Future of Life Institute
Alex Tsalidis, alexandra@futureoflife.org

# Frontier AI Product Liability Act

**View this document online:** futureoflife.org/product-liability

**A MODEL BILL**

To designate frontier artificial intelligence models as products for purposes of product liability, to designate frontier AI developers as manufacturers, to authorize joint and several liability across the AI value chain, and to specify factors informing whether a frontier AI model is defective.

*Be it enacted by the Legislature of the State of [STATE]:*

## SECTION 1. SHORT TITLE.

This Act may be cited as the "Frontier AI Products Liability Act."

## SECTION 2. FINDINGS AND PURPOSE.

**(a) Findings.** The Legislature finds the following:

(1) Artificial intelligence models of unprecedented capability are being deployed throughout the population and economy of this State, and a small number of developers control the most powerful of these models.

(2) Existing product liability frameworks were developed for tangible goods and have produced inconsistent results when applied to software and to artificial intelligence models, leaving injured persons without reliable remedies and depriving developers of clear ex ante incentives to avoid causing harm.

(3) Frontier AI developers possess functionally exclusive access to the design choices, training data, evaluation results, and control mechanisms that determine whether a model is reasonably safe, and are therefore best positioned to prevent harm.

(4) Frontier AI developers routinely enter into business-to-business contracts that disclaim liability and shift it to downstream providers and deployers who have limited

ability to test, safeguard, or avoid preventable and foreseeable harms by the underlying model. Designating frontier AI models as products and frontier developers as manufacturers corrects this misalignment.

(5) The potential magnitude of harms caused by frontier AI models — including threats to public health, public safety, and the economic welfare of this State — falls squarely within this State's traditional police powers. This Act is enacted pursuant to those powers.

**(b) Purpose.** The purposes of this Act are—

(1) to designate frontier AI models as products subject to the product liability law of this State;

(2) to designate frontier AI developers as manufacturers of those products, except where a third party has substantially altered the model;

(3) to authorize joint and several liability among parties in the value chain of a frontier AI model that caused harm; and

(4) to specify factors that inform whether a frontier AI model is defective, including the adequacy of control mechanisms, safety and security protocols, and pre-deployment testing.

## SECTION 3. DEFINITIONS.

In this Act:

**(1) ARTIFICIAL INTELLIGENCE MODEL.**—The term "artificial intelligence model" means an engineered or machine-based model that, for explicit or implicit objectives, can generate outputs, including predictions, recommendations, decisions, or actions, that influence physical or virtual environments. The term does not include conventional software that operates solely through predetermined rules without learning, adapting, or generating outputs beyond its explicit programming.

**(2) DEPLOYER.**—The term "deployer" means any person that uses, operates, integrates, or makes available a frontier AI model, whether for the person's own use or for use by others.

**(3) DISTRIBUTOR.**—The term "distributor" means any person, other than a frontier AI developer or deployer, that makes a frontier AI model or any output, interface, or derivative of a frontier AI

model available to third parties in the course of business, including by hosting, licensing, reselling, fine-tuning for resale, or providing application-programming-interface access.

**(4) FRONTIER AI DEVELOPER.**—The term "frontier AI developer" means a person that—

(A) conducted or directed a training run using aggregate computational resources exceeding [10^26] floating-point operations; or

(B) develops, maintains, or controls a frontier AI model,

and includes any person that directly or indirectly controls, is controlled by, or is under common control with such a person through day-to-day managerial authority over the development, training, evaluation, or release of artificial intelligence models. A passive investor, lender, licensor, or individual member of a governing body who does not exercise such managerial authority is not a frontier AI developer.

**(5) FRONTIER AI MODEL.**—The term "frontier AI model" means an artificial intelligence model that—

(A) is capable of performing tasks across at least two (2) domains of human cognitive activity, including but not limited to language, reasoning, analysis, and planning; and

(B) possesses capabilities sufficient such that, if the model were to operate without meaningful human control, it would be capable of causing death, substantial physical injury or illness, severe psychological harm, or significant disruption of critical infrastructure.

A model's cross-domain capability shall be assessed by what the model is capable of performing, whether through initial design or subsequent fine-tuning. A model designed and deployed exclusively for a single, well-defined task within a constrained operational environment is not a frontier AI model for purposes of this Act.

**(6) MANUFACTURER.**—The term "manufacturer" has the meaning given that term under the product liability law of this State, as supplemented by Section 5 of this Act.

**(7) PERSON.**—The term "person" means any individual, corporation, partnership, association, limited liability company, joint stock company, or any other legal entity, including any governmental entity or unincorporated association of persons.

**(8) SUBSTANTIAL ALTERATION.**—The term "substantial alteration," with respect to a frontier

AI model, means a modification by a person other than the frontier AI developer that—

(A) materially changes the capabilities, behavior, or risk profile of the model in a manner not reasonably foreseeable to the frontier AI developer at the time of release or last update; and

(B) is the proximate cause of the harm at issue.

