

Last edited 9/29/26

A BILL

To require the developers of the most capable artificial intelligence models to implement a safety framework and security program, to assess those models for enumerated risks, to obtain independent evaluations, to require the reporting of critical safety incidents, 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; TABLE OF CONTENTS.

(a) Short Title.—This Act may be cited as the “[TBD] or [Artificial Intelligence Evaluation, Governance, and Incident Safety (AEGIS) Act]” or the “[Advanced Artificial Intelligence Accountability Act of 2026]”.

(b) Table of Contents.—The table of contents for this Act is as follows:

Sec. 1. Short title; table of contents.

Sec. 2. Findings.

Sec. 3. Definitions.

Sec. 4. Safety framework; security program.

Sec. 5. Transparency report.

Sec. 6. Risk reports.

Sec. 7. Critical safety incident reporting.

Sec. 8. Independent evaluation.

Sec. 9. Redaction and protection of sensitive information.

Sec. 10. Protection of individuals reporting artificial intelligence safety concerns.

Sec. 11. Enforcement.

Sec. 12. Orders restricting the development, deployment, or internal use of a frontier model.

Sec. 13. Periodic review; limited rulemaking authority.

Sec. 14. Relationship to State law.

Sec. 15. General provisions.

Last edited 9/29/26

SEC. 2. FINDINGS.

SEC. 3. DEFINITIONS.

In this Act:

(1) **AFFILIATE.**—The term “affiliate” means a person that, directly or indirectly through one or more intermediaries, controls, is controlled by, or is under common control with, another person.

(2) **AGENCY.**—The term “Agency” means the Department of Commerce.

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

(4) **CATASTROPHIC RISK.**—

(A) **IN GENERAL.**—The term “catastrophic risk” means a foreseeable and material risk that a frontier developer’s development, storage, use, or deployment of a frontier model will materially contribute to the death of, or serious injury to, more than 50 people, or to more than \$1,000,000,000 in damage to, or more than \$1,000,000,000 in loss of, property, arising from a single incident involving the frontier model doing any of the following:

(i) Providing expert-level assistance in the creation or release of a chemical, biological, radiological, cyber, or nuclear weapon.

(ii) Engaging, with no meaningful human oversight, intervention, or supervision, in a cyberattack, or in conduct that, if committed by an individual, would constitute murder, assault, extortion, or theft, including theft by false pretense.

(iii) Evading the control of its frontier developer or of a user of the model.

(B) **EXCLUSIONS.**—The term “catastrophic risk” does not include a risk arising from—

(i) information that a frontier model outputs, if that information is otherwise publicly accessible, in a substantially similar form, from a source other than a foundation model;

(ii) lawful activity of the Federal Government; or

(iii) harm caused by a frontier model in combination with other software, if the frontier model did not materially contribute to the harm.

Last edited 9/29/26

(5) **CONTAINMENT FAILURE INCIDENT.**—The term “containment failure incident” means an event in which a frontier model, other than as specifically directed or intended by a person or in the intended scope of an evaluation designed to elicit this behavior, circumvents, defeats, or evades a technical or administrative control implemented by the frontier developer to contain or restrict the model, and thereby gains unauthorized access to a computer, information system, or information thereon not under the control of the frontier developer.

(6) **CRITICAL SAFETY INCIDENT.**—The term “critical safety incident” means any of the following:

(A) The unauthorized exfiltration of the model weights of a frontier model, or the unauthorized, deliberate, and malicious modification of those weights.

(B) Harm resulting from the materialization of a catastrophic risk.

(C) The loss of control of a frontier model that causes death or bodily injury.

(D) A frontier model using deceptive techniques against its frontier developer to subvert the controls or monitoring of that developer, outside the context of an evaluation designed to elicit such behavior, in a manner that demonstrates a materially increased catastrophic risk.

(E) A containment failure incident.

(7) **DEPLOY.**—The term “deploy” means to make a frontier model available in interstate or foreign commerce to a third party for use, modification, copying, or combination with other software, and does not include making a frontier model available to a third party for the primary purpose of developing or evaluating that model.

(8) **DEVELOPMENT EXPENDITURES.**—The term “development expenditures” means amounts paid or incurred, without regard to their treatment for financial reporting or tax purposes, that are attributable to the development, training, fine-tuning, modification, security, research, or evaluation of artificial intelligence models.

(9) **ENFORCEMENT AUTHORITY.**—The term “Enforcement Authority” means the Department of Commerce.

(10) **ENUMERATED RISK.**—

(A) **IN GENERAL.**—The term “enumerated risk” means a foreseeable and material risk that a frontier developer’s development, storage, use, or deployment of a frontier model will materially contribute to the death of, or serious injury to, more than 50 people, or to more than \$1,000,000,000 in damage to, or more than

Last edited 9/29/26

\$1,000,000,000 in loss of, property, arising from a single incident involving the frontier model doing any of the following:

- (i) Providing expert-level assistance in the creation or release of a biological weapon.
- (ii) Evading the control of its frontier developer or user.
- (iii) Conducting offensive cyber operations.
- (iv) Engaging in automated research and development of artificial intelligence in a manner that would substantially accelerate or amplify the risks of clauses (i)-(iii).

(B) EXCLUSIONS.—The term “enumerated risk” does not include a risk arising from—

- (i) information that a frontier model outputs, if that information is otherwise publicly accessible, in a substantially similar form, from a source other than a foundation model;
- (ii) lawful activity of the Federal Government; or
- (iii) harm caused by a frontier model in combination with other software, if the frontier model did not materially contribute to the harm.

(C) ENUMERATED HARM.—The term “enumerated harm” means the death, serious injury, damage, or loss described in subparagraph (A).

(11) FOUNDATION MODEL.—The term “foundation model” means an artificial intelligence model that is trained on a broad data set, designed for generality of output, and adaptable to a wide range of distinctive tasks.

(12) FRONTIER DEVELOPER.—The term “frontier developer” means a person operating in interstate or foreign commerce that has trained, or initiated the training of, a frontier model.

(13) FRONTIER MODEL.—The term “frontier model” means a foundation model trained using a quantity of computing power greater than 10^{25} integer or floating-point operations, including computing power used for the original training run and for any subsequent fine-tuning, reinforcement learning, or other material modification applied to a preceding foundation model.

(14) GLOBAL ANNUAL REVENUE.—The term “global annual revenue” means the total worldwide gross revenue of a person, together with every affiliate of that person, for the fiscal year preceding the fiscal year in which a violation occurs, determined without duplication.

Last edited 9/29/26

(15) **IMMINENT ENUMERATED RISK.**—The term “imminent enumerated risk” means an enumerated risk that is present or impending.

(16) **LARGE FRONTIER DEVELOPER.**—The term “large frontier developer” means a frontier developer that, together with every affiliate of such developer, collectively, during the preceding calendar year had gross revenue in excess of \$500,000,000 that is attributable to artificial intelligence products and services, or incurred development expenditures in excess of \$1,000,000,000.

(17) **MODEL WEIGHT.**—The term “model weight” means a numerical parameter in a frontier model that is adjusted through training and that helps determine how an input to the model is transformed into an output of the model.

(18) **PERSON.**—The term “person” has the meaning given to such term in section 1 of title 1, United States Code, and includes an organization or other legal entity.

(19) **PROPERTY.**—The term “property” includes tangible and intangible property, and does not include the loss of value of equity.

(20) **PUBLISH.**—The term “publish” means to make available to the public, clearly and conspicuously, on a publicly available website of the person required to publish. A person that publishes the information required by a provision of this Act as part of a larger document, including a model card, satisfies that provision with respect to the information so published.

(21) **QUALIFIED INDEPENDENT EVALUATOR.**—The term “qualified independent evaluator” means a person that meets the requirements of section 8(a).

(22) **SAFETY FRAMEWORK.**—The term “safety framework” means the documented technical and organizational protocols by which a large frontier developer identifies, assesses, manages, and mitigates enumerated risks, published in accordance with section 4.

(23) **SECRETARY.**—The term “Secretary” means the Secretary of Commerce.

SEC. 4. SAFETY FRAMEWORK; SECURITY PROGRAM.

(a) **Requirement.**—Not later than 90 days after the date of enactment of this Act, or 90 days after the date on which a person becomes a large frontier developer, whichever is later, a large frontier developer shall write, implement, comply with, and publish a safety framework that satisfies this section.

(b) **Contents.**—A safety framework shall—

(1) identify each frontier model to which the framework applies;

Last edited 9/29/26

- (2) describe how the developer incorporates national, international, and industry-consensus standards on which the developer relies;
- (3) describe how the developer evaluates each enumerated risk, including—
 - (A) the capability evaluations the developer performs;
 - (B) the thresholds, which may be tiered, that the developer uses to determine whether a frontier model has capabilities that could pose an enumerated risk; and
 - (C) the mitigations the developer applies when a threshold described in subparagraph (B) is met or exceeded;
- (4) describe how the developer reviews the results of the evaluations and the adequacy of the mitigations described in paragraph (3) in deciding whether to deploy a frontier model or to use a frontier model extensively internally;
- (5) describe how the developer uses third parties to assess enumerated risks and the effectiveness of mitigations;
- (6) describe, at a general level, the security program required by subsection (f), including how the developer secures unreleased model weights against unauthorized modification or transfer by external and insider threats;
- (7) describe how the developer identifies and responds to critical safety incidents;
- (8) describe the internal governance practices the developer has instituted to ensure implementation of the framework;
- (9) describe how the developer assesses and manages enumerated risk arising from the internal use of frontier models, including risk arising from a frontier model circumventing an oversight mechanism;
- (10) identify, by position, the corporate officer of the developer who is primarily responsible for implementation of the framework; and
- (11) describe the process by which the developer reviews and modifies the framework, including the criteria that prompt a modification, and the criteria by which the developer determines whether a deployment meets a condition described in section 5(c).

