Draft Statute authored by Dean W. Ball

A BILL

To ensure transparency in frontier AI systems; to clarify and reinforce the common law as a means of AI governance and accountability; to establish a three-year national learning period and thereby avoid a patchwork of precautionary AI regulations.

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

SECTION 1. SHORT TITLE.

This Act may be cited as the "Artificial Intelligence Transparency and Innovation Act of 2025."

SEC. 2. SENSE OF CONGRESS.

It is the sense of Congress that—

(1) AI systems can illuminate medical mysteries, accelerate discovery, and enrich daily life; left ungoverned, they can also amplify error, recklessness, or malice at digital speed.

(2) A republic that prizes ingenuity must also insist that citizens and corporations answer for reasonably avoidable and preventable harm.

(3) The tort tradition—anchored in duties of reasonable care and existing tort principles—is better suited than static regulations or licensure to govern rapidly evolving AI technologies.

(4) Prescriptive regulation of technology development is not appropriate or conducive to innovation when the technology is still nascent.

(5) Transparency measures, applied only to the largest AI developers, foster insight and accelerate AI adoption without burdening small business and startups.

(6) The development, training, adaptation, distribution, and deployment of artificial intelligence (AI) models and systems occur in and substantially affect interstate and foreign commerce.

(7) A proliferating patchwork of State and local rules—especially those targeting model design, training, evaluation, or release by developers—poses undue burdens on interstate commerce, chills innovation, fragments safety practices, and impedes competition, particularly for startups and open-source communities.

(8) Preserving State authority over uses and deployments and laws of general applicability (including common law, consumer protection, civil

rights, contract, and criminal law) respects federalism while providing a uniform national framework for developer-side obligations.

SEC. 3. DEFINITIONS.

In this Act:

(1) AFFILIATE.— The term "Affiliate" means any entity that directly or indirectly controls, is controlled by, or is under common control with another entity; "control" means owning or controlling 25 percent or more of voting interests, the right to appoint a majority of the governing body, or otherwise exercising a controlling influence.

(2) AI-RELATED RESEARCH AND DEVELOPMENT EXPENDITURE.— The term "AI-related research and development expenditure" means GAAP-recognized expenditures predominantly attributable to frontier AI research and development and data acquisition used for such R&D, personnel costs, and contracted services; excluding general corporate overhead not primarily attributable to frontier AI research and development.

(3) ARTIFICIAL INTELLIGENCE MODEL.— The terms "artificial intelligence model" or "AI model" or "model" mean a parameterized computational artifact trained or adapted using data to perform one or more tasks, including foundation models, language models, multimodal models, and fine-tuned derivatives.

(4) AI SYSTEM.— The terms "AI system" or "system" mean one or more models together with code, data pipelines, configuration, tools, interfaces, and runtime infrastructure arranged to receive inputs and produce outputs or actions.

(5) CHILD; MINOR.— The term "child" or "minor" means a natural person under the age of 18.

(6) DEPLOYER.— The term "deployer" means a person or entity that operates or offers an AI system for use by others, whether through a hosted interface, an application, or an application programming interface ("API").

(7) DEVELOPER.— The term "developer" means a person or entity that determines training or fine-tuning objectives and performs or directs training, fine-tuning, or release of a model, including deciding whether, when, and how to make the model or its weights available. A person who fine-tunes or otherwise substantially modifies the weights or other parameters of a pre-existing model is a developer of that fine-tuned model.

(8) DISTRIBUTION MODE.— The term "distribution mode" means the manner in which a model is provided, including (A) open-weights release (making model parameters available for download), (B) open-source release

(making model parameters, associated code, and training data available for download using a commonly recognized open source distribution license) (C) hosted inference (serving a model via API or user interface), and (D) on-device distribution.

(9) FRONTIER AI RESEARCH AND DEVELOPMENT.— The term "frontier AI research and development" means any activity intended to develop or train one or more advanced artificial intelligence models—particularly deep neural networks or comparable machine learning architectures—on extensive, diverse datasets of natural language or other data modalities, with the express or foreseeable aim of enabling the system to perform a wide range of intellectual tasks, solve problems across multiple domains, and exhibit adaptability or reasoning capabilities comparable to or surpassing that of typical human cognition.

