H. R. 118TH CONGRESS 2D SESSION

To require the National Institute of Standards and Technology to establish task forces to facilitate and inform the development of technical standards and guidelines relating to the identification of content created by generative artificial intelligence, to ensure that audio or visual content created or substantially modified by generative artificial intelligence includes a disclosure acknowledging the generative artificial intelligence origin of such content, and for other purposes.

IN THE HOUSE OF REPRESENTATIVES

Ms. Eshoo introduced the following bill; which was referred to the Committee on

A BILL

To require the National Institute of Standards and Technology to establish task forces to facilitate and inform the development of technical standards and guidelines relating to the identification of content created by generative artificial intelligence, to ensure that audio or visual content created or substantially modified by generative artificial intelligence includes a disclosure acknowledging the generative artificial intelligence origin of such content, and for other purposes.
Be it enacted by the Senate and House of Representa-
tives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE.

This Act may be cited as the “Protecting Consumers
from Deceptive AI Act”.

SEC. 2. FINDINGS.

This Congress finds the following:

(1) The majority of Americans consume most of
their information online from social media platforms.
A 2023 Pew Research survey found that a large ma-
majority of U.S. adults (86%) say they often or some-
times get news from a smartphone, computer or tab-
let, including 56% who say they do so often.

(2) The increasing capabilities of generative ar-
tificial intelligence models has led to a marked in-
crease in the creation of convincing “deepfakes” and
greater difficulty for everyday Americans in telling
real and deepfake images, audio, and videos apart.
A December 2022 study found participants were
only 62% accurate when asked to determine whether
images were deepfake or real, and even worse, that
their self-reported confidence in their answers was
high, and unrelated to accuracy.

(3) Deepfakes create consumer deception issues,
where persons can create “deepfake” images and
videos to fool consumers about information related to products they may wish to purchase. Fake celebrity endorsements of various products and scams have proliferated in the past year, including an ad in which a deepfake of famous actor Tom Hanks endorsed a dental insurance plan.

(4) The proliferation of deepfakes can also create national security issues, such as a deepfake image of an explosion at the Pentagon that was shared widely last year and caused enough confusion that the stock market briefly dipped.

(5) Deepfakes used in political advertising can also create confusion, such as when someone used an AI model that had been trained to replicate President Biden’s voice was used to make robocalls to voters in New Hampshire ahead of a primary election, and provided false information intended to discourage potential voters from voting in the election. As the Supreme Court found 8-1 in Citizens United v. Federal Election Commission, 558 U.S. 310 (2010), the government has an interest in “‘insur[ing] that the voters are fully informed’ about the person or group who is speaking”.

(6) Requiring deepfakes to be clearly labeled is important to protect consumers from deception, pro-
tect our national security, and to maintain an in-
formed electorate.

SEC. 2. GUIDELINES TO FACILITATE DISTINGUISHING CON-
TENT GENERATED BY GENERATIVE ARTIFI-
CIAL INTELLIGENCE.

(a) Task Forces for Development of Guide-
lines and Promoting Standards.—

(1) In general.—Not later than 90 days after
the date of the enactment of this Act, the Director
of the National Institute of Standards and Tech-
nology shall establish task forces to accomplish the
following goals:

(A) Supporting the development of tech-
ical standards and guidelines to provide con-
tent provenance metadata, watermarking, dig-
ital fingerprinting for audio or visual content,
and other technical measures that the task
forces determine significant. To the extent tech-
ically feasible, such task forces should seek to
make content provenance metadata crypt-
tographically verifiable, and to make water-
marks difficult to remove or obscure.

(B) Supporting the development of tech-
ical standards and guidelines to assist online
application and content providers and operators
in identifying and labeling audio or visual content created or substantially modified by generative artificial intelligence, including exploring interoperable standards that assist social media and other online platforms with identifying, maintaining, interpreting, and displaying watermarks, digital fingerprinting, and secure content provenance metadata associated with audio or visual content, while considering circumvention techniques and enforcement.

(C) Supporting the development of technical standards and guidelines to identify and label text-based content created or substantially modified by generative artificial intelligence. Such support may include developing standards to embed content provenance data or metadata, watermarking, digital fingerprinting, or other technical measures when creating such content.

