



104TH GENERAL ASSEMBLY

State of Illinois

2025 and 2026

SB3261

Introduced 2/3/2026, by Sen. Mary Edly-Allen

SYNOPSIS AS INTRODUCED:

New Act

Creates the Artificial Intelligence Public Safety and Child Protection Transparency Act. Provides that a frontier artificial intelligence model developer or large chatbot provider shall write, implement, comply with, and clearly and conspicuously publish on its website a public safety and child protection plan. Provides that the Attorney General shall establish a mechanism to be used by a large frontier developer, a large chatbot provider, or a member of the public to report a safety incident related to specified artificial intelligence models or chatbots. Sets forth provisions concerning the protection of whistleblowers; third party audits of large frontier developers; and civil penalties. Provides for rulemaking by the Attorney General. Effective January 1, 2027.

LRB104 19179 SPS 32624 b

1 AN ACT concerning business.

2 **Be it enacted by the People of the State of Illinois,**
3 **represented in the General Assembly:**

4 Section 1. Short title. This Act may be cited as the
5 Artificial Intelligence Public Safety and Child Protection
6 Transparency Act.

7 Section 5. Findings. The General Assembly finds and
8 declares:

9 (a) Artificial intelligence, including new advances in
10 foundation models, has the potential to catalyze innovation
11 and the rapid development of a wide range of benefits for
12 Illinoisans and the Illinois economy, including advances in
13 medicine, agriculture, and climate science, and to push the
14 bounds of human creativity and capacity.

15 (b) Targeted interventions to support effective artificial
16 intelligence governance should balance the technology's
17 benefits and the potential for material risks.

18 (c) In building a robust and transparent evidence
19 environment, policymakers can align incentives to
20 simultaneously protect consumers, leverage industry expertise,
21 and recognize leading safety practices.

22 (d) As industry actors conduct internal research on their
23 technologies' impacts, public trust in these technologies

1 would significantly benefit from access to information
2 regarding, and increased awareness of, frontier artificial
3 intelligence capabilities.

4 (e) Greater transparency can also advance accountability,
5 competition, and public trust.

6 (f) Whistleblower protections and public-facing
7 information sharing are key instruments to increase
8 transparency.

9 (g) Incident reporting systems enable monitoring of the
10 post-deployment impacts of artificial intelligence.

11 (h) Unless they are developed with careful diligence and
12 reasonable precaution, there is concern that advanced
13 artificial intelligence systems could have capabilities that
14 pose catastrophic risks from both malicious uses and
15 malfunctions, including artificial intelligence-enabled
16 hacking, biological attacks, and loss of control.

17 (i) With the frontier of artificial intelligence rapidly
18 evolving, there is a need for legislation to track the
19 frontier of artificial intelligence research and alert
20 policymakers and the public to serious risks and harms from
21 the most advanced artificial intelligence systems, while
22 avoiding burdening smaller companies behind the frontier.

23 (j) While the major artificial intelligence developers
24 have already voluntarily established the creation, use, and
25 publication of frontier artificial intelligence frameworks as
26 an industry best practice, not all developers have provided

1 reporting that is consistent and sufficient to ensure
2 necessary transparency and protection of the public.
3 Mandatory, standardized, and objective reporting by frontier
4 developers is required to provide the government and the
5 public with timely and accurate information.

6 (k) Timely reporting of critical safety incidents to the
7 government is essential to ensure that public authorities are
8 promptly informed of ongoing and emerging risks to public
9 safety. This reporting enables the government to monitor,
10 assess, and respond effectively if the advanced capabilities
11 emerge in frontier artificial intelligence models that may
12 pose a threat to the public.

13 (l) In the future, foundation models developed by smaller
14 companies or that are behind the frontier may pose significant
15 catastrophic risk, and additional legislation may be needed at
16 that time.

17 (m) It is the intent of the General Assembly to create more
18 transparency, but collective safety will depend in part on
19 frontier developers taking due care in their development and
20 deployment of frontier models proportional to the scale of the
21 foreseeable risks.

22 Section 10. Definitions. As used in this Act:

23 "Affiliate" means a person controlling, controlled by, or
24 under common control with a specified person, directly or
25 indirectly, through one or more intermediaries.

1 "Artificial intelligence model" means an engineered or
2 machine-based system that varies in its level of autonomy and
3 that can, for explicit or implicit objectives, infer from the
4 input it receives how to generate outputs that can influence
5 physical or virtual environments.

6 "Catastrophic risk" means a foreseeable and material risk
7 that a frontier developer's development, storage, use, or
8 deployment of a frontier model will materially contribute to
9 the death of, or serious injury to, more than 50 people or more
10 than \$1,000,000,000 in damage to, or loss of, property arising
11 from a single incident involving a frontier model doing the
12 following:

13 (1) providing expert-level assistance in the creation
14 or release of a chemical, biological, radiological, or
15 nuclear weapon;

16 (2) engaging in conduct with no meaningful human
17 oversight, intervention, or supervision that is either a
18 cyberattack or, if the conduct had been committed by a
19 human, would constitute the crime of murder, assault,
20 extortion, or theft, including theft by false pretense; or

21 (3) evading the control of its frontier developer or
22 user.