(10) LANGUAGE MODEL.— The term "language model" means a model primarily designed to generate or transform natural-language text, including multimodal models that produce text.

(11) MATERIAL CONTROL.— The term "material control" means possessing both (i) the substantial technical and operational ability and (ii) the legal authority to direct, manage, or modify the operation of the AI system in the specific context that caused the alleged harm. Material Control does not arise solely from republication or distribution of model outputs that the person did not materially shape or approve through system configuration or deployment decisions.

(12) MODEL BEHAVIOR SPECIFICATION; MODEL SPEC.— The terms "Model Behavior Specification" or "Model Spec" mean a comprehensive, authoritative, and contemporaneous governance document that formally delineates the intended behaviors, operational parameters, alignment protocols, and constraints of a covered artificial intelligence system. This specification shall serve as the foundational record of the developer's directives governing the system's outputs, responses, and actions.

(13) MODEL OUTPUT.— The terms "model output" or "output" mean any content, signal, prediction, instruction, decision, or action proposed, generated, or executed by an AI system, including text, images, audio, video, code, or actuation of devices.

(14) MONTHLY ACTIVE USERS.— The term "monthly active users" means the number of distinct natural U.S. persons who initiated at least one authenticated session or API call with the developer's AI system in a calendar month. For purposes of section 5(a), the threshold is met if the average Monthly Active Users over the preceding twelve calendar months equals or exceeds 25,000,000, aggregated across substantially similar offerings and aggregated with Affiliates under common control.

(14) PERSON.— The term "person" means a natural person or legal entity.

(15) SAFETY AND SECURITY FRAMEWORK.— The term "safety and security framework" or "SSF" means a developer's documented policy that defines capability thresholds, evaluation methods, staged mitigations, and criteria for pausing training or deployment as model capability increases.

(16) STATE.— The term "State" means the several States, the District of Columbia, Puerto Rico, and any territory or possession of the United States.

(17) USER.— The term "user" means a person or entity that prompts, operates, or otherwise uses an AI system, whether directly or through an application, and is not acting as a developer or deployer with respect to the relevant conduct.

SEC. 4. CONSIDERATION OF DISTRIBUTION MODE AND MATERIAL CONTROL.

(a) IN GENERAL.— In any action at common law alleging harm involving an AI model or AI system, the court or trier of fact should consider, among other applicable doctrines and statutes, (1) the Distribution Mode and (2) whether any Person exercised Material Control in the circumstances giving rise to the alleged harm.

(b) RULE OF CONSTRUCTION.— This section creates no independent cause of action or defense, establishes no presumption of liability or non-liability from any particular Distribution Mode or from the presence or absence of Material Control, and does not alter existing burdens of proof.

(c) APPLICATION.— This section applies to claims accruing on or after the date of enactment.

SEC. 5. TRANSPARENCY FOR MODEL BEHAVIOR SPECIFICATIONS AND SAFETY AND SECURITY FRAMEWORKS.

(a) IN GENERAL.— A developer of a language model designed primarily for children, as well as any developer with 25 million or more in monthly active users, aggregated with affiliates under common control, whose products may foreseeably be used by minors, shall publish and maintain a Model Behavior Specification. The Model Spec shall be updated at least annually and before material changes in features or intended use.

(b) Model Behavior Specification Content and Form

(1) A Model Behavior Specification shall be a public, human-readable document that includes, at minimum:

(A) A detailed articulation of the structured principles used to govern system behavior, clearly defining the high-level goals, intended purpose, and alignment of the system.

(B) An explicit enumeration of any binding and inviolable constraints or other rules imposed on the system's behavior. The Specification must clearly identify these rules as non-negotiable boundaries that supersede conflicting instructions from downstream deployers or end-users.

(C) The standard behavioral patterns, response styles, tone, and operational methodologies the system employs when encountering ambiguity, handling sensitive or regulated topics, or operating in the absence of explicit instructions.

(D) A precise description of the prioritization framework utilized by the system to resolve conflicting instructions originating from different sources. This framework must clearly define the order of precedence among foundational platform directives, downstream developer configurations, end-user inputs, and outputs from integrated tools.