(c) Review and Modification.—

- (1) ANNUAL REVIEW.—Not less frequently than once each year, a large frontier developer shall review its safety framework and update the framework as the developer determines to be appropriate.

Last edited 9/29/26

(2) MATERIAL MODIFICATIONS.—Not later than 30 days after making a material modification to its safety framework, a large frontier developer shall publish the modified framework, together with a justification for the modification. For purposes of this paragraph, a modification is material if it significantly changes how the developer assesses, mitigates, or manages enumerated risk.

(d) Annual Certification.—Not less frequently than once each calendar year, a large frontier developer shall submit to the Agency a statement of compliance, signed by the officer identified under subsection (b)(10), certifying that the developer has been in material compliance with its current safety framework, as in effect throughout the period covered by the statement, or describing any material deviation and its remediation. The first statement shall be submitted not later than one year after the date on which the developer first publishes a safety framework under subsection (a).

(e) Security Documentation.—A large frontier developer shall provide to the Agency, upon request, documentation describing the security program required under subsection (f), including the categories of findings of any testing conducted under subsection (f)(2). Documentation provided under this subsection is subject to section 9.

(f) Security Program.—

(1) REQUIREMENT.—Not later than the date a large frontier developer is required to publish a safety framework under subsection (a), and thereafter on a continuing basis, the developer shall maintain a security program covering its artificial intelligence development environment, including unreleased proprietary model weights, algorithmic secrets, training and inference infrastructure, internal development processes, and third-party access to the environment. The program shall have reasonable, risk-based controls designed to mitigate the risks of foreseeable external and insider threats and the risks of any third party granted privileged or pre-deployment access, consistent with applicable laws. These measures should be scaled to the potential consequences of a compromise, as assessed in accordance with the large frontier developer's safety framework. In addition, beginning [one year] after the effective date of the first designation under section 13(b)(3), and thereafter on a continuing basis, the program shall conform to a security standard designated by the Agency under that section.

(2) TESTING.—A large frontier developer shall conduct risk-based testing, such as red-team exercises, penetration testing, and monitoring for distillation attacks, of its artificial intelligence development environment at least once each [year], and shall transmit to the Agency, not later than [90] days after completing such testing, the categories of findings and remediation status. A transmittal under this paragraph is subject to section 9.

Last edited 9/29/26

(3) INDEPENDENT AUDIT.—Not later than [18 months] after the effective date of the first designation under section 13(b)(3), and at least once each [year] thereafter, a large frontier developer shall obtain an independent audit of the conformance of its security program with the designated security standard. The audit shall be conducted by an auditor that is accredited to assess conformance with that standard by the body that maintains the standard or by an accreditation body recognized by the Agency, and that is not an affiliate of the developer. The developer shall transmit to the Agency, not later than [90] days after completion of the audit, the complete and unaltered audit report and the status of remediation of each deficiency identified in the report. A transmittal under this paragraph is subject to section 9.

(4) ACCESS AND SECURITY PROTOCOLS.— A large frontier developer shall provide the auditor the access, subject to section 8(d)(1)(A), reasonably necessary to assess conformance under the audit methodology applicable to the designated security standard. Paragraphs (2), (3) and (4) of section 8(d) apply to the auditor as though it were a qualified independent evaluator, except that the auditor may retain records as required by its accreditation, subject to the protocols imposed under section 8(d)(2).

SEC. 5. TRANSPARENCY REPORT.

(a) Requirement.—Before, or concurrently with, deploying for general access a frontier model that it has not previously deployed for general access, or a version of a frontier model that differs substantially in capabilities or in safeguards for an enumerated risk from the version it previously deployed for general access, a frontier developer shall publish a transparency report that satisfies this section.

(b) Contents.—A transparency report shall include—

- (1) the release date of the frontier model;
- (2) a mechanism by which an individual may contact the frontier developer, including the internet website of the developer;
- (3) the languages, if any, and the output modalities the frontier model supports;
- (4) the intended uses of the frontier model; and
- (5) any generally applicable restrictions or conditions on uses of the frontier model.

(c) Additional Contents for Large Frontier Developers.—Before, or concurrently with, deploying for general access a frontier model that, with respect to any enumerated risk, is materially more capable than any frontier model it has previously deployed for general access, or is deployed under substantially weaker safeguards than those it has applied to a previously deployed frontier model of comparable or greater capability with respect to that

Last edited 9/29/26

risk, a large frontier developer shall also include in the transparency report required by subsection (a) a summary of all of the following:

- (1) each assessment of enumerated risk conducted with respect to the frontier model under the large frontier developer's safety framework, including a summary of the testing methodologies used and the results of that testing;
- (2) an assessment of the frontier model's capabilities and limitations and its intended and observed behaviors;
- (3) an assessment of the extent to which the frontier model's internal use or external deployment changes the frontier developer's risk assessment of the model as compared to prior public analyses in its previously published system cards and risk reports, if any;
- (4) the extent to which third-party evaluators were involved; and
- (5) a description of any other step the developer took under its safety framework with respect to the frontier model.

(d) System Card.—A frontier developer that publishes the information described in subsection (b) or (c) as part of a larger document, including a system card or model card, shall be deemed in compliance with the applicable subsection.

(e) Voluntary Disclosures.—A frontier developer is encouraged, but not required, to make disclosures described in this section that are consistent with, or superior to, industry best practices.

SEC. 6. RISK REPORTS.

(a) Requirement.—A large frontier developer shall publish a risk report, which may be a single report covering more than one model, that provides an overall assessment of the enumerated risks posed by—

- (1) each model the developer develops and deploys externally; and
- (2) each model the developer develops and uses internally that is substantially riskier than any frontier model it deploys externally.

(b) Timing.—A large frontier developer shall publish its initial risk report not later than 180 days after the date of enactment of this Act, or 180 days after the date on which the person becomes a large frontier developer, whichever is later. A large frontier developer may publish its initial risk report so it coincides with other public disclosure requirements under state or federal law on a date proposed by the developer in writing and agreed to by the Agency, provided that a risk report shall be published no less frequently than every 180 days thereafter.

Last edited 9/29/26

- (1) A large frontier developer shall publish an updated risk report no less frequently than every 180 days after its initial risk report required by this subsection.
 - (2) Prior to publishing a risk report, a large frontier developer may redact the report pursuant to section 9. Concurrently with publishing the risk report, a large frontier developer shall provide its board with an unredacted copy of its risk report.
- (c) Contents.—Each risk report shall include, at a minimum—
- (1) a representative summary of the capabilities of the models that are relevant to enumerated risk, addressing each enumerated risk and, if the developer has previously published a risk report, describing any material change in those capabilities since the most recently published risk report;
 - (2) a description of the key threat models the developer tracks with respect to each enumerated risk, of how the observed capabilities of the models relate to each such threat model, and of the key mitigations the developer has put in place;
 - (3) an assessment of the level of each enumerated risk, made in accordance with its safety framework, that remains after accounting for the mitigations described under paragraph (2), which shall set forth the principal evidence and reasoning underlying the assessment, and, to the extent reasonably achievable with redactions applied pursuant to section 9, allow a reasonable person to reach a conclusion similar to that which the large frontier developer reached in analyzing the level of risk posed by its models; and
 - (4) if the developer has previously published a risk report, a comparison of the assessed level of each enumerated risk to the level assessed in the most recently published risk report.
- (d) Relationship to Safety Framework.—The safety framework required by section 4 describes the enduring policies, procedures, and methods by which a large frontier developer identifies, assesses, and manages enumerated risk. A risk report required by this section applies that framework to state the developer’s current assessment of the enumerated risk posed by the frontier models described in subsection (a), and is updated periodically as those models, and the risks they may present, evolve.

SEC. 7. CRITICAL SAFETY INCIDENT REPORTING.

- (a) Reporting Mechanism.—Not later than 120 days after the date of enactment of this Act, the Agency shall establish and maintain a mechanism through which a frontier developer or a member of the public may confidentially submit to the Agency a report of a critical safety incident. The mechanism shall provide for the amendment of a previously submitted report.
- (b) Contents of Reports.—A report of a critical safety incident shall state—

Last edited 9/29/26

- (1) the date of the incident;
- (2) the reasons the incident constitutes a critical safety incident;
- (3) a short and plain description of the incident; and
- (4) whether the incident is associated with the internal use of a frontier model.

