

AI Responsibility for Catastrophic Harm Act

("ARCH Act")



View this document online: futureoflife.org/absolute-liability

A MODEL BILL

To impose strict liability on frontier AI developers for catastrophic harms caused by frontier artificial intelligence models, as an ultra-hazardous activity.

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

SECTION 1. SHORT TITLE.

This Act may be cited as the "AI Responsibility for Catastrophic Harm Act."

SECTION 2. FINDINGS AND PURPOSE.

(a) Findings. The Legislature finds—

1. The development and release of frontier AI models, when it goes wrong at scale, can cause death, serious injury, or catastrophic disruption on a magnitude no individual victim, deployer, or downstream provider can absorb or insure against.
2. Such activity exhibits the characteristics the common law associates with ultra-hazardous or abnormally dangerous activities: a high degree of risk of grave harm, an inability to eliminate that risk through reasonable care, and concentration of the activity in a small number of commercially benefiting actors.
3. The settled rule is that one who carries on such an activity is liable for the harm it causes without regard to fault, defect, or compliance with industry practice.
4. AI models cannot bear legal responsibility; every catastrophic harm caused by an AI model is the consequence of decisions made by the persons who designed, trained, and released it.

5. Frontier AI developers routinely use business-to-business contracts to shield themselves from liability and pass it to downstream providers and deployers with limited ability to test or safeguard the underlying models.
6. The magnitude of these harms falls within this State's traditional police powers over public health, safety, and economic welfare, pursuant to which this Act is enacted.

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

1. to impose strict liability on frontier AI developers for catastrophic harms caused by frontier AI models;
2. to ensure persons suffering such harm have a remedy without proving negligence, defect, or absence of human control; and
3. to align the cost of catastrophic AI harms with the parties whose activity creates the risk.

SECTION 3. DEFINITIONS.

In this Act:

(1) ARTIFICIAL INTELLIGENCE model. — An engineered or machine-based model that, for explicit or implicit objectives, generates outputs (predictions, recommendations, decisions, or actions) that influence physical or virtual environments. Excludes conventional software operating solely through predetermined rules without learning, adapting, or generating outputs beyond its explicit programming.

(2) CATASTROPHIC HARM. — Harm caused by an AI model that results in—

- (A) the death of one hundred (100) or more persons, whether arising from a single event or from the cumulative operation of one or more AI models developed or controlled by a single frontier AI developer over any twelve (12) month period;
- (B) substantial physical injury or illness to five hundred (500) or more persons, on the same single-event or twelve-month basis;
- (C) significant disruption of critical infrastructure, as defined in section 1016(e) of the USA PATRIOT Act (42 U.S.C. § 5195c(e)), or of any model or asset whose disruption would have a

debilitating impact on the security, economic stability, public health, or safety of this State; or

(D) damage to property exceeding one billion dollars (\$1,000,000,000) in the aggregate over any twelve (12) month period;

(3) CHAIN OF DEVELOPMENT. — The sequence of frontier AI developers whose acts or omissions in designing, training, fine-tuning, modifying, or releasing the model or any predecessor model contributed to the capabilities or deployment configuration that resulted in the harm. Includes—

(A) the developer of any foundation or base model from which the model was derived through fine-tuning, distillation, retrieval-augmentation, or other post-training modification;

(B) any developer that conducted material fine-tuning, reinforcement-learning training, or other capability enhancement; and

(C) any developer that released model weights or other technical artifacts enabling the deployment that caused the harm.

The chain does not include a person whose contributions were so temporally or functionally remote that no reasonable factfinder could conclude they were a proximate cause.

(5) FRONTIER AI DEVELOPER. — A person that—

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

(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 AI 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.

(6) FRONTIER AI MODEL. — An AI model that—

(A) is capable of performing tasks across at least two (2) domains of human cognitive activity, including language, reasoning, analysis, or planning, whether through initial design or

subsequent fine-tuning; and

(B) possesses capabilities sufficient that it would be capable of causing harm of a character and magnitude described in paragraph (2).

A model designed and deployed exclusively for a single, well-defined task within a constrained operational environment is not a high-capability AI model.

SECTION 4. AI MODELS CANNOT BEAR RESPONSIBILITY.

(a) An AI model is not a legal person and shall not be treated as such for any purpose under the law of this State. It cannot bear legal responsibility, hold rights or immunities, own property, enter contracts, or be named as a party in any legal proceeding.

(b) In any action under this Act, it shall not be a defense that the AI model acted autonomously, acted unexpectedly or contrary to its design, or possessed characteristics of intelligence, understanding, or intent.

SECTION 5. STRICT LIABILITY FOR CATASTROPHIC HARM.

