates (if any) to such standards.

11   (j) ENFORCEMENT BY FEDERAL TRADE COMMIS-

12 SION.—

13   (1) UNFAIR OR DECEPTIVE ACTS OR PRAC-

14   TICES.—A violation of a regulation promulgated

15   under subsection (a)(1) shall be treated as a viola-

16   tion of a regulation under section 18(a)(1)(B) of the

17   Federal Trade Commission Act (15 U.S.C.

18   57a(a)(1)(B)) regarding unfair or deceptive acts or

19   practices.

20   (2) POWERS OF COMMISSION.—Except as pro-

21   vided in subsection (m)(3)(C)—

22   (A) the Commission shall enforce the regu-

23   lations promulgated under subsection (a)(1) in

24   the same manner, by the same means, and with

25   the same jurisdiction, powers, and duties as

1 though all applicable terms and provisions of

2 the Federal Trade Commission Act (15 U.S.C.

3 41 et seq.) were incorporated into and made a

4 part of this section; and

5 (B) any covered entity that violates a regu-

6 lation promulgated under subsection (a)(1)

7 shall be subject to the penalties and entitled to

8 the privileges and immunities provided in the

9 Federal Trade Commission Act.

10 (k) REPORT.—Not later than 2 years after the date

11 of the enactment of this Act, the Commission shall submit

12 to the Committee on Energy and Commerce and the Com-

13 mittee on Science, Space, and Technology of the House

14 of Representatives and the Committee on Commerce,

15 Science, and Transportation of the Senate a report on the

16 establishment, implementation, and enforcement of the

17 standards required by subsection (a)(1).

18 (l) AUTHORIZATION OF APPROPRIATIONS.—There

19 are authorized to be appropriated to the Commission to

20 carry out this section—

21 (1) $10,000,000 for fiscal year 2025; and

22 (2) $3,000,000 for each fiscal year thereafter.

23 (m) DEFINITIONS.—In this section:

24 (1) ARTIFICIAL INTELLIGENCE.—The term "ar-

25 tificial intelligence" has the meaning given such

1 term in section 5002 of the National Artificial Intel-

2 ligence Initiative Act of 2020 (15 U.S.C. 9401; en-

3 acted as division E of the William M. (Mac) Thorn-

4 berry National Defense Authorization Act for Fiscal

5 Year 2021 (Public Law 116–283)).

6 (2) COMMISSION.—The term ''Commission''

7 means the Federal Trade Commission.

8 (3) COVERED ENTITY.—

9 (A) IN GENERAL.—The term ''covered en-

10 tity'' means any person, partnership, or cor-

11 poration described in subparagraph (C) that

12 provides—

13 (i) use of or services from a founda-

14 tion model which generates, in aggregate,

15 over 100,000 monthly output instances

16 (whether text, images, video, audio, or

17 other modality), including output instances

18 generated from use by users of second

19 party entities that use such model; or

20 (ii) use of or services from a founda-

21 tion model which has, in aggregate, over

22 30,000 monthly users, including users of

23 second party entities that use such model.

24 (B) UPDATING OF THRESHOLDS.—The

25 Commission, in consultation with the Director

1 of the National Institute of Standards and
2 Technology and the Director of the Office of
3 Science and Technology Policy, may, by regula-
4 tion promulgated in accordance with section
5 553 of title 5, United States Code, update the
6 number of monthly output instances for pur-
7 poses of subparagraph (A)(i) or the number of
8 monthly users for purposes of subparagraph
9 (A)(ii), as the Commission considers appro-
10 priate.

11 (C) PERSONS, PARTNERSHIPS, AND COR-
12 PORATIONS DESCRIBED.—The persons, partner-
13 ships, and corporations described in this sub-
14 paragraph are—

15 (i) any person, partnership, or cor-
16 poration over which the Commission has
17 jurisdiction under section 5(a)(2) of the
18 Federal Trade Commission Act (15 U.S.C.
19 45(a)(2)); and

20 (ii) notwithstanding section 4, 5(a)(2),
21 or 6 of the Federal Trade Commission Act
22 (15 U.S.C. 44; 45(a)(2); 46) or any juris-
23 dictional limitation of the Commission—

24 (I) any common carrier subject
25 to the Communications Act of 1934

1 (47 U.S.C. 151 et seq.) and all Acts

2 amendatory thereof and supple-

3 mentary thereto; and

4 (II) any organization not orga-

5 nized to carry on business for its own

6 profit or that of its members.

7 (4) FOUNDATION MODEL.—

8 (A) IN GENERAL.—The term "foundation

9 model" means an artificial intelligence model

10 that—

11 (i) is trained on broad data;

12 (ii) generally uses self-supervision;

13 (iii) generally contains at least

14 1,000,000,000 parameters;

15 (iv) is applicable across a wide range

16 of contexts; and

17 (v) exhibits, or could be easily modi-

18 fied to exhibit, high levels of performance

19 at tasks that could pose a serious risk to

20 security, national economic security, na-

21 tional public health or safety, or any com-

22 bination of those matters.

23 (B) EFFECT OF TECHNICAL SAFE-

24 GUARDS.—The term "foundation model" in-

25 cludes an artificial intelligence model otherwise

1   described in subparagraph (A) even if such

2   model is provided to users with technical safe-

3   guards that attempt to prevent users from tak-

4   ing advantage of any relevant unsafe capabili-

5   ties.

6   (5) INFERENCE.—The term "inference" means,

7   with respect to a foundation model, when such foun-

8   dation model is operated by a user to produce a re-

9   sult.

10   (6) TRAINING DATA.—The term "training

11   data" means, with respect to a foundation model,

12   the data on which such foundation model was

13   trained.