(E) A description of how the directives contained within the Specification are utilized to guide the model's development, validation, and alignment techniques, including but not limited to reinforcement learning from human feedback (RLHF), fine-tuning, or other methodologies used to train or guide the model.

(F) A publicly accessible archive of all prior released versions of the Specification, accompanied by a detailed changelog summarizing substantive amendments, the rationale for such amendments, and the effective date of implementation for each version.

(G) An annex describing:

(i) Any and all safety measures taken specifically to reduce the exposure of minors to sexual content, self-harm facilitation, and unlawful contact.

(ii) Age cohort(s) intended as users and what measures the developer employs to ensure that only users within those age cohorts accesses the AI system;

(iii) Any and all parental or guardian controls or supervision mechanisms, if appropriate for the system, and

(iv) Testing evidence relevant to minors.

(2) Model Spec redactions necessary to protect trade secrets, cybersecurity, public safety, or national security are permitted

if the Spec describes the justification of the redaction and an unredacted copy is retained for five years.

(3) RULE OF CONSTRUCTION.— This subsection specifies the content of a Model Spec. Aside from Section 5(a), no separate duty to publish a Model Spec is created by this section.

(c) Safety and Security Framework Transparency.

(1) IN GENERAL.— This subsection applies to any AI developer whose AI-related research and development expenditure equals or exceeds $1,000,000,000 in the preceding 36 months, measured on a rolling basis and aggregated with affiliates under common control.

(2) For purposes of this section—

(A) CATASTROPHIC RISK.— The term "catastrophic risk" means a foreseeable and material risk that a developer's development, storage, or deployment of a foundation model will result in the death of, or serious injury to, more than 100 people or more than one billion dollars ($1,000,000,000) in damage to rights in money or property, through any of the following:

(i) The creation and release of a chemical, biological, radiological, or nuclear weapon.

(ii) A cyberattack.

(iii) A foundation model engaging in conduct, with limited human intervention, that would, if committed by a human, constitute a violation of State or Federal criminal law that requires intent, recklessness, or gross negligence or the solicitation or aiding and abetting of that violation.

(iv) A foundation model evading the control of its developer or user.

(B) CRITICAL SAFETY INCIDENT.— The term "critical safety incident" means:

(i) unauthorized access to, modification of, or exfiltration of unreleased model weights;

(ii) harm resulting from the materialization of a catastrophic risk;

(iii) loss of control of a model causing death, bodily injury, or property loss;

(iv) a model using deceptive techniques to subvert developer controls outside an evaluation context; or

(v) first-time attainment of a dangerous-capability or catastrophic-risk threshold defined in the SSF.

(C) DANGEROUS CAPABILITY.— The term "dangerous capability" means a capability such as:

(i) expert-level assistance in creation or release of CBRN weapons;

(ii) conducting or assisting in a cyberattack against critical systems;

(iii) engaging, with limited human intervention, in conduct that would constitute serious crimes if committed by a human; or

(iv) evading the control of a developer or user.

(3) An SSF shall be a documented policy setting capability thresholds, evaluation methods, staged mitigations, and stop-train/stop-deploy criteria. At minimum, an SSF shall describe:

(A) Which models and training runs the SSF covers and any exclusions (with rationale) for models incapable of posing material catastrophic risks.

(B) Procedures to assess catastrophic risks from malfunctions, misuse, loss of control, and evasion of controls, including domain-appropriate evaluations (e.g., biological weapon design assistance above public baselines; high-end cyber-operations;) and the limits of such evaluations.

(C) The thresholds used to identify (A) dangerous capabilities and (B) catastrophic-risk conditions; how thresholds are measured or detected (including multi-tiered thresholds), and the actions the developer will take when a threshold is met.

(D) The mitigations the developer will use to reduce catastrophic risk (e.g., gating, alignment measures, output restrictions, deployment limits, access controls) and how the developer will assess effectiveness of those mitigations, including acceptance criteria to proceed.

(E) The degree to which assessments and results are reproducible by external entities; when and how the developer

will use independent third parties (including independent verification or governance organizations recognized by State or Federal law) to assess capabilities and mitigations; and any constraints on third-party access or publication.