(2) STANDARDS BODIES.—To the extent possible, the outcome and output of the task forces established pursuant to paragraph (1) should inform development of technical standards developed by private, consensus organizations, as referred to in section 2 of the National Institute of Standards and

(3) Membership.—The Director of the National Institute of Standards and Technology shall include in the memberships of each of the task forces described in paragraph (1) appropriate representatives of the following:

(A) Relevant Federal agencies.

(B) Developers of generative artificial intelligence technology.

(C) Entities, including standards development organizations, engaged in the development of content detection standards and technology, including authentication and traceability.

(D) Social networking service providers and online instant messaging service providers.

(E) Online search engine service providers.

(F) Developers of web browsers and mobile operating systems.

(G) Academic entities, civil society and advocacy groups, and other related entities, especially such entities and groups engaged in the development or implementation of content detection standards and technology.

(H) Privacy advocates and experts.
(I) Human rights lawyers and advocates with expertise in the effects of technology in countries around the world.

(J) Media organizations, including news publishers and image providers.

(K) Creator associations and organizations representing the interests of other copyright owners.

(L) Artificial intelligence testing experts, such as those with privacy expertise in artificial intelligence red-teaming.

(M) Technical experts in digital forensics, cryptography, and secure digital content and delivery.

(N) Any other entity the Director determines appropriate.

(4) DUTIES.—

(A) SUBMISSION TO DIRECTOR.—Each of the task forces established pursuant to paragraph (1) shall, not later than 270 days after the establishment of each such task force, submit to the Director of the National Institute of Standards and Technology a report containing recommendations relating to the technical
standards and guidelines each such task force is
supporting.

(B) SUBMISSION TO CONGRESS.—Each of
the task forces established pursuant to para-
graph (1) shall, not later than one year after
the establishment of each such task force and
annually thereafter for five years, submit to the
Committee on Science, Space, and Technology
and the Committee on Energy and Commerce
of the House of Representatives and the Com-
mittee on Commerce, Science, and Transpor-
tation of the Senate a report on the activities
of such task force for the immediately preceding
one year period.

(5) PRIVACY.—The task forces established pur-
suant to paragraph (1) shall consider issuing guid-
ance for online service and application providers and
operators to store and display content provenance
data and metadata in a privacy-preserving manner,
including clear guidance on how such providers and
operators can indicate to users when such users are
sharing content that contains content provenance
data and metadata, indicate the information con-
tained in the data and metadata such users are
sharing, and provide options to limit the data and
metadata such users are sharing that may have privacy implications.

(b) **Informing Consumers of Content Generated by Artificial Intelligence.**—

(1) **Providers of Generative Artificial Intelligence Applications.**—A person who makes available to users a software application based on generative artificial intelligence technology shall—

(A) ensure that audio or visual content created or substantially modified by such application incorporates (as part of such content and in a manner that may or may not be perceptible by unaided human senses) a disclosure that—

(i) is machine-readable; and

(ii) acknowledges the generative artificial intelligence origin of such content;

(B) establish and implement reasonable measures to prevent a disclosure described in subparagraph (A) from being removed or otherwise tampered with;

(C) collaborate with providers of covered online platforms to assist such providers in identifying and accessing the information of disclosures described in subparagraph (A); and
(D) ensure that such application makes available to users the ability to incorporate, within the metadata of content created or modified by such application, information regarding the generative artificial intelligence origin of such content, including tamper-evident information regarding—

(i) the name of such application;

(ii) the name and version of the generative artificial intelligence model utilized by such application to create or modify such content;

(iii) the date and time associated with the creation or modification of such content by such application; and

(iv) the portion of such content that was created or modified by such application.

(2) PROVIDERS OF COVERED ONLINE PLATFORMS.—A person who makes available for use a covered online platform—

(A) shall clearly and conspicuously provide to a user of such platform, with respect to audio or visual content accessed by such user through such platform that incorporates a dis-
closure described in paragraph (1)(A), the information included in such disclosure; and

(B) may not, with respect to audio or visual content accessed by such user through such platform that incorporates a disclosure described in paragraph (1)(A), remove such disclosure or any incorporated information described in paragraph (1)(D).

(3) REGULATIONS.—

(A) IN GENERAL.—Not later than 2 years after the date of the enactment of this Act, the Commission shall promulgate regulations under section 553 of title 5, United States Code, to carry out this subsection.

(B) CONSULTATION.—In carrying out subparagraph (A), the Commission shall consult with the National Institute of Standards and Technology and the task forces established under subsection (a)(1).