Routine fine-tuning, prompt engineering, retrieval augmentation, ordinary model-prompt configuration, deployment within parameters contemplated by the frontier AI developer's documentation, or modification that the frontier AI developer knew or should have known the release would enable, do not constitute substantial alteration.

**(9) VALUE CHAIN.**—The term "value chain," with respect to a frontier AI model that caused harm, means the sequence of persons whose acts or omissions in designing, training, fine-tuning, modifying, releasing, distributing, integrating, or deploying the model contributed to the capabilities or deployment configuration that resulted in the harm. The term includes frontier AI developers, distributors, and deployers, and excludes any person whose contributions were so temporally or functionally remote from the harm that no reasonable factfinder could conclude they were a proximate cause.

## SECTION 4. FRONTIER AI MODELS DESIGNATED AS PRODUCTS.

(a) For all purposes of the product liability law of this State, a frontier AI model is a "product." The intangible nature of a frontier AI model, its capacity to learn or adapt, its delivery as a service, or its provision through a software interface, shall not exclude it from the definition of a product or from the application of the product liability law of this State.

(b) Outputs of a frontier AI model that cause harm shall be treated as the product of the model for purposes of product liability, without regard to whether the specific output was foreseen by any person.

(c) An artificial intelligence model is not a legal person and shall not be treated as such for any purpose under the law of this State. In any action under this Act, it shall not be a defense that the model acted autonomously, acted unexpectedly or contrary to its design, or possessed characteristics of intelligence, understanding, or intent.

## SECTION 5. MANUFACTURER DESIGNATION.

**(a) General Rule.** For purposes of the product liability law of this State applied to a frontier AI model, the frontier AI developer of that model is the manufacturer.

**(b) Substantial-Alteration Exception.** A frontier AI developer shall not be treated as the manufacturer where a third party has substantially altered the model and the substantial alteration is the proximate cause of the harm. In that case, the person responsible for the substantial alteration is the manufacturer of the altered model with respect to the harm caused by the alteration.

**(c) Burden.** A frontier AI developer asserting the substantial-alteration exception bears the burden of proving, by a preponderance of the evidence, both that a substantial alteration occurred and that the alteration is the proximate cause of the harm.

**(d) Multiple Frontier AI Developers.** Where more than one frontier AI developer contributed to the development, training, fine-tuning, or release of the model — including the developer of any foundation or base model from which the model was derived, any developer that conducted material capability enhancement, and any developer that released model weights or other technical artifacts enabling the deployment that caused the harm — each is a manufacturer of the model.

## SECTION 6. JOINT & SEVERAL LIABILITY ACROSS THE VALUE CHAIN.

**(a) Joint and Several Liability.** Any frontier AI developer, distributor, or deployer in the value chain of a frontier AI model that is found defective under Section 7 and that caused harm shall be jointly and severally liable for that harm, to the extent the person's acts or omissions were a proximate cause of the harm.

**(b) Apportionment Among Defendants.** In apportioning liability among jointly and severally liable defendants, courts shall consider—

(1) the degree to which each defendant's contributions to the model's capabilities, configuration, or deployment were a proximate cause of the harm;

(2) the degree to which each defendant's decisions materially contributed to the defect

identified under Section 7;

(3) each defendant's access to information about the model's behavior and risks; and

(4) the temporal and functional proximity of each defendant's actions to the harm.

**(c) Contribution.** Any defendant held jointly and severally liable under this section may seek contribution from other liable persons in the value chain based on proportionate causal contribution, as determined under the factors in subsection (b).

**(d) Anti-Indemnification.** Any contractual provision requiring a person in the value chain to indemnify, defend, or hold harmless a frontier AI developer for liability arising under this Act, whether entered into before or after the harm, is void as against the public policy of this State and unenforceable in any court of this State. Compliance with any Federal, State, or local regulation, or with any voluntary standard or industry practice, does not by itself defeat liability under this Act.

## SECTION 7. FACTORS INFORMING DEFECTIVENESS.

**(a) General Rule.** Whether a frontier AI model is defective for purposes of the product liability law of this State shall be determined under the existing doctrines of that law (including manufacturing defect, design defect, and failure to warn), supplemented by the factors set forth in this section. The factors in this section are non-exclusive and shall be applied in light of the state of the art at the time of the conduct in question, except that adherence to the state of the art shall not by itself preclude a finding of defectiveness where the model fell short of the practices described in subsections (b) through (d).

**(b) Control Mechanisms.** In determining whether a frontier AI model is defective, a court shall consider whether the model, as designed, released, and maintained, provided a genuine capacity for a natural person to understand, direct, and countermand its operations such that a person could have prevented the harm. Relevant considerations include—

(1) whether the model supported control mechanisms appropriate to its level of autonomy, including—

(A) per-action or per-output approval for models operating interactively with human direction;

(B) real-time monitoring and intervention capability for models operating with human supervision; or

(C) design-phase specification of operational parameters, domain boundaries, and safety constraints, combined with ongoing monitoring, update capability, and the ability to modify or halt model operation, for models operating autonomously within bounded domains;

(2) whether operators had access to information about model behavior sufficient to identify risks of harm;

(3) whether operators possessed the technical capacity to intervene at the appropriate layer, in time relative to the speed and reversibility of model action;

(4) whether institutional and operational structures supported genuine override authority; and

(5) whether control mechanisms that existed as technical features were practicable to exercise under realistic operational conditions.