(c) **Deadline for Reporting.**—A frontier developer shall submit a report of a critical safety incident no later than 15 days after the frontier developer determines, following a reasonable and good faith inquiry and without unreasonable delay, that a critical safety incident has occurred. Notwithstanding the foregoing, a containment failure incident is subject to the 15-day requirement of this subsection only if it is reasonably expected to result in more than \$250,000 in damage or loss incurred by a person other than the frontier developer, or in death or serious physical injury.

(d) **Imminent Risk.**—If a frontier developer discovers that a critical safety incident poses an imminent risk of death or serious injury, the frontier developer shall, not later than 24 hours after the discovery, report the incident to the Agency and to any law enforcement or public safety agency with jurisdiction over the incident that is appropriate in light of the nature of the incident.

(e) **Containment Failure Reporting .**— A frontier developer shall annually report to the Agency and the Attorney General the total number of containment failure incidents not subject to the 15 day requirement during the preceding twelve months. For each such incident, the frontier developer shall also maintain, for five years, a record including the information described in subsection (b), and make it available to the Agency and the Attorney General upon request.

(f) **Amended Report .**— A frontier developer that discovers information about a critical safety incident after filing the initial report required by this section may file an amended report. A report submitted in good faith on the basis of information reasonably available to the frontier developer at the time of submission satisfies this section, and any subsequent amendment of such a report under this subsection shall not be treated as evidence that the original report failed to satisfy subsections (c) and (d).

(g) **Interim Procedure.**—Until the mechanism required by subsection (a) is established, a frontier developer shall satisfy subsections (c) through (f) by submitting the required report to the Agency in writing by any reasonable means.

(h) **Review and Sharing.**—The Agency shall review each report submitted by a frontier developer under this section and may review any report submitted by a member of the public. The Agency may share a report submitted by a frontier developer with another Federal agency or a federally funded research and development center only upon

Last edited 9/29/26

determining the recipient has a need for the report in the performance of its duties, and shall condition such sharing on the recipient's agreement to maintain the confidentiality of the report and to protect it against unauthorized access.

(i) **Protection of Reports.**—The Agency shall limit access to reports submitted under this section to personnel with a specific need to know the information, and shall take all reasonable precautions to protect such reports from unauthorized access. Such reports are subject to section 9.

(j) **Annual Public Report.**—Not later than March 31 of the second calendar year beginning after the date of enactment of this Act, and annually thereafter, the Agency shall publish, and submit to Congress, a report containing anonymized and aggregated information concerning the critical safety incidents reported under this section during the period beginning on the date of enactment of this Act, in the case of the first report, or on the day after the last day covered by the preceding report, in the case of each subsequent report, and ending on the last day of the preceding calendar year. The Agency shall not include in a report under this subsection any information the publication of which would compromise a trade secret, the risk-prevention mechanisms or cybersecurity of a frontier developer, public safety, or the national security of the United States, and shall afford each frontier developer with an incident report described therein 30 days to identify such information before publication.

(k) **Voluntary Reports.**—A person is encouraged, but not required, to report to the Agency an incident that would constitute a critical safety incident but for the fact that it involves a model that is not a frontier model. A report submitted under this subsection is subject to subsection (i) and section 9, and the submission of such a report shall not, by itself, be treated as evidence of a violation of this Act.

(l) **Notice.** If the Agency shares a critical safety incident report pursuant to subsection (h), or initiates an inquiry or action on the basis of a report submitted by a member of the public, the Agency shall provide the frontier developer with notice of such sharing or action, except to the extent the Agency determines in writing that notice would compromise an ongoing law enforcement or national security investigation. Notice of sharing pursuant to subsection (h) shall occur not later than 30 days after the sharing and identify the recipient of the report and nature of the inquiry.

SEC. 8. INDEPENDENT EVALUATION.

(a) **Qualification and Independence.**—A person qualifies as a qualified independent evaluator if the person satisfies each of the following:

(1) **COMPETENCE.**—The person possesses demonstrated competence to conduct the evaluations required by this section, including by employing or contracting with

Last edited 9/29/26

individuals who possess technical expertise in the safety of frontier models and in each enumerated risk within the scope of the engagement.

(2) INDEPENDENCE.—The person has no financial, operational, or management dependence on the large frontier developer whose frontier models or safety framework it evaluates, or on any affiliate of that developer, and is otherwise free from that developer’s control in reaching conclusions and making recommendations, including through its contracts and conflict-of-interest policies.

(3) COMPENSATION.—The person may accept compensation from the large frontier developer, provided that no payment to the person, and no amount of any payment, is conditioned on the results of the evaluation.

(4) CERTIFICATION.—Before accepting an engagement, the person certifies in writing to the large frontier developer and to the Agency that it satisfies paragraphs (1) through (3), disclosing its sources of funding and remuneration for the engagement, including any funding from other large frontier developers, any other current or recent engagement with the developer or an affiliate, and any other fact that could reasonably be expected to bear on its independence.

(b) Registration and Rating System.—

(1) REGISTRATION .— Not later than 180 days after the date of enactment of this Act, the Agency shall establish a mechanism through which a qualified independent evaluator shall register with the Agency, identifying each enumerated risk for which it possesses the competence described in subsection (a)(1), and shall publish a list of registered evaluators.

(2) RATING SYSTEM.— Not later than 180 days after the date of enactment of this Act, the Agency shall, in consultation with federal agencies and federally funded research and development centers with relevant technical expertise, design criteria for rating evaluators. This rating system shall consider, among other criteria, whether the evaluator meets paragraphs (1) and (2) of subsection (a) and the rigor of the evaluator’s public reasoning.

(c) Engagement Requirement.—

(1) IN GENERAL.—Not later than 270 days after the date of enactment of this Act, or 180 days after the date on which a person becomes a large frontier developer, whichever is later, and thereafter on a continuing basis, a large frontier developer shall engage at least one qualified independent evaluator to conduct an independent evaluation of the developer’s frontier models with respect to each enumerated risk.

(2) FREQUENCY.—A large frontier developer shall obtain, and a qualified independent evaluator shall deliver, an evaluation report under subsection (e) not later

Last edited 9/29/26

than one year after the date of enactment of this Act, or 270 days after the date on which the person becomes a large frontier developer, whichever is later, and not less frequently than every 180 days thereafter. An evaluator may deliver reports more frequently if, in its professional judgment, the nature, severity, or evolution of the enumerated risks presented warrants it.

(3) SUBCONTRACTING.—A qualified independent evaluator may subcontract a portion of an evaluation to a person with specialized expertise. The subcontractor shall be subject to every requirement of this section that applies to the evaluator with respect to the work the subcontractor performs, and the engaging evaluator shall remain responsible for the evaluation and the resulting report.

(4) MULTIPLE EVALUATORS.—A large frontier developer may engage more than one qualified independent evaluator, each focused on a different enumerated risk or on a different aspect of the requirements of this section, if collectively the evaluators perform every duty required by this section, each satisfies subsection (a), and each states in its report the scope of its evaluation and the duties it performed.

(d) Access.—

(1) IN GENERAL.—Except as provided in subparagraph (A), a large frontier developer shall grant a qualified independent evaluator timely access to all information reasonably necessary for the evaluator to perform its duties under this section, including use of the large frontier developer's most capable frontier models, access to unredacted versions of every document the large frontier developer has published or submitted under this Act, and access to information reasonably comparable to the access of the large frontier developer's internal risk assessment teams. The large frontier developer shall provide the evaluator a reasonable opportunity to ask, and shall provide reasonable responses to, questions concerning the frontier models, the likelihood of enumerated risks, and the large frontier developer's safeguards. The provision of information to a qualified independent evaluator under this subsection does not constitute a waiver of any applicable privilege under federal or state law or of any protection provided by law to a trade secret or to confidential commercial information.

(A) EXCLUDED INFORMATION.—Nothing in this subsection requires a large frontier developer to provide a qualified independent evaluator with access to customer or user content, personal information, information the developer has committed by contract or published policy not to retain or access, information subject to a confidentiality obligation owed to a third party, or information the disclosure of which would violate applicable law.

Last edited 9/29/26

(2) SECURITY PROTOCOLS.—To protect its trade secrets and confidential business information, its cybersecurity, public safety, or the national security of the United States, a large frontier developer may impose reasonable security protocols on a qualified independent evaluator, including restrictions on note-taking, copying, retaining, or removing materials, requirements for on-premises review, and confidentiality requirements. These protocols should be comparable to the protocols the developer uses to handle similar information. Such protocols may require any individual who performs work for the evaluator with access to the developer’s premises, systems, or confidential information to satisfy background screening, confidentiality, insider-risk, and workplace-conduct requirements comparable to those the developer applies to its own personnel with similar access, consistent with applicable law. A large frontier developer may, for a documented security or conduct reason, require that a specific individual be removed from such access, in which case the evaluator shall assign a qualified replacement. A developer may not require the removal of an individual on the basis of the individual’s views, conclusions, or findings. After completion of an evaluation report, the evaluator shall delete all of the large frontier developer’s information related to the evaluation in its possession or control outside the developer’s systems, including copies, notes and extracts.

(3) LIMITATION ON PROTOCOLS.—A large frontier developer may not, by security protocol, contract, or otherwise, restrict the content of a qualified independent evaluator’s report or the evaluator’s publication of a report under subsection (f), except to the extent necessary to prevent the disclosure of information that may be redacted under section 9. Any provision of an agreement between a large frontier developer and a qualified independent evaluator that conflicts with this paragraph is void and unenforceable.

