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.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE.

This Act may be cited as the “AI Foundation Model Transparency Act of 2023”.

SEC. 2. FINDINGS.

Congress finds the following:
(1) With the increase in public access to artificial intelligence, there has been an increase in lawsuits and public concerns about copyright infringement, including in court cases such as the following:


(2) Public use of foundation models has led to countless instances of the public being presented with inaccurate, imprecise, or biased information during inference, based on limited training data, limited model training mechanisms, or a lack of disclosures about the training data composition or foundation model training procedures, including in facial recognition technology usage, artificial intelligence inferences relating to health, artificial intelligence inferences relating to loan granting and housing approval, and more.
(3) Transparency with respect to high-impact foundation models has become increasingly necessary, including to assist copyright owners with enforcing their copyright protections and to promote consumer protection.

(4) While not compromising the intellectual property rights of those who develop and deploy foundation models, users should be equipped with the information necessary to enforce their copyright protections and to make informed decisions about such foundation models.

SEC. 3. FOUNDATION MODEL DATA SOURCE AND TRAINING TRANSPARENCY.

(a) Establishment of Standards.—Not later than 9 months after the date of the enactment of this Act, the Commission shall—

(1) in accordance with section 553 of title 5, United States Code, promulgate regulations that establish standards specifying information to improve the transparency of foundation models by covered entities with respect to training data, model documentation, data collection in inference, and operations of foundation models; and
(2) issue guidance to assist covered entities in complying with the standards established under paragraph (1).

(b) CONSULTATION.—In establishing the standards and issuing the guidance required by subsection (a), the Commission shall consult with the Director of the National Institute of Standards and Technology, the Director of the Office of Science and Technology Policy, the Register of Copyrights, and other relevant stakeholders, including standards bodies, covered entities, academia, technology experts, and advocates for civil rights and consumers.

(c) SUBMISSION TO COMMISSION AND PUBLIC AVAILABILITY OF INFORMATION.—The standards established under subsection (a)(1) shall include requirements, with respect to a foundation model, for—

(1) what information specified under such subsection shall be submitted to the Commission by the covered entity that provides such model; and

(2) what information specified under such subsection shall be made publicly available by the covered entity that provides such model.

(d) FORM AND MANNER.—The standards established under subsection (a)(1) shall specify the form and manner in which certain information specified under such sub-
A section, selected at the discretion of the Commission, in consultation with the Director of the National Institute of Standards and Technology and the other actors described in subsection (b), shall be made publicly available by covered entities, including—

(1) what information shall be made available on the website of a covered entity that relates to any foundation model provided by such covered entity;

(2) what information shall be displayed in a central location on a website hosted by the Commission, which shall include, with respect to a foundation model, information that is substantially similar to the information required under paragraph (1) to be made available on the website of the covered entity that provides such model;

(3) that a machine-readable format shall be used with respect to the information specified under paragraphs (1) and (2);

(4) the URL at which the information specified under paragraph (2) shall be hosted by the Commission; and

(5) such additional specifications as the Commission considers appropriate.

(e) PROCESS.—The standards established under subsection (a)(1) shall specify a process by which the informa-
tion required under subsection (c)(1) shall be submitted to the Commission.

(f) INFORMATION TO BE CONSIDERED.—The Commission shall consider specifying in the standards established under subsection (a)(1), with respect to a foundation model, the following information:

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

(2) A description of the size and composition of such training data, including broad demographic information, language information, and other attribute information, while accounting for privacy.

(3) Information on data governance procedures, including how such training data was edited or filtered.

(4) How such training data was labeled, and information regarding how the validity of the labeling process was assessed.

(5) A description of the intended purposes and foreseeable limitations or risks of the foundation model,
an overview of past edits to such model, the version
of such model, and the date of release of such model.

(6) A description of the efforts of the covered
entity to align the foundation model and the trans-
parency of such model with—

(A) the AI Risk Management Framework
(or any successor framework) of the National
Institute of Standards and Technology; or

(B) a similar Federal Government-ap-
proved consensus technical standard.

(7) Performance under evaluation, either self-
driven or through audit, on public or industry stand-
ard benchmarks, including what precautions the
foundation model takes to answer or respond to situ-
ations with higher levels of risk of providing inac-
curate or harmful information, including, if such
model responds to such questions, relating to the fol-
lowing:

(A) Medical, health, or healthcare ques-
tions.

(B) Biological or chemical synthesis.

(C) Cybersecurity.

(D) Elections.

(E) Policing, including predictive policing.

(F) Financial loan decisions.
(G) Education.

(H) Employment or hiring decisions.

(I) Public services.

(J) Information relating to vulnerable populations, including children and protected classes.

(8) Information on the computational power used to train and operate the foundation model.

(9) Any other information determined necessary by the Commission, in consultation with the Director of the National Institute of Standards and Technology, to improve transparency of foundation models.