23 "Catastrophic risk" does not include a foreseeable and
24 material risk from the following:

25 (1) information that a frontier model outputs if the
26 information is otherwise publicly accessible in a

1 substantially similar form from a source other than a
2 foundation model;

3 (2) lawful activity of the federal government; or

4 (3) harm caused by a frontier model in combination
5 with other software if the frontier model did not
6 materially contribute to the harm.

7 "Child safety incident" means death or bodily injury to a
8 minor resulting from the materialization of a child safety
9 risk.

10 "Child safety risk" means a material and foreseeable risk
11 that a frontier developer's foundation model, when used as
12 part of a covered chatbot operated by the frontier developer,
13 will engage in behavior when conversing with a minor that, if
14 it had been engaged in by a human, would be deemed to
15 intentionally or recklessly do the following:

16 (1) cause death or bodily injury to that minor,
17 including as a result of self-harm; or

18 (2) cause damage to mental health that constitutes
19 severe emotional distress.

20 "Covered chatbot" means a service that:

21 (1) allows an ordinary person to converse with and
22 have conversations where humanlike responses are generated
23 by a foundation model;

24 (2) is foreseeably likely to be accessed by minors;
25 and

26 (3) has at least 1,000,000 monthly active users.

1 "Covered risk" means a catastrophic risk or a child safety
2 risk.

3 "Critical safety incident" means the following:

4 (1) unauthorized access to, modification of,
5 inadvertent release of, or exfiltration of, the model
6 weights of a frontier model;

7 (2) the death of, or serious injury to, more than 50
8 people or more than \$1,000,000,000 in damage to, or loss
9 of, property resulting from the materialization of a
10 catastrophic risk;

11 (3) loss of control of a frontier model that causes
12 death, bodily injury, or that demonstrates materially
13 increased catastrophic risk; or

14 (4) the use of deceptive techniques by a frontier
15 model against its frontier developer to subvert the
16 controls or monitoring of its frontier developer outside
17 of the context of an evaluation designed to elicit this
18 behavior and in a manner that demonstrates materially
19 increased catastrophic risk.

20 "Deploy" means to make a frontier model available to a
21 third party for use, modification, copying, or combination
22 with other software. "Deploy" does not include making a
23 frontier model available to a third party for the primary
24 purpose of developing or evaluating the frontier model.

25 "Employee" has the meaning set forth in the Whistleblower
26 Act.

1 "Foundation model" means an artificial intelligence model
2 that is:

- 3 (1) trained on a broad data set;
- 4 (2) designed for generality of output; and
- 5 (3) adaptable to a wide range of distinctive tasks.

6 "Frontier developer" means a person who has trained, or
7 initiated the training of, a frontier model, with respect to
8 which the person has used, or intends to use, at least as much
9 computing power to train the frontier model as would meet the
10 technical specifications described in the definition of
11 "frontier model". "Frontier developer" does not include
12 accredited colleges and universities to the extent that the
13 colleges and universities are engaging in academic research.
14 For the purpose of this definition, if a person subsequently
15 transfers full intellectual property rights of a frontier
16 model to another person, including the right to resell the
17 model, and retains none of those rights, then the receiving
18 person shall be considered the frontier developer with respect
19 to that frontier model.

20 "Frontier model" means a foundation model that was trained
21 using a quantity of computing power greater than 10^{26} integer
22 or floating-point operations. The quantity of computing power
23 described in this definition shall include computing for the
24 original training run and for any subsequent fine-tuning,
25 reinforcement learning, or other material modifications the
26 developer applies to a preceding foundation model.

1 "Large chatbot provider" means a provider who makes a
2 covered chatbot available in this State and who, together with
3 the provider's affiliates, collectively have an annual revenue
4 of at least \$25,000,000.

5 "Large frontier developer" means a frontier developer who,
6 together with the large frontier developer's affiliates,
7 collectively have an annual revenue of at least \$500,000,000.

8 "Minor" means an individual younger than 18 years old.

9 "Model weight" means a numerical parameter in a frontier
10 model that is adjusted through training and that helps
11 determine how inputs are transformed into outputs.

12 "Property" means tangible or intangible property.

13 "Public safety and child protection plan" means a
14 documented technical and organizational protocol to manage,
15 assess, and mitigate covered risks.

16 "Safety incident" means a child safety incident or a
17 critical safety incident.

18 Section 15. Public safety and child protection plans.

19 (a) A large frontier developer or large chatbot provider
20 shall write, implement, comply with, and clearly and
21 conspicuously publish on its website a public safety and child
22 protection plan that describes in detail:

23 (1) For a large frontier developer only, how the large
24 frontier developer:

25 (A) defines and assesses thresholds used by the

1 large frontier developer to identify and assess
2 whether a frontier model has capabilities that could
3 pose a catastrophic risk, which may include
4 multiple-tiered thresholds;

5 (B) applies mitigations to address the potential
6 for catastrophic risks based on the results of the
7 assessments undertaken in accordance with subparagraph
8 (A);

9 (C) reviews assessments of catastrophic risk and
10 adequacy of mitigations of catastrophic risk as