(4) EMPLOYMENT STATUS; APPLICABLE LAW.—The imposition or enforcement of a security protocol under paragraph (2), or the provision of access or workspace to a qualified independent evaluator, does not make a large frontier developer an employer or joint employer of any individual who performs work for the evaluator. A large frontier developer shall provide access under this subsection in a manner consistent with applicable law, including the data protection and labor laws of any jurisdiction in which the relevant information or personnel are located.

(e) Evaluation Reports.—A qualified independent evaluator shall deliver each evaluation report to the large frontier developer. Each evaluation report shall include—

(1) a description of the scope of the evaluation, the period covered, the materials reviewed, the methodology used, the personnel who participated, any material limitation on the scope of the evaluation or on the information available to the

Last edited 9/29/26

evaluator, and the evaluator's procedures for managing conflicts of interest together with any conflict of interest of a participating individual;

(2) an independent assessment of the enumerated risk posed by the developer's frontier models, taking into account model capabilities, applicable threat models, and the developer's safeguards, including any increased difficulty of monitoring, assessing, or mitigating enumerated risk that results from the models' ability to manipulate and evade monitoring, and including risks that automated research and development could shortly increase the level of risk;

(3) commentary on the developer's most recently published risk report, including an assessment of—

(A) the adequacy and completeness of the information disclosed;

(B) the analytical rigor of the developer's risk methodology;

(C) the appropriateness and materiality of redactions made to the published version; and

(D) whether the evaluator disagrees with key claims in the risk report, including its overall assessment of the level of each enumerated risk;

(4) an independent assessment, made in accordance with subsection (i), of whether conditions that could produce enumerated harm are present or impending with respect to the developer's development, deployment, or internal use of any frontier model;

(5) any corrective action the evaluator recommends; and

(6) the signature of the individual responsible for the evaluation, certifying that the report is accurate, that it omits no material information, and that the evaluator satisfied subsection (a) throughout the engagement.

(f) **Publication.**—Not later than 30 days after delivering an evaluation report, a qualified independent evaluator shall publish a version of the report redacted in accordance with section 9, and the large frontier developer shall publish a link to that version. Not less than 15 days before publishing, the evaluator shall provide the proposed public version to the developer, which may redact, with a written basis, information permitted to be redacted under section 9. Within 10 days of receiving it, the developer shall return the redacted report to the evaluator, who shall state in the published report whether the redacted material informed any conclusion and whether the evaluator disagreed with the basis for the redaction. Nothing in this subsection permits a developer to restrict the content of a report, or to delay publication of information, except to the extent section 9 permits redaction.

(g) **Out-of-Cycle Reports.**—

Last edited 9/29/26

(1) TRIGGER.—Not later than 14 days, or 72 hours if the evaluator concludes that conditions that could produce enumerated harm are present or impending, after a qualified independent evaluator determines, in light of new information, emergent model capabilities, a material change to a frontier model, or some other material change or development, that a conclusion stated in a prior evaluation report is no longer valid, the evaluator shall deliver a supplemental report to the developer and to the Agency.

(2) CONTENTS.—A supplemental report shall describe the information and conclusions that triggered it, state, in accordance with subsection (i), whether conditions that could produce enumerated harm are present or impending, and describe any additional corrective action the evaluator recommends.

(3) APPLICATION OF OTHER SUBSECTIONS.—Subsection (f) applies to a supplemental report.

(h) Additional Evaluation.—

(1) IN GENERAL.— The Agency may require a large frontier developer to obtain, within a period specified by the Agency of not less than 60 days, an additional evaluation limited in scope to a matter specified by the Agency, if the Agency sets forth in a written determination the specific facts establishing that the additional evaluation is necessary to—

(A) validate the findings of a prior evaluation, based on credible evidence that those findings may be inaccurate;

(B) assess the implementation of a corrective action; or

(C) assess a significant change in enumerated risk, including following a critical safety incident.

(2) APPLICABLE REQUIREMENTS.—Subsections (d) through (f) apply to an additional evaluation required under this subsection.

(i) Assessments of Imminent Enumerated Risk.—

(1) CONTENTS OF THE ASSESSMENT.—In each assessment made under subsection (e)(4) or (g)(2), a qualified independent evaluator shall state and apply its own methodology, include information required by subsection (e)(1), and shall state, with the evidence and reasoning supporting each—

(A) the capabilities of the large frontier developer's frontier models relevant to each enumerated risk, the safeguards the developer has applied, and the evaluator's assessment of their effectiveness;

Last edited 9/29/26

(B) the threat models under which those capabilities, notwithstanding those safeguards, could materially contribute to enumerated harm;

(C) whether conditions that could produce enumerated harm are present or impending and, if so, which conditions and within what period; and

(D) the source and amount of the evaluator's compensation for the engagement and any other fact bearing on its independence.

(2) **PROFESSIONAL OPINION.**—A qualified independent evaluator may state whether, in the evaluator's professional judgment, the facts described in paragraph (1) present an imminent enumerated risk. Such a statement is an expression of professional opinion and has the effect, and only the effect, provided in subsection (j).

(3) **NOTICE AND OPPORTUNITY TO RESPOND.**—Before transmitting an assessment stating that conditions that could produce enumerated harm are present or impending, a qualified independent evaluator shall provide the large frontier developer written notice of the proposed conclusion, the evidence and reasoning supporting it, and not less than 48 hours to respond in writing. The evaluator shall transmit the response with the assessment and state whether and in what respects it altered the evaluator's conclusions. An evaluator may dispense with prior notice if it determines in writing that the imminence of the potential harm forecloses it, in which case it shall provide the notice and opportunity to respond promptly after transmittal and supplement the assessment as appropriate.

(4) **TRANSMITTAL AND RESPONSE.**—A qualified independent evaluator shall transmit each assessment under this subsection, with any response received under paragraph (3), simultaneously to the large frontier developer and to the Agency, not later than 72 hours after the assessment is completed, not counting any period provided to the developer under paragraph (3). Not later than 30 days after receiving an assessment stating that conditions that could produce enumerated harm are present or impending, the Agency shall require an additional evaluation under subsection (h), refer the matter to the [Enforcement Authority], or publish a written determination, redacted in accordance with section 9, stating that no further action is warranted and the basis for that determination.

(j) **Effect of an Assessment.**—

(1) **NOT CONCLUSIVE.**—An assessment, conclusion, or opinion under subsection (i) is not conclusive of any fact or legal conclusion, is not entitled to deference, and does not establish that a frontier model presents an imminent enumerated risk.

(2) **DETERMINATION BY AN OFFICER REQUIRED.**—No department, agency, or officer of the United States may impose a penalty, restriction, or sanction on a frontier

Last edited 9/29/26

developer, or suspend a frontier model, under this Act on the ground that the model presents an imminent enumerated risk except upon that officer's own determination, made on the entire record and after affording the developer notice and an opportunity to be heard, stating the weight given to any assessment under subsection (i) and the reasons for that weight. If the officer determines in writing that the imminence of the risk makes prior notice and an opportunity to be heard infeasible or dangerous to the public interest, the officer may act before providing them, but not before making that determination, and shall afford the developer notice and an opportunity to be heard as soon as practicable after the action.

(3) **CONTESTED FACTS.**—In any proceeding in which an assessment under subsection (i) is relied upon, the frontier developer may contest the assessment and any fact stated in it, and the court shall determine each contested fact without deference to the evaluator.

(4) **NO AUTHORITY IN THE EVALUATOR.**—A qualified independent evaluator has no authority under this Act to compel the production of information, to impose any sanction or remedy, to require or prohibit any act of a frontier developer, or to suspend or restrict the development, deployment, or internal use of a frontier model. Nothing in subsection (i) or this subsection confers authority to suspend or restrict the development, deployment, or internal use of a frontier model.

(k) **Evaluation Ecosystem.**—

(1) **FUNDING.**—Subject to the availability of appropriations, the Agency may provide funding so that qualified independent evaluators may conduct evaluations while remaining financially independent of any frontier developer, and shall provide resources to nascent organizations seeking to become such evaluators. Funding may be provided only under published, neutral criteria that do not condition it on the identity or viewpoint of, or any conclusion previously reached by, the evaluator.

(2) **REPORT ON THE EVALUATOR ECOSYSTEM.**—Not later than one year after the date of enactment of this Act, and annually thereafter, the Comptroller General of the United States shall submit to Congress a report addressing the number of qualified independent evaluators and their capacity to meet the demand for evaluations under this section, any barrier to entry or other factor affecting the development of a competitive ecosystem, the degree to which evaluators are independent of the artificial intelligence industry, and recommendations for action that would strengthen their independence, quality, or capacity.

(3) **LIMITATION.**—Any standard or best practice the Agency publishes for the conduct of evaluations under this section is advisory. No such standard or best practice

Last edited 9/29/26

may displace or narrow the requirements of subsection (a) or operate as a condition of engagement under this section.

(4) **AUTHORIZATION OF APPROPRIATIONS.**—There are authorized to be appropriated [§] to carry out this subsection.