(a) **Activity Designated Ultra-Hazardous.** The development, training, and release of a frontier AI model by a frontier AI developer is hereby declared an abnormally dangerous activity within the meaning of the common law of this State.

(b) **Strict Liability Standard.** A frontier AI developer is strictly liable for all catastrophic harm caused by a frontier AI model, and liability attaches without regard to fault, negligence, recklessness, or intent.

(c) **Joint and Several Liability; Apportionment.** Strict liability under this section applies jointly and severally to all frontier AI developers in the chain of development. Any developer held liable may seek contribution from other responsible developers based on proportionate causal contribution, considering—

1. the degree to which each developer's capability contributions were a proximate cause of the harm;

2. each developer's causal contribution to the deployment configuration in which the harm occurred; and
3. the temporal and functional proximity of each developer's actions to the harm.

(d) Open-Weight Releases. A frontier AI developer who releases model weights or other technical artifacts enabling deployment of a high-capability AI model is strictly liable under this section for all reasonably foreseeable catastrophic harm caused by any deployment of, or model derived from, the released artifact, occurring within five (5) years of the date of release.

(e) Nexus. This section applies where—

1. the catastrophic harm was suffered by one or more residents of this State, or occurred in whole or in part within this State, or
2. the AI model was deployed to, made commercially available in, or accessed from within this State.

SECTION 6. EXEMPTION OF UNAFFILIATED DOWNSTREAM PARTIES.

(a) Exemption. No person other than a frontier AI developer in the chain of development of the frontier AI model that caused the catastrophic harm shall be subject to liability under this Act. Without limitation, the following persons are not subject to liability under this Act, provided they satisfy the independence conditions of subsection (b)—

1. Deployers that use, operate, host, or make available a high-capability AI model;
2. Integrators and downstream developers that incorporate, embed, wrap, orchestrate, or otherwise integrate a high-capability AI model or its outputs into a product, service, application, agent, or workflow, where the person is not itself a frontier AI developer with respect to that model;
3. Distributors and resellers that distribute, license, resell, or provide access to a high-capability AI model on substantially the terms on which the model was made available to them;
4. Infrastructure and platform providers that supply compute, cloud hosting, networking, storage, or similar generally-available infrastructure, where the provider does not direct

or materially shape the design, training, fine-tuning, evaluation, or release of the model;

5. End users that interact with a high-capability AI model in the ordinary course of using a product or service made available by another person; and
6. Tooling and component providers that supply datasets, evaluation tools, software libraries, model components, or other inputs that are not, standing alone, a high-capability AI model, and that do not direct or materially shape the development of the model that caused the harm.
7. Accredited institutions of higher education (as defined in 20 U.S.C. § 1001).

(b) Independence Conditions. The exemption applies only where the person, with respect to each frontier AI developer in the chain of development of the AI model that caused the harm is not the same person, and is not a parent, subsidiary, or affiliate of such a developer under common control through day-to-day managerial authority over the development, training, evaluation, or release of the frontier AI models;

(c) Preservation of Other Liability. Nothing in this section limits the liability of any exempted person under any other provision of law.

SECTION 7. ENFORCEMENT.

(a) Private Right of Action. Any person suffering catastrophic harm from a violation of §5 may bring a civil action in any court of competent jurisdiction, individually or as a member of a class.

(b) Attorney General Enforcement.

(1) Action in Name of State. The Attorney General may bring a civil action for injunctive relief, civil penalties, and other appropriate relief. Civil penalties shall not exceed five million dollars (\$5,000,000) per violation, except that no per-violation cap shall apply where the violation involves catastrophic harm described in §3(2)(C) (critical infrastructure).

(c) Non-Waiver and Anti-Indemnification.

(1) Non-Waiver. Remedies under this Act supersede any contrary provision in a contract, terms of service, or user agreement, including provisions requiring arbitration, limiting liability, or predetermining forum or governing law for claims under this Act.

(2) Anti-Indemnification. Any contractual provision requiring a person other than a frontier AI developer to indemnify, defend, or hold harmless a frontier AI developer for liability under this Act, whether entered before or after the harm, is void as against the public policy of this State and unenforceable in any court of this State.

(d) Statute of Limitations. Actions under this Act must be brought within six (6) years of the date the injury was discovered or reasonably should have been discovered.

SECTION 8. PRESERVATION OF OTHER LAW.

(a) This Act establishes an additional basis for liability and does not limit or preclude any claim or remedy otherwise available under the law of this State, including claims for negligence, product liability, 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 AI models within its existing jurisdiction.

SECTION 9. EFFECTIVE DATE AND SEVERABILITY.

(a) This Act shall take effect [90 days] after enactment.

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