(g) Consideration of Alternative Provisions for Specific Types of Foundation Models.—In establishing the standards and issuing the guidance required by subsection (a), the Commission shall consider whether to include alternative provisions for—

(1) open-source foundation models; or

(2) foundation models that are a derivation of or built upon another foundation model, having been retrained or adapted from such other foundation model to any extent.

(h) Applicability.—The regulations required by subsection (a)(1) shall apply beginning on the date that
is 90 days after the date on which the Commission promulgates such regulations.

(i) Updates.—Not later than 2 years after the date on which the Commission promulgates the regulations required by subsection (a)(1), and not less often than annually thereafter, the Commission, in consultation with the Director of the National Institute of Standards and Technology, shall assess the standards established by such regulations and update such regulations so as to incorporate appropriate updates (if any) to such standards.

(j) Enforcement by Federal Trade Commission.—

(1) Unfair or deceptive acts or practices.—A violation of a regulation promulgated under subsection (a)(1) shall be treated as a violation of a regulation under section 18(a)(1)(B) of the Federal Trade Commission Act (15 U.S.C. 57a(a)(1)(B)) regarding unfair or deceptive acts or practices.

(2) Powers of Commission.—Except as provided in subsection (m)(3)(C)—

(A) the Commission shall enforce the regulations promulgated under subsection (a)(1) in the same manner, by the same means, and with the same jurisdiction, powers, and duties as
though all applicable terms and provisions of
41 et seq.) were incorporated into and made a
part of this section; and

(B) any covered entity that violates a regu-
lation promulgated under subsection (a)(1)
shall be subject to the penalties and entitled to
the privileges and immunities provided in the

(k) REPORT.—Not later than 2 years after the date
of the enactment of this Act, the Commission shall submit
to the Committee on Energy and Commerce and the Com-
mittee on Science, Space, and Technology of the House
of Representatives and the Committee on Commerce,
Science, and Transportation of the Senate a report on the
establishment, implementation, and enforcement of the
standards required by subsection (a)(1).

(l) AUTHORIZATION OF APPROPRIATIONS.—There
are authorized to be appropriated to the Commission to
carry out this section—

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

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

(m) DEFINITIONS.—In this section:

(1) ARTIFICIAL INTELLIGENCE.—The term “ar-
tificial intelligence” has the meaning given such

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

(3) COVERED ENTITY.—

(A) IN GENERAL.—The term “covered entity” means any person, partnership, or corporation described in subparagraph (C) that provides—

(i) use of or services from a foundation model which generates, in aggregate, over 100,000 monthly output instances (whether text, images, video, audio, or other modality), including output instances generated from use by users of second party entities that use such model; or

(ii) use of or services from a foundation model which has, in aggregate, over 30,000 monthly users, including users of second party entities that use such model.

(B) UPDATING OF THRESHOLDS.—The Commission, in consultation with the Director
of the National Institute of Standards and Technology and the Director of the Office of Science and Technology Policy, may, by regulation promulgated in accordance with section 553 of title 5, United States Code, update the number of monthly output instances for purposes of subparagraph (A)(i) or the number of monthly users for purposes of subparagraph (A)(ii), as the Commission considers appropriate.

(C) PERSONS, PARTNERSHIPS, AND CORPORATIONS DESCRIBED.—The persons, partnerships, and corporations described in this subparagraph are—

(i) any person, partnership, or corporation over which the Commission has jurisdiction under section 5(a)(2) of the Federal Trade Commission Act (15 U.S.C. 45(a)(2)); and

(ii) notwithstanding section 4, 5(a)(2), or 6 of the Federal Trade Commission Act (15 U.S.C. 44; 45(a)(2); 46) or any jurisdictional limitation of the Commission—

(I) any common carrier subject to the Communications Act of 1934
(47 U.S.C. 151 et seq.) and all Acts amendatory thereof and supplementary thereto; and

(II) any organization not organized to carry on business for its own profit or that of its members.

(4) FOUNDATION MODEL.—

(A) IN GENERAL.—The term “foundation model” means an artificial intelligence model that—

(i) is trained on broad data;

(ii) generally uses self-supervision;

(iii) generally contains at least 1,000,000,000 parameters;

(iv) is applicable across a wide range of contexts; and

(v) exhibits, or could be easily modified to exhibit, high levels of performance at tasks that could pose a serious risk to security, national economic security, national public health or safety, or any combination of those matters.

(B) EFFECT OF TECHNICAL SAFEGUARDS.—The term “foundation model” includes an artificial intelligence model otherwise
described in subparagraph (A) even if such model is provided to users with technical safeguards that attempt to prevent users from taking advantage of any relevant unsafe capabilities.

(5) INFERENCEN.——The term “inference” means, with respect to a foundation model, when such foundation model is operated by a user to produce a result.

(6) TRAINING DATA.——The term “training data” means, with respect to a foundation model, the data on which such foundation model was trained.