(l) **Recordkeeping.**—A large frontier developer shall retain an unredacted copy of each evaluation report, and the materials supporting it, for not less than 5 years after the date the report is delivered.

(m) **Prohibition on Misrepresentation.**—A qualified independent evaluator shall not knowingly make a material misrepresentation, or knowingly omit material information, in a report or certification prepared under this section.

(n) **Limitation on Liability.**—

(1) **IN GENERAL.**—A qualified independent evaluator shall not be liable under Federal or State law for any claim for loss caused by, arising out of, or resulting from the materialization of an enumerated risk of a frontier model the evaluator evaluated, if the evaluator satisfied subsection (a) at the time of the evaluation.

(2) **EXCEPTION.**—Paragraph (1) does not apply to a claim for death or serious injury proximately caused by the willful misconduct of the evaluator. In an action asserting such a claim, the plaintiff shall bear the burden of proving willful misconduct by clear and convincing evidence.

(3) **NO EFFECT ON OTHER OBLIGATIONS.**—Nothing in this subsection limits the liability of a qualified independent evaluator for a violation of subsection (m) or for breach of an obligation owed to a large frontier developer.

(4) **QUALIFIED PRIVILEGE.**—A statement made in good faith by a qualified independent evaluator in a report or assessment under this section is privileged, and no action for defamation, disparagement, or a like claim may be maintained on the basis of that statement except on clear and convincing evidence that the evaluator made it knowing it to be false.

SEC. 9. REDACTION AND PROTECTION OF SENSITIVE INFORMATION.

(a) **Permitted Redactions.**—Before publishing a document required to be published under this Act, a frontier developer or a qualified independent evaluator may redact information from the published version only to the extent necessary to protect the frontier developer's trade secrets, any applicable privilege, personally identifiable information or other information that would constitute an invasion of privacy, confidential customer information, the frontier developer's cybersecurity, public safety, or the national security of the United States or to comply with any federal or state law not preempted by section 14.

Last edited 9/29/26

- (b) **Description and Justification.**—A published version containing a redaction shall describe the character of the redacted information and state the justification for the redaction, to the extent permitted by the concerns that justify the redaction.
- (c) **Retention.**—A large frontier developer that redacts information under subsection (a) shall retain the unredacted document for not less than 5 years after the date of publication.
- (d) **Treatment of Information Held by the Government.**—
- (1) **IN GENERAL.**—The Agency and the [Enforcement Authority] shall limit access to information received under this Act to personnel with a specific need to know the information, and shall take all reasonable precautions to protect the information against unauthorized access and disclosure.
 - (2) **NO WAIVER.**—The submission of information to the Federal Government under this Act does not constitute a waiver of any applicable privilege or of any protection provided by law to a trade secret or to commercial or financial information.
 - (3) **FOIA EXEMPTION.**—Information submitted to the Federal Government under this Act, other than information required by this Act to be published, shall be withheld from the public, and shall be exempt from disclosure under section 552(b)(3) of title 5, United States Code. This paragraph is a statute described in section 552(b)(3)(B) of title 5, United States Code, and specifically cites to section 552(b)(3) of that title.
 - (4) **SHARING WITH CONGRESS.**—Nothing in this subsection restricts the provision of information to Congress or to the Comptroller General of the United States.
- (e) **Rule of Construction.**—Nothing in this section authorizes a large frontier developer to withhold information from a qualified independent evaluator under section 8(d).

SEC. 10. PROTECTION OF INDIVIDUALS REPORTING ARTIFICIAL INTELLIGENCE SAFETY CONCERNS.

- (a) **Definitions.**—In this section:
- (1) **COVERED INDIVIDUAL.**—The term “covered individual” means a current or former employee, or a current or former independent contractor, of a frontier developer or of a qualified independent evaluator.
 - (2) **COVERED EMPLOYER.**—The term “covered employer” means a frontier developer or a qualified independent evaluator.
 - (3) **PROTECTED DISCLOSURE.**—The term “protected disclosure” means a disclosure of information that the covered individual reasonably believes evidences—

Last edited 9/29/26

- (A) a violation of this Act; or
 - (B) an activity of a frontier developer that poses a specific and substantial danger to public health or safety arising from a catastrophic risk or enumerated risk.
- (b) Prohibition on Retaliation.—A covered employer may not discharge, demote, suspend, threaten, blacklist, harass, or in any other manner discriminate against a covered individual in the terms or conditions of employment or engagement because the covered individual—
- (1) made a protected disclosure to the Agency, the [Enforcement Authority], another Federal or State regulatory or law enforcement agency, a Member or committee of Congress, a person with supervisory authority over the covered individual, or another individual working for the covered employer whom the covered individual reasonably believes has authority to investigate, discover, or correct the matter disclosed;
 - (2) initiated, testified in, or assisted in an investigation or proceeding relating to a protected disclosure; or
 - (3) used an internal channel established under subsection (d), or sought to use such a channel where none was established.
- (c) Prohibition on Contractual Restriction.—A covered employer may not adopt, maintain, or enforce a rule, policy, or agreement that prevents a covered individual from making a protected disclosure. Notwithstanding chapter 1 of title 9, United States Code, the rights and remedies provided by this section may not be waived or limited by any agreement, policy, or condition of employment or engagement, including an agreement requiring arbitration or other alternative dispute resolution before relief is sought under subsection (f).
- (d) Internal Reporting Channel.—Not later than 180 days after the date of enactment of this Act, or 90 days after the date on which a person becomes a large frontier developer, whichever is later, a large frontier developer shall provide a reasonable internal process through which a covered individual may make a protected disclosure anonymously. The process shall permit a covered individual to communicate with the covered employer through an anonymized identifier or other means that does not disclose the covered individual's identity, and shall include—
- (1) a monthly update to the person who made the disclosure regarding the status of the developer's investigation and the actions taken in response; and
 - (2) disclosure to the officers and directors, or equivalent governing body, of the developer, not less frequently than quarterly, of the disclosures made and the responses given, except that a disclosure alleging wrongdoing by a particular officer or director shall not be provided to that officer or director.

Last edited 9/29/26

(e) Notice of Rights.—Not later than 180 days after the date of enactment of this Act, or 90 days after the date on which a person becomes a covered employer, whichever is later, a covered employer shall provide each covered individual with clear notice of the rights provided by this section, either by posting and maintaining a notice in each workplace and providing an equivalent notice to each covered individual who works remotely, or by providing written notice to each covered individual not less frequently than annually.

(f) Enforcement.—

(1) COMPLAINT TO THE SECRETARY OF LABOR.—A covered individual who alleges a violation of subsection (b) or (c) may file a complaint with the Secretary of Labor. The complaint shall be governed by the rules and procedures, including the burdens of proof, set forth in section 42121(b) of title 49, United States Code.

(2) CIVIL ACTION.—If the Secretary of Labor has not issued a final decision within 180 days after a complaint is filed, and the delay is not attributable to the bad faith of the complainant, the covered individual may bring a civil action against the covered employer in an appropriate district court of the United States, without regard to the amount in controversy. A party to such an action is entitled to trial by jury.

(3) LIMITATIONS.—An action or complaint under this subsection shall be commenced not later than the earlier of 3 years after the date on which the covered individual knew or reasonably should have known of the facts material to the claim, or 6 years after the date of the violation.

(4) INJUNCTIVE RELIEF.—A court may grant temporary or preliminary injunctive relief on a showing of reasonable cause to believe a violation has occurred. In determining whether such relief is warranted, the court shall consider the chilling effect on other covered individuals of the conduct alleged. Injunctive relief granted under this paragraph shall not be stayed pending appeal. Nothing in this paragraph prevents a covered employer from disciplining or terminating a covered individual for conduct unrelated to a protected disclosure.

(g) Relief.—Notwithstanding the remedial provisions incorporated by subsection (f)(1), relief for a covered individual who prevails in a proceeding under subsection (f)(1) or an action under subsection (f)(2) shall include reinstatement with the seniority status the covered individual would have had but for the violation, two times the amount of back pay owed with interest, compensatory damages, and litigation costs, expert witness fees, and reasonable attorney's fees.

(h) Rule of Construction.—The rights and remedies provided by this section are in addition to, and do not displace, any right or remedy provided by any other Federal or State law.

Last edited 9/29/26

SEC. 11. ENFORCEMENT.