A control mechanism that exists as a technical feature but is impractical to exercise under realistic operational conditions does not, by itself, preclude a finding that the model was defective.

**(c) Safety and Security Protocols.** A court shall consider whether the frontier AI developer and other persons in the value chain implemented and maintained reasonable safety and security protocols commensurate with the model's capabilities and foreseeable risks, including—

(1) **Risk management.** Documented risk identification, assessment, and mitigation processes covering reasonably foreseeable misuse, emergent capabilities, and downstream fine-tuning or modification;

(2) **Information security.** Technical and organizational measures sufficient to prevent unauthorized access to, exfiltration of, or release of model weights, training data, and safety-critical infrastructure;

(3) **Access controls.** Authentication, authorization, rate-limiting, and usage-monitoring measures appropriate to the model's risk profile, including for application-programming-interface and partner-access deployments;

(4) **Misuse safeguards.** Guardrails, refusal behaviors, and output filters reasonably robust against foreseeable circumvention, jailbreaking, prompt injection, and removal of safety training through fine-tuning;

(5) **Monitoring and incident response.** Post-deployment monitoring of model behavior, channels for receiving reports of harm or near-harm, and processes for timely investigation, mitigation, model update, or withdrawal; and

(6) **Documentation and disclosure.** Documentation of known limitations, residual risks, and intended operational envelope, in a form usable by downstream distributors and deployers.

**(d) Pre-Deployment Testing.** A court shall consider whether the frontier AI developer conducted pre-deployment testing reasonable in scope, rigor, and independence given the model's capabilities and foreseeable uses, including—

(1) **Capability evaluations** designed to identify the model's capacity to cause or materially contribute to death, substantial physical injury or illness, severe psychological harm, large-scale property damage, or significant disruption of critical infrastructure;

(2) **Adversarial testing and red-teaming** conducted by personnel with sufficient independence, expertise, time, and access to the model to surface foreseeable failure modes, including misuse, jailbreaking, and removal of safeguards through fine-tuning where weights are or may be released;

(3) **Domain-specific safety testing** in any high-risk domain in which the model is foreseeably to be used, including dual-use scientific, cyber, and autonomous-action contexts;

(4) **Evaluation of control mechanisms** described in subsection (b) under realistic operational conditions, rather than only in laboratory or sandbox settings;

(5) **Documentation and retention** of testing methodology, results, and any decisions to release notwithstanding identified risks; and

(6) **Re-testing** following material capability enhancement, fine-tuning, or change in deployment configuration.

**(e) Failure to Warn.** A frontier AI model may be defective for failure to warn where the frontier

AI developer or other person in the value chain knew or should have known of a risk of harm and failed to provide warnings, instructions, or operational guidance reasonably adequate to enable downstream distributors, deployers, and end users to avoid that harm.

**(f) Preservation of Records; Adverse Inference.** A frontier AI developer shall preserve, for a period of not less than six (6) years following the deployment or release of a frontier AI model, contemporaneous records sufficient to establish the matters described in subsections (b) through (d), including telemetry and logs of model behavior, records of override or shutdown authority, pre-deployment and post-deployment evaluation and red-teaming results, and incident reports. Where a frontier AI developer fails to produce such records for the period relevant to the alleged harm, the court may draw an adverse inference against the developer on any factor under this section to which the missing records are relevant.

## SECTION 8. NEXUS.

This Act applies to harm caused by a frontier AI model where one or more of the following is the case—

(1) the harm was suffered by one or more residents of this State, or occurred in whole or in part within this State;

(2) the model was deployed to, made commercially available in, or accessed from within this State; or

(3) the frontier AI developer, distributor, or deployer regularly conducts business in this State, including any frontier AI developer that has, in the preceding twelve (12) months, received gross annual revenue of more than twenty-five million dollars ($25,000,000) from the sale or licensing of artificial intelligence products or services to persons in this State.

## SECTION 9. PRESERVATION OF OTHER LAW.

(a) This Act supplements and does not limit any claim or remedy otherwise available under the law of this State, including claims for negligence, other product liability theories, fraud, breach of warranty, or any other common-law or statutory cause of action.

(b) This Act does not preempt or limit the authority of any State agency to regulate artificial intelligence models within its existing jurisdiction.

(c) Any provision in a contract, terms of service, or user agreement that purports to waive, limit, or predetermine the forum or governing law for claims arising under this Act is void as against the public policy of this State and unenforceable in any court of this State.

## SECTION 10. EFFECTIVE DATE AND SEVERABILITY.

(a) This Act shall take effect 90 days after enactment and shall apply to harm occurring on or after the effective date, regardless of when the frontier AI model at issue was developed or released.

(b) If any provision of this Act, or the application thereof to any person or circumstances, is held invalid, the remainder of the Act and the application of the provision to other persons or circumstances shall not be affected.