(4) ENFORCEMENT BY COMMISSION.—

(A) UNFAIR OR DECEPTIVE ACTS OR PRACTICES.—A violation of this subsection or a regulation promulgated under this subsection shall be treated as a violation of a regulation under section 18(a)(1)(B) of the Federal Trade Com-
mission Act (15 U.S.C. 57a(a)(1)(B)) regarding unfair or deceptive acts or practices.

(B) POWERS OF COMMISSION.—The Commission shall enforce this subsection and the regulations promulgated under this subsection in the same manner, by the same means, and with the same jurisdiction, powers, and duties as though all applicable terms and provisions of the Federal Trade Commission Act (15 U.S.C. 41 et seq.) were incorporated into and made a part of this Act. Any person who violates such subsection or a regulation promulgated under such subsection shall be subject to the penalties and entitled to the privileges and immunities provided in the Federal Trade Commission Act.

(C) AUTHORITY PRESERVED.—Nothing in this subsection may be construed to limit the authority of the Commission under any other provision of law.

(5) EFFECTIVE DATE.—Paragraphs (1) and (2) of this subsection shall take effect on the date that is 90 days after the date on which the regulations promulgated under paragraph (3) take effect.

(6) SAFE HARBORS.—
(A) IN GENERAL.—A person who makes available for use a generative artificial intelligence application or a covered online platform may satisfy the requirements of this subsection (including regulations promulgated under this subsection) by following self-regulatory guidelines that are approved by the Commission under subparagraph (B).

(B) SELF-REGULATORY GUIDELINES.—

(i) INCENTIVES.—In promulgating regulations under this subsection, the Commission may provide incentives for self-regulation.

(ii) DEEMED COMPLIANCE.—Incentives described in clause (i) shall include provisions for ensuring that a person will be deemed to be in compliance with the requirements of this subsection (including regulations promulgated under this subsection) if that person complies with guidelines that, after provision of notice and an opportunity for comment, are approved by the Commission upon a determination that such guidelines satisfy the requirements of
this subsection (including regulations promulgated under this subsection).

(iii) Expedited Response to Requests.—The Commission shall act upon a request for approval of guidelines under this paragraph not later than 180 days after the date on which such request is filed and shall set forth in writing conclusions with regard to such request.

(C) Appeals.—Final action by the Commission on a request for approval of guidelines under this paragraph, or the failure to act within the time period described in subparagraph (B)(iii), may be appealed to a district court of the United States of appropriate jurisdiction as provided for in section 706 of title 5, United States Code.

(7) Privacy and Interoperability.—The Commission shall consider privacy concerns and the interoperability of standards when promulgating regulations under paragraph (3) and considering the approval of guidelines under paragraph (6).

(e) Definitions.—In this section:
(1) **Audio or visual content.**—The term “audio or visual content” means content in the form of a digital image, a video, or audio.

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

(3) **Content provenance.**—The term “content provenance” means the chronology of the origin and history associated with digital content.

(4) **Covered online platform.**—The term “covered online platform” means a website, internet application, or mobile application available to users in the United States, including a social networking site, video sharing service, search engine, or content aggregation service available to users in the United States, that—

   (A) generates at least $50,000,000 in annual revenue; or

   (B) had at least 25,000,000 monthly active users for not fewer than 3 of the preceding 12 months.

(5) **Digital fingerprinting.**—The term “digital fingerprinting” means the process by which an identifier is derived from a piece of digital content and stored in a database, for the purpose of
identifying, matching against, or verifying such content, or similar content, at a later date.

(6) GENERATIVE ARTIFICIAL INTELLIGENCE.—The term “generative artificial intelligence” means the class of models and algorithms that use deep learning algorithms or other statistical techniques to generate new data that has similar characteristics and properties to the data with respect to which such models and algorithms have been trained, including any form of digital content.

(7) MACHINE-READABLE.—The term “machine-readable” has the meaning given such term in section 3502 of title 44, United States Code.

(8) METADATA.—The term “metadata” has the meaning given such term in section 3502 of title 44, United States Code.

(9) WATERMARKING.—The term “watermarking” means the act of embedding tamper-resistant information into digital content (perceptibly or imperceptibly) which may be used to establish some aspect or aspects of the content provenance of the content or to store reference information.