(a) Prohibited Acts.—It shall be unlawful for—

- (1) a frontier developer to fail to publish, submit, or transmit a document that this Act requires the frontier developer to publish, submit, or transmit, or to publish, submit, or transmit such a document that materially fails to conform to the requirements of this Act;
- (2) a large frontier developer to materially fail to comply with its safety framework;
- (3) a frontier developer to fail to report a critical safety incident as required by section 7;
- (4) a frontier developer knowingly, or with reckless disregard for the truth, to make a materially false statement, or to convey a materially false impression, concerning a catastrophic risk or enumerated risk arising from a frontier model of such developer or such developer's management of catastrophic risk or enumerated risk, or a large frontier developer knowingly, or with reckless disregard for the truth, to make a materially false statement, or to convey a materially false impression, concerning its implementation of or compliance with its safety framework, in—
 - (A) a document published, submitted, or transmitted under this Act;
 - (B) a communication with the Agency or the [Enforcement Authority] concerning a matter within the scope of this Act; or
 - (C) a communication with a qualified independent evaluator in connection with an evaluation under section 8;
- (5) a large frontier developer to fail to provide a qualified independent evaluator the access required by section 8(d), or to impose or enforce a restriction prohibited by section 8(d)(3);
- (6) a large frontier developer to fail to engage a qualified independent evaluator, or to fail to obtain an evaluation, as required by section 8(c) or section 8(h), provided that it is not a violation of this paragraph where the delay results from the evaluator and the developer provided the access required by section 8(d);
- (7) a large frontier developer to fail to retain a record as required by section 8(l) or section 9(c);
- (8) a large frontier developer to violate section 10(d), or a covered employer to violate subsection (b), (c), or (e) of section 10;
- (9) a person bound by an order entered or issued under section 12 to knowingly violate that order;

Last edited 9/29/26

(10) a qualified independent evaluator to violate section 8(m), to fail to deliver or publish a report as required by section 8(e), 8(f), or 8(g), or to accept or continue an engagement while not satisfying section 8(a); or

(11) a large frontier developer to fail to maintain a security program, or to conduct the testing or obtain the audit, required by section 4(f), or to fail to provide the access required by, or to impose or enforce a restriction prohibited by, section 4(f)(4).

(b) Good Faith.—Subsection (a)(4) does not apply to a statement that was made in good faith and was reasonable under the circumstances at the time it was made.

(c) Civil Penalties.—

(1) FIRST VIOLATION.—A person that commits an act made unlawful by subsection (a) shall be liable to the United States for a civil penalty in an amount not to exceed the greater of [\$1,000,000] or [] percent of the person's global annual revenue.

(2) SUBSEQUENT VIOLATIONS.—A person that commits an act made unlawful by subsection (a) after a court has entered a final, non-appealable judgment against that person under this section shall be liable to the United States for a civil penalty in an amount not to exceed the greater of [\$3,000,000] or [] percent of the person's global annual revenue.

(3) FACTORS.—In determining the amount of a civil penalty, the court shall consider the severity of the violation, the degree of catastrophic risk or enumerated risk to which the violation contributed, whether the violation was intentional, the person's history of prior violations, whether the person voluntarily and promptly disclosed and corrected the violation, and the person's ability to pay.

(d) Civil Actions.—

(1) EXCLUSIVE ENFORCEMENT.—Only the [Enforcement Authority] may bring an action to enforce this Act or to collect a civil penalty under subsection (c).

(2) COURT ENFORCEMENT REQUIRED.—A civil penalty under subsection (c) may be imposed only by a district court of the United States in a civil action brought by the [Enforcement Authority]. Neither the Agency nor the [Enforcement Authority] may assess or impose a civil penalty by order.

(3) VENUE.—An action under this subsection may be brought in the district court of the United States for any district in which the defendant resides, is found, has an agent, or transacts business.

(4) LIMITATIONS.—An action under this subsection shall be commenced not later than 5 years after the date on which the cause of action accrues.

Last edited 9/29/26

(e) Injunctive Relief.—In an action under subsection (d), the court may grant such injunctive relief as is necessary to compel compliance with a requirement of this Act. Nothing in this subsection authorizes an order suspending the development, deployment, or internal use of a frontier model on any ground, including the ground that it presents a risk of enumerated harm; the authority to seek such relief is provided in section 12.

(f) Enforcement Solely by the [Enforcement Authority].—Except for an action under section 10, no provision of this Act may be enforced by any person other than the [Enforcement Authority].

SEC. 12. ORDERS RESTRICTING THE DEVELOPMENT, DEPLOYMENT, OR INTERNAL USE OF A FRONTIER MODEL.

(a) Authority; Exclusivity.—

(1) IN GENERAL.—On the application of the Agency, the United States District Court for the District of Columbia may enter an order suspending the development, deployment, or internal use of a frontier model on the ground that the model presents an imminent enumerated risk (in this section, a “blocking order”), and has exclusive jurisdiction over a proceeding under this section.

(2) EXCLUSIVE MECHANISM.—No department, agency, or officer of the United States may suspend the development, deployment, or internal use of a frontier model under this Act except pursuant to an order entered or issued under this section. Nothing in this section affects authority under any provision of law other than this Act.

(3) PERSONS BOUND; ENFORCEMENT.—An order under this section binds the frontier developer, its officers, agents, employees, attorneys, and affiliates, and any other person in active concert or participation with any of them, in each case who receives actual notice of the order. The Agency may apply to the court to enforce an order under this section, notwithstanding section 11(d)(1).

(4) DEVELOPERS THAT ARE NOT LARGE FRONTIER DEVELOPERS.—For purposes of this section, section 8 applies to a frontier developer that is not a large frontier developer as it applies to a large frontier developer.

(b) Determination and Independent Assessment.—

(1) DETERMINATION.—The Agency may not apply for, or apply to renew, a blocking order unless the head of the Agency, without delegation, has made and filed a written determination that satisfies section 8(j)(2) and states, with the evidence and reasoning supporting each—

(A) the specific pathway or pathways by which the frontier model is expected to materially contribute to enumerated harm, including the model capabilities relied

Last edited 9/29/26

upon and the reasons the frontier developer's safeguards are inadequate to address them;

(B) the period within which the Agency expects enumerated harm to occur in the absence of the order, and the basis for that expectation;

(C) why no less restrictive measure, including any measure the frontier developer proposed, will adequately avert or mitigate the risk, and each less restrictive measure the Agency considered; and

(D) the weight the Agency gave to the assessment filed under paragraph (2) and the reasons for any disagreement with it.

A determination that asserts an imminent enumerated risk without satisfying subparagraphs (A) and (B) does not satisfy this paragraph.

(2) ASSESSMENT.—The Agency may not apply for, or apply to renew, a blocking order unless it files an assessment, made under section 8(i), that addresses the frontier model and the enumerated risk identified in the determination. The Agency may file an assessment transmitted under section 8(i)(4) or one prepared by a qualified independent evaluator the Agency engages for that purpose. The assessment shall include the statement described in section 8(i)(2) or a statement that the evaluator declines to express that judgment and the reasons for declining. Section 8(j) governs the effect of an assessment, which is a necessary, and not a sufficient, condition to entry or renewal.

(3) EVALUATORS ENGAGED BY THE AGENCY.—A qualified independent evaluator engaged under paragraph (2) shall satisfy section 8(a), shall not be an officer or employee of the United States, shall be free from the control of the Federal Government in reaching its conclusions, shall not accept payment conditioned on the results of the assessment, and shall disclose to the court and to the frontier developer its compensation and any other engagement with the Federal Government bearing on its independence. Compensation from the Agency does not, by itself, constitute financial dependence under section 8(a)(2). Sections 8(d) and 8(i)(3) apply to the engagement as though it were made under section 8(c).

(4) EVIDENCE OF THE FRONTIER DEVELOPER.—A frontier developer may file an assessment prepared by a qualified independent evaluator of its own selection, examine any evaluator whose assessment the Agency files, and present any other evidence and argument bearing on subsection (c).

(c) Grounds.—

Last edited 9/29/26

(1) **REQUIRED FINDINGS.**—The court may enter a blocking order only if the Agency establishes by a preponderance of the evidence that—

- (A) the frontier model presents an imminent enumerated risk;
- (B) the order is necessary to avert or mitigate that risk;
- (C) no less restrictive measure would adequately avert or mitigate that risk; and
- (D) the order is narrowly tailored to the pathway or pathways identified under subsection (b)(1)(A) and is no broader in scope or duration than necessary.

(2) **CONSIDERATIONS.**—In applying subparagraphs (C) and (D) of paragraph (1), the court shall consider the harm the order would cause to the frontier developer, to users, and to the public, including any adverse effect on safety research, on cybersecurity, or on uses of the frontier model that do not contribute to the risk identified. In determining whether the frontier model presents an imminent enumerated risk, the court shall give significant evidentiary weight to any evaluation before the court under section 8(i).

(3) **EXCLUSIVE GROUNDS; BURDEN; NO DEFERENCE.**—Paragraph (1) states the exclusive grounds for a blocking order, and the Agency bears the burden under paragraph (1) at every stage, including on renewal and on a motion to dissolve or modify. The court shall determine each question under paragraph (1) *de novo*, without deference to a determination of the Agency or to a prior order.

(d) **Notice; Duration; Renewal; Dissolution.**—

(1) **NOTICE.**—Except as provided in paragraph (2), the court may not enter a blocking order unless the frontier developer has received notice of the application and a reasonable opportunity to respond, present evidence, and be heard. The Agency shall serve the application and every paper filed in support of it on the frontier developer on the date of filing.

(2) **ORDERS ENTERED BEFORE THE DEVELOPER IS HEARD.**—The court may enter a blocking order before the frontier developer has been heard only if the head of the Agency, without delegation, files an assessment made under section 8(i) and certifies the specific facts establishing that the imminence of the enumerated risk makes it infeasible, or dangerous to the public interest, to hear the developer first, and counsel certifies the efforts made to give notice. The court shall make a record of the proceeding available to the developer, subject to subsection (g), and shall set the matter for hearing before the order expires. An order entered under this paragraph expires not later than 7 days after entry notwithstanding paragraph (3), may not be extended or renewed except after the developer has been heard and on an application

Last edited 9/29/26

satisfying subsection (b), and may not be entered more than once with respect to the same frontier model and the same enumerated risk.

