.....

(Original Signature of Member)

118TH CONGRESS
2D SESSION

H. R. _____

To direct the National Institute of Standards and Technology to catalog and evaluate emerging practices and norms for communicating certain characteristics of artificial intelligence systems, including relating to transparency, robustness, resilience, security, safety, and usability, and for other purposes.

IN THE HOUSE OF REPRESENTATIVES

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

A BILL

To direct the National Institute of Standards and Technology to catalog and evaluate emerging practices and norms for communicating certain characteristics of artificial intelligence systems, including relating to transparency, robustness, resilience, security, safety, and usability, 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 Development Practices Act of 2024''.

SEC. 2. NIST RESEARCH ON DEVELOPMENT BEST PRACTICES.

Section 22A of the National Institute of Standards and Technology Act (15 U.S.C. 278h–1) is amended—

(1) by redesignating subsection (h) as subsection (i); and

(2) by inserting after subsection (g) the following new subsection:

''(h) ASSESSMENT OF THE PRACTICES OF ARTIFICIAL INTELLIGENCE DEVELOPMENT.—

''(1) IN GENERAL.—The Director of the National Institute of Standards and Technology (in this subsection referred to as the 'Director') shall, subject to the availability of appropriations, develop, and periodically update, in collaboration with other public and private sector organizations, voluntary guidance for practices and guidelines relating to the development, release, and assessment of artificial intelligence systems. Such guidelines shall satisfy the following:

''(A) Define methods and guidelines for developing reasonable risk tolerances for various

use cases of artificial intelligence systems based on the following:

''(i) The risks associated with the intended and unintended applications, use cases, and outcomes of the artificial intelligence system at issue, based on the guidelines specified in the voluntary risk management framework for trustworthy artificial intelligence systems, or successor framework, authorized under subsection (c), which may include different categories of risk, such as the following:

''(I) Security risks, including threats to national security.

''(II) Economic risks, including threats to economic opportunities.

''(III) Social risks, including infringement upon constitutional rights, privileges, or liberties.

''(ii) Such other factors as the Director determines appropriate and consistent with this subsection.

''(B) Categorize and list practices and norms for communicating relevant characteristics, including robustness, resilience, security,

safety, fairness, privacy, validation, reliability, accountability, and usability, of artificial intelligence systems, and including any characteristics identified by the voluntary risk management framework for trustworthy artificial intelligence systems, or successor framework, authorized under subsection (c). Such practices and norms may relate to the following:

''(i) Documentation of training and evaluation datasets, such as information and statistics about a dataset's size, curation, annotation, and sources, and the protocols for a dataset's selection, creators, provenance, processing, augmentation, filters, inclusion of personally identifiable information, and intellectual property usage.

''(ii) Documentation of model information, such as a model's development stages, training objectives, training strategies, inference objectives, capabilities, reproducibility of capabilities, input and output modalities, components, size, and architecture.

''(iii) Evaluation of benchmarks for multi-metric assessments, such as an as-

sessment of an appropriate combination of robustness, resilience, security, safety, fairness, privacy, accuracy, validity, reliability, accountability, usability, transparency, efficiency, and calibration, and any characteristics identified by the voluntary risk management framework for trustworthy artificial intelligence systems, or successor framework, authorized under subsection (c).

''(iv) Metrics and methodologies for evaluations of artificial intelligence systems, such as establishing evaluation datasets.

''(v) Public reporting of artificial intelligence systems' capabilities, limitations, and possible areas of appropriate and inappropriate use.

''(vi) Disclosure of security practices, such as artificial intelligence red teaming and third-party assessments, that were used in the development of an artificial intelligence system.

''(vii) How to release to the public components of an artificial intelligence sys-

tem or information about an artificial intelligence system, including aspects of the model, associated training data, and license agreements.

''(viii) Approaches and channels for collaboration and knowledge-sharing of best practices across industry, governments, civil society, and academia.

''(ix) Such other categories as the Director determines appropriate and consistent with this subsection.

''(C) For each practice and norm categorized and listed in accordance with subparagraph (B), provide recommendations and practices for utilizing such practice or norm.

''(2) IMPLEMENTATION.—In conducting the Director's duties under paragraph (1), the Director shall carry out the following:

''(A) Update the voluntary risk management framework for trustworthy artificial intelligence systems, or successor framework, authorized under subsection (c) as the Director determines appropriate.

''(B) Ensure that voluntary guidance developed in paragraph (1) is based on inter-

national standards and industry best practices to the extent possible and practical.

''(C) Not prescribe or otherwise require the use of specific information or communications technology products or services.

''(D) Collaborate with public, industry, and academic entities as the Director determines appropriate, including conducting periodic outreach to receive public input from public, industry, and academic stakeholders.

''(3) REPORT.—In conducting the Director's duties under paragraph (1), the Director shall, not later than 18 months after the date of the enactment of this subsection, brief the Committee on Science, Space, and Technology of the House of Representatives and the Committee on Commerce, Science, and Transportation of the Senate on the following:

''(A) New or updated materials, programs, or systems that were produced as a result of carrying out this subsection.

''(B) Policy recommendations of the Director that could facilitate and improve communication and coordination between the private sector and relevant Federal agencies regarding

implementing the recommended practices identified in this subsection.

''(4) DEFINITIONS.—In this subsection:

''(A) ARTIFICIAL INTELLIGENCE RED TEAMING.—The term 'artificial intelligence red teaming' means a structured testing of adversarial efforts to find flaws and vulnerabilities in an artificial intelligence system and identify risks, flaws, and vulnerabilities of artificial intelligence systems, such as harmful outputs from such system, unforeseen or undesirable system behaviors, limitations, and potential risks associated with the misuse of such system.

''(B) ARTIFICIAL INTELLIGENCE SYSTEM.—The term 'artificial intelligence system' has the meaning given such term in section 7223 of the Advancing American AI Act (40 U.S.C. 11301 note; as enacted as part of title LXXII of division G of the James M. Inhofe National Defense Authorization Act for Fiscal Year 2023; Public Law 117–263).''.