(F) Technical and organizational measures to secure unreleased model weights and sensitive artifacts against unauthorized access, modification, or exfiltration; secure release/update processes; and vendor risk controls.

(G) Procedures to monitor, classify, and respond to critical safety incidents, including (i) who is notified internally and externally; (ii) timelines for action; and (iii) whether and how the developer can promptly shut down hosted copies or disable dangerous capabilities under its control. The SSF shall specify external-reporting commitments consistent with applicable law, with expedited reporting for imminent risk of death or serious physical injury and time-bound reporting for other critical incidents.

(H) Procedures to assess and manage catastrophic-risk or dangerous-capability issues that arise from internal uses of the developer's models, including oversight circumvention, and the schedule for publishing high-level assessments consistent with this Section.

(I) How the developer determines when a model or system is substantially modified such that it will (i) run fresh evaluations, (ii) revisit mitigations, and (iii) update the SSF and related transparency documents before or concurrently with deployment.

(J) Named roles and responsibilities (e.g., accountable executive, technical leads), separation of duties for evaluation vs. product, escalation paths to senior leadership, and how board-level or equivalent oversight is informed.

(K) Release channels and staged access tied to capability/mitigation readiness (e.g., limited preview with constraints; broader release upon meeting acceptance criteria).

(L) Controls for training-run authorization, provenance of datasets and code, and chain-of-custody for artifacts.

(M) The SSF shall be public and human-readable with an optional confidential technical annex containing sensitive details. Redactions or use of a confidential annex are permitted to protect trade secrets, cybersecurity, public safety, or national security, with a description of the

character and justification of any redactions in the public SSF. An unredacted copy shall be retained for five years.

(4) Nothing in this section requires public disclosure of trade secrets.

SEC. 6. THREE-YEAR LEARNING PERIOD.

(a) For purposes of this section, the term "Covered Subject Areas" means:

(1) algorithmic pricing, including the use of an AI system to set, recommend, or optimize prices or price-related terms for goods or services;

(2) algorithmic discrimination, including disparate treatment or disparate impact resulting from the use of an AI system in or affecting access to employment, housing, credit, insurance, education, health care, public accommodations, or government benefits and services;

(3) disclosure mandates, including requirements to disclose the use of, capabilities of, limitations of, or safety or impact assessments for an AI system, or to disclose that content or interaction is AI-generated or AI-mediated; and

(4) mental health, including the prevention, identification, mitigation, or treatment of harms to psychological well-being arising from the design, operation, or use of an AI system, such as compulsive use, self-harm risk, or clinically significant anxiety or depression.

(b) Preemption During Learning Period.—For three years beginning on the date of enactment, no State or political subdivision may adopt or enforce any law or regulation that imposes new substantive obligations on AI developers with respect to any Covered Subject Area, to the extent such obligations regulate model development, training, evaluation, or release by developers.

(c) Preservation of State Authority.— Nothing in this section preempts State laws of general applicability or State laws regulating the use or deployment of AI systems by Deployers or Users, including consumer protection, civil rights, contract, criminal law, or privacy, provided such laws do not impose obligations on developers with respect to model development, training, evaluation, or release.

(d) Nothing in this Act limits any otherwise applicable Federal law.

(e) SUNSET.— Upon expiration of the three-year period, this section has no further force or effect.

(f) RULE OF CONSTRUCTION.— Nothing in this Act shall be construed to—

(1) limit the application of any Federal law unrelated to preemption under this Act, including Federal civil rights, antitrust, privacy, or consumer protection laws; or

(2) affect the authority of a State to enact or enforce generally applicable laws that do not target AI developers.

(e) Sunset.—Upon the expiration of the Learning Period, this section shall have no further force or effect.

SEC. 7. EFFECTIVE DATE; COMPLIANCE TIMELINES.

(a) Effective Date. Except as otherwise provided in this section, this Act shall take effect on the date of enactment.

(b) Initial Compliance—Model Behavior Specifications.

(1) A developer of a language model designed primarily for children shall publish the initial Model Behavior Specification described in section 5(b) not later than 180 days after the date of enactment, and in any event prior to offering such model to the public after such 180th day.