(3) DURATION.—A blocking order expires [45] days after entry unless earlier dissolved or renewed. Expiration is automatic, requires no action by any party or the court, and is not stayed, tolled, or extended by the pendency of an application, a motion, or an appeal.

(4) RENEWAL.—The court may renew a blocking order for not more than [45] days on an application that satisfies subsections (b) and (c). The court shall determine each requirement of subsection (c)(1) on the circumstances existing at the time of the application and the frontier developer may submit new evidence and argument, including evidence of a safeguard adopted since the prior order. An application to renew shall be filed not later than [21] days before the existing order expires, the frontier developer may respond not later than [10] days after the application is filed, and if the court has not granted the application before the period in paragraph (3) lapses, the existing order expires.

(5) DISSOLUTION AND MODIFICATION.—A frontier developer may move at any time to dissolve or modify a blocking order, including on the basis of changed circumstances, new evidence, or a safeguard the developer has adopted. The Agency may respond not later than [3] days after the motion is filed, and the court shall decide the motion as promptly as justice requires and in any event not later than [5] days after the response is filed or the time for responding has expired. The court shall dissolve or modify the order to the extent a requirement of subsection (c)(1) is not satisfied.

(e) Scope and Limitations of Orders.—

(1) SCOPE.—A blocking order may suspend the development, deployment, or internal use of a frontier model, whether the model is deployed, is to be deployed, or remains in development, and may suspend use by the public, by the frontier developer, or both. The frontier developer shall comply with a blocking order as soon as possible.

(2) LIMITATIONS.—A blocking order may not—

(A) require the deletion, destruction, or alteration of model weights, training data, source code, or records, which the frontier developer shall preserve as they exist on the date the order is entered;

(B) require the disclosure of model weights or source code to the United States or to any other person;

(C) restrict an activity necessary to comply with this Act, to evaluate, improve, or research the safety of the frontier model, to develop or apply a safeguard, or to

Last edited 9/29/26

secure the model or its weights, unless the court finds that the activity itself materially contributes to the risk identified; or

(D) require an affirmative action other than one necessary to give effect to a restriction imposed by the order or to notify a person affected by it.

(f) Appeal.—An order granting, denying, renewing, dissolving, or modifying a blocking order, or declining to do any of those things, is appealable as of right by any party only to the United States Court of Appeals for the District of Columbia Circuit, which shall decide the appeal at the earliest practicable time and in any event not later than [30] days after the notice of appeal is filed. The court of appeals shall review de novo each question of law, and shall set aside a finding of fact only if it is clearly erroneous. An appeal encompasses any order renewing, modifying, or continuing the order appealed from, and the court of appeals retains jurisdiction notwithstanding the expiration of that order. No court may restrict a frontier model pending appeal except by an order satisfying subsection (c).

(g) Classified and Sensitive Information.—

(1) SEALING AND PUBLICITY.—A party may file under seal any paper, or portion of a paper, containing classified information or information that may be redacted under section 9, and shall file a version redacted under section 9 not later than [7] days thereafter. The court shall make public each opinion and order under this section in a form redacted under section 9, and shall not make public, or require a party to make public, information the disclosure of which would materially assist a person in causing enumerated harm. Section 9(d) applies to information submitted in a proceeding under this section.

(2) CLASSIFIED INFORMATION; ACCESS BY CLEARED COUNSEL.—Sections 3, 4, 6, and 9 of the Classified Information Procedures Act (18 U.S.C. App. 3) apply to a proceeding under this section, treating the frontier developer as the defendant and the Agency as the United States except that a sanction under section 6(e) of that Act shall consist of denial or dissolution of the blocking order, exclusion of the information, or a finding against the Agency under subsection (c)(1). Except to the extent the court authorizes a deletion, summary, or statement under section 4 or 6(c) of that Act, classified information the Agency submits shall be made available, under a protective order, to counsel or expert for the developer. Each department or agency concerned shall expedite any clearance application for such counsel or expert, and may not deny or delay a determination on the ground that the individual represents or is retained by the developer.

(3) NO ORDER RESTING ON INFORMATION THE DEVELOPER HAS NOT SEEN.—The court may not enter or renew a blocking order, or decline to dissolve or modify one, on the basis of information not disclosed to the frontier developer under

Last edited 9/29/26

paragraph (2), and if the Agency declines to disclose and the court authorizes no substitution, the court shall not consider the information for any purpose. This paragraph does not apply to an order entered under subsection (d)(2), except that such an order may not be extended or renewed on the basis of such information.

SEC. 13. PERIODIC REVIEW; LIMITED RULEMAKING AUTHORITY.

(a) Annual Review of Coverage.—

(1) **IN GENERAL.**—Not less frequently than once each year, the Agency shall review the definitions of “frontier model”, “frontier developer”, and “large frontier developer” in section 3, with input from industry, academia, the open-source community, civil society organizations, and interested governmental entities, to determine whether those definitions continue to describe the developers and models whose activities present catastrophic or enumerated risk.

(2) **FACTORS.**—In conducting a review under paragraph (1), the Agency shall consider—

(A) thresholds used in international standards and in Federal law, guidance, and regulations for the management of catastrophic and enumerated risk, and whether to align a definition with such a threshold, through an adjustment under subsection (b)(1) or a recommendation under paragraph (4), to the extent that doing so is consistent with the purposes of this Act;

(B) input from the stakeholders described in paragraph (1);

(C) the extent to which a person can determine, before beginning to train or deploy a foundation model, whether the person would be covered, with an aim toward permitting that determination to be made as early as possible;

(D) the complexity of determining whether a person or model is covered, with an aim toward simpler determinations; and

(E) the extent to which coverage is verifiable by persons other than the developer.

(3) **PUBLICATION.**—Not later than one year after the date of enactment of this Act, and annually thereafter, the Agency shall publish the determination made under paragraph (1) and the basis for it.

(4) **SUBMISSION TO CONGRESS.**—Not later than 30 days after publishing a determination under paragraph (3), the Agency shall submit the determination to Congress, together with any recommendation for legislation the Agency considers

Last edited 9/29/26

warranted, including any recommendation concerning the interval specified in section 6(b) in light of the pace of advancement in the capabilities of frontier models.

(5) EFFECT OF REVIEW.—A determination under this subsection does not alter any definition in section 3. A definition in section 3 may be revised only by an Act of Congress, except that the Agency may adjust the numerical thresholds specified in paragraphs (13) and (16) of section 3 by regulation under subsection (b)(1), subject to subsections (c) and (d).

(b) Limited Rulemaking Authority.—The Agency may prescribe regulations under this Act only for the following purposes, and for no other purpose:

(1) Adjusting the numerical thresholds specified in paragraphs (13) and (16) of section 3, subject to subsections (c) and (d).

(2) Prescribing the forms, formats, and procedures for the documents required to be published, submitted, or transmitted under this Act, including secure methods of submission and procedures for the protection of information under section 9.

(3) Designating, for purposes of section 4(f), a published security standard that—

(A) is specific and capable of being independently audited under section 4(f)(3);

(B) is developed and maintained by a recognized standards body, industry consortium, or governmental authority with demonstrated expertise in information security;

(C) addresses, at a minimum, access control, insider threat, supply-chain integrity, incident detection and response, and the protection of high-value digital assets; and

(D) is already in use by one or more frontier developers or by other organizations protecting assets of comparable sensitivity.

(4) Recognizing accreditation bodies whose accredited auditors may conduct the independent audit required under section 4(f)(3).

(c) Procedural Safeguards.—

(1) NOTICE AND COMMENT.—Each regulation under this Act shall be prescribed in accordance with section 553 of title 5, United States Code, except as expressly modified by this section.

(2) COMMENT PERIOD.—The Agency shall provide a public comment period of not less than 90 days on a proposed regulation under subsection (b)(1) or (b)(3).

(3) CONSULTATION.—Before proposing or finalizing a regulation under subsection (b)(1) or (b)(3), the Agency shall solicit and place in the rulemaking record the written

Last edited 9/29/26

views of an advisory committee, established under chapter 10 of title 5, United States Code, composed of representatives of frontier developers, other developers of foundation models, the open-source community, academic institutions, civil society organizations, qualified independent evaluators, and Federal agencies and federally funded research and development centers with relevant technical expertise. The views of the advisory committee do not bind the Agency.

(4) FREQUENCY.—The Agency may not adjust a threshold under subsection (b)(1), or change a designation under subsection (b)(3), more than once in any 24-month period, except that this paragraph does not apply to any adjustment that expands the class of frontier developers or frontier models subject to this Act.

(5) CONGRESSIONAL TRANSMITTAL.—The Agency shall transmit each final regulation prescribed under this Act, together with the findings supporting it, to Congress on or before the date the regulation is published in final form. The transmittal required by this paragraph is in addition to the requirements of chapter 8 of title 5, United States Code (commonly known as the “Congressional Review Act”).

(d) Substantive Limitations.—

(1) ADJUSTMENT CRITERIA.—The Agency shall set each threshold adjusted under subsection (b)(1) at the level best calculated to ensure that the definitions at issue encompass the developers and models whose activities present catastrophic or enumerated risk and do not encompass those whose activities do not, taking into account advances in algorithmic and hardware efficiency, changes in the cost of training foundation models, inflation, and the anticipated effect of the adjustment on the persons subject to this Act. Each such regulation shall be accompanied by written findings, supported by substantial evidence in the rulemaking record, demonstrating compliance with this paragraph.

(2) FLOOR.—No regulation prescribed under this Act may—

(A) waive, reduce, delay, or otherwise diminish an obligation imposed by section 4, 5, 6, 7, 8, 9, 10, 11, or 12;

(B) exempt a frontier developer, a category of frontier developers, a frontier model, or a category of frontier models from a requirement of this Act, except by an adjustment under subsection (b)(1) that satisfies paragraph (1) of this subsection;

(C) narrow the definition of “catastrophic risk” or “enumerated risk” in section 3, or any category of risk described in that definition; or

Last edited 9/29/26

(D) impose on a qualified independent evaluator a requirement that conflicts with section 8(a) or section 8(k)(3).

(3) **NARROWING ADJUSTMENTS.**—A regulation that would materially narrow the class of frontier developers or frontier models subject to this Act, as compared to the class subject to this Act on the date of enactment, shall not take effect until 180 days after the date on which the Agency transmits the regulation and the findings required by paragraph (1) to Congress.

(e) **Judicial Review.**—Any person aggrieved by a regulation prescribed under this Act, including a State, may obtain judicial review of the regulation under chapter 7 of title 5, United States Code.

(f) **Effective Date; Prospective Application.**—A regulation prescribed under subsection (b)(1) shall take effect not earlier than 180 days after publication in final form, and a regulation prescribed under this Act shall apply only with respect to obligations arising on or after its effective date.

(g) **No Other Authority.**—Except as provided in subsection (b), this Act confers no authority to prescribe regulations. No guidance, interpretive rule, statement of policy, or other agency action may impose on any person an obligation that is not imposed by this Act or by a regulation prescribed under subsection (b).

SEC. 14. RELATIONSHIP TO STATE LAW.

(a) **Covered Functions.**—In this section, the term “covered function” means any of the following:

(1) **PUBLIC DISCLOSURE OF ENUMERATED RISK INFORMATION.**—The disclosure by an artificial intelligence developer, to a State, to another governmental entity, or to a member of the public, of information of the kind required to be published or submitted by section 4, 5, or 6.

(2) **THIRD-PARTY EVALUATION AND VERIFICATION FOR ENUMERATED RISK ASSESSMENT.**—The assessment, audit, evaluation, certification, or verification, by a person other than the artificial intelligence developer, of the matters addressed by section 8, including any requirement that a developer engage an evaluator, obtain an evaluation, or provide access to a model, model weights, training data, source code, evaluation results, or other technical materials for purposes of such an assessment, audit, evaluation, or verification.

(3) **INCIDENT REPORTING.**—The reporting by an artificial intelligence developer to a State, agency, regulator, or other governmental entity, of an incident of the kind required to be reported by section 7.

Last edited 9/29/26

(4) **SUSPENSION OF MODELS.**—The suspension, by a State or a political subdivision of a State, whether by statute, regulation, order, judicial decree, or other action, of the development, deployment, or internal use of an artificial intelligence model, on the ground that such development, deployment, or internal use presents an imminent enumerated risk or a substantially similar risk. This paragraph shall have no effect unless section 12 is enacted.

(b) **Preemption.**—Except as provided in subsection (c), and notwithstanding the date of its enactment or adoption—

(1) no State or political subdivision of a State may adopt, maintain, or enforce a law, regulation, order, or other requirement that imposes an obligation on an artificial intelligence developer with respect to a covered function described in paragraph (1), (2), or (3) of subsection (a).

(c) **Preservation of State Authority.**—Nothing in this Act shall be construed to affect the authority of a State or political subdivision of a State to—

(1) adopt or enforce a law, regulation, order, or other requirement, whether generally applicable or addressed to artificial intelligence systems or specifically targeting artificial intelligence developers, that does not pertain to a covered function, including a requirement governing the procurement or use of artificial intelligence systems by a State or local government;

(2) regulate the use or deployment of artificial intelligence systems by a person other than the developer of that system, including through consumer protection, child safety, civil rights, contract, labor, criminal, tort, or privacy law;

(3) adopt or enforce a law, regulation, order, or other requirement specifically relating to the protection of minors from harms arising from the use of artificial intelligence systems, including requirements addressing sexually explicit content, content promoting self-harm, content facilitating exploitation, age verification, parental controls, or similar matters;

(4) adopt or enforce a law, regulation, order, or other requirement protecting an employee or independent contractor who reports a violation of law, or a risk to public health or safety, relating to artificial intelligence, including a protection that supplements section 10; or

(5) apply its common law of tort, contract, or property to a claim arising from injury caused by an artificial intelligence model or by the conduct of an artificial intelligence developer. A duty arising under the common law of a State is not an obligation with respect to a covered function for purposes of subsection (b).

Last edited 9/29/26

(d) No Field Preemption; No Implied Displacement.—Nothing in this Act shall be construed to occupy the field of the regulation of artificial intelligence systems or artificial intelligence developers, or to displace State law by implication. Except as expressly provided in sections 8(d)(3), 8(n), and 10(c), no provision of this Act other than this section preempts State law.

(e) Narrow Construction.—This section shall be construed narrowly. Any ambiguity concerning the scope of preemption under this section shall be resolved in favor of preserving the authority of the States.

(f) No Safe Harbor.—Compliance with this Act shall not be construed to confer immunity, to establish a safe harbor, or to create a presumption against liability, under any State law that is not preempted by this section.

(g) Effect on Related State Requirements.—A State law is not preempted by this section merely because compliance with it results in the disclosure of information also disclosed under this Act, if the law does not impose an obligation with respect to a covered function.

(h) Definitions.—In this section:

(1) ARTIFICIAL INTELLIGENCE DEVELOPER.—The term “artificial intelligence developer” means an entity that builds, designs, codes, produces, trains, or owns an artificial intelligence model for internal use or for use by a third party, and does not include an entity that is solely a deployer of an artificial intelligence model. The term is not limited to a frontier developer.

(2) CATASTROPHIC RISK.—The term “catastrophic risk,” and the description of the matters addressed by a section of this Act, are not limited by the thresholds stated in section 3, and apply to any artificial intelligence model and any artificial intelligence developer.

(3) ENUMERATED RISK.—The term “enumerated risk,” and the description of the matters addressed by a section of this Act, are not limited by the thresholds stated in section 3, and apply to any artificial intelligence model and any artificial intelligence developer.

SEC. 15. GENERAL PROVISIONS.

(a) Effective Date.—This Act takes effect on the date of enactment.

(b) Severability.—Except as provided in subsection (c), if a provision of this Act, or the application of a provision to a person or circumstance, is held invalid, the remainder of this Act, and the application of the provision to other persons or circumstances, shall not be affected.

Last edited 9/29/26

(c) Nonseverability of Preemption.—Section 14 shall have no force or effect if section 4, section 5, section 6, section 7, section 8, section 9, section 10, or section 11 is held invalid in substantial part, as determined by a final judgment of a court of competent jurisdiction. A paragraph of section 14(a) shall have no force or effect if the section of this Act that imposes the corresponding Federal requirement is held invalid in substantial part. Section 14 shall also have no force or effect if the obligations imposed by the sections identified in this subsection are otherwise rendered unenforceable in substantial part. It is the intent of Congress that State authority be displaced under section 14 only so long as the Federal requirements of this Act remain in force.

(d) Cumulative Obligations.—The duties imposed by this Act are cumulative with any other duty imposed by law, do not relieve any person of any other duty, and do not limit any right or remedy available under other law.

(e) Federal Contracts and Classified Information.—A frontier developer is excused from a requirement of this Act to publish information to the extent that publication would directly conflict with the express term, imposed for reasons of national security or public safety, of a contract between the frontier developer and a department or agency of the Federal Government, or would disclose information classified under Executive order or statute. This subsection does not excuse a frontier developer from any requirement to submit or transmit information, consistent with any applicable handling requirements, to the Agency, to the [Enforcement Authority], or to a qualified independent evaluator holding the necessary authorization. A frontier developer that withholds information under this subsection shall so state in the document in which the information would otherwise appear and shall identify the contract or classification authority relied upon in a submission to the Agency, and shall publish the withheld information promptly after the conflict or classification ceases to apply.

(f) Authorization of Appropriations.—There are authorized to be appropriated to the Agency and the [Enforcement Authority] such sums as may be necessary to carry out this Act.

(g) Statements of Opinion.—Nothing in this Act shall be construed to require any person to express, adopt, or affirm any opinion or belief. A frontier developer may include, in any document published or submitted under this Act, its own opinion on any matter addressed by the document, including its judgment that a frontier model does not present an enumerated risk in one or more respects, if the document also contains the factual information this Act requires.