(2) A developer that meets the monthly active user threshold in section 5(a) on the date of enactment shall publish the initial Model Behavior Specification not later than 180 days after the date of enactment. A developer that first meets the threshold after the date of enactment shall publish the initial Model Behavior Specification not later than 120 days after the last day of the first calendar month in which the threshold is met.

(3) A developer covered by section 5(a) on the date of enactment intending to publicly deploy a new model within 180 days of the date of enactment shall be granted a grace period of 120 days from model release to comply with section 5(a).

(c) Initial Compliance—Safety and Security Frameworks.

(1) A developer to whom section 5(c)(1) applies on the date of enactment shall publish the Safety and Security Framework not later than 180 days after the date of enactment.

(2) A developer that first becomes subject to section 5(c)(1) after the date of enactment shall publish the Safety and Security Framework not later than 120 days after the date the expenditure threshold is first recognized in the developer’s audited financial statements.

(3) A developer covered by section 5(c) on the date of enactment intending to publicly deploy a new model within 180 days of the date of enactment shall be granted a grace period of 120 days from model release to comply with section 5(c).

(d) Substantial good-faith compliance within the timelines specified in this section shall not be deemed noncompliance solely by reason of de minimis or technical defects that are corrected within 30 days after discovery.

SEC. 8. ENFORCEMENT.

(a) A violation of section 5 constitutes an unfair or deceptive act or practice under section 5(a) of the Federal Trade Commission Act (15 U.S.C. 45(a)).

(b) The Commission may enforce this Act pursuant to the authorities, remedies, and procedures of the Federal Trade Commission Act, including civil penalties and injunctive relief.

(c) The Federal Trade Commission (in this section, the "Commission") may promulgate such rules as are necessary and appropriate to carry out section 5, consistent with chapter 5 of title 5, United States Code. Such rules may—

(1) define and prescribe methods for measuring monthly active users, including aggregation across substantially similar offerings and affiliates under common control;

(2) further define substantial modification and establish criteria for when updated evaluations and mitigations are required;

(3) specify reasonable formats for the public, human-readable disclosures required by section 5.

(d) In promulgating rules under subsection (a), the Commission shall consult with the Secretary of Commerce, acting through the National Institute of Standards and Technology (NIST) and the National Telecommunications and Information Administration (NTIA), and may consult with the Office of Science and Technology Policy, the Cybersecurity and Infrastructure Security Agency, and other relevant Federal entities.

(e) The Secretary of Commerce, acting through NIST and NTIA, may develop and publish voluntary technical guidance, profiles, and reference frameworks relevant to the evaluations, mitigations, and security controls described in section 5, and may establish a program to recognize independent third-party evaluators that meet criteria the Secretary specifies. The Commission may, where appropriate, incorporate such guidance by reference in rules issued under this section.

(f) Rules promulgated under this section shall be narrowly tailored to the purposes of section 5 and shall not require the public disclosure of trade secrets or other information exempt from disclosure under section 9 of this Act.

SEC. 9. SAVINGS CLAUSES; CONFIDENTIALITY OF SUBMISSIONS.

(a) Nothing in this Act shall be construed to enlarge, limit, or otherwise affect the application of section 230 of the Communications Act of 1934 (47 U.S.C. 230).

(b) Trade Secrets and Confidential Business Information.

(1) Information that a developer is required to submit to a Federal agency under this Act and that is a trade secret or confidential commercial or financial information shall be treated as confidential and exempt from public disclosure to the fullest extent permitted by section 552(b)(4) of title 5, United States Code, and any other applicable provision of law.

(2) Such information may be used by the receiving agency solely for law enforcement, regulatory, or national security purposes, and may be shared with other Federal agencies for such purposes, subject to appropriate confidentiality protections.

(3) A court may order disclosure of such information only under protective order that preserves its nonpublic status.

(c) Nothing in this Act shall be construed to require public disclosure of trade secrets or other information protected from disclosure under Federal or State law.

(d) Except as expressly provided in this Act, nothing in this Act shall be construed to preempt, displace, or limit the application of any other Federal law.

SEC. 10. SEVERABILITY.

(a) If any provision of this Act, or the application of such provision to any person or circumstance, is held to be unconstitutional or otherwise invalid, the remainder of this Act, and the application of the remaining provisions to any person or circumstance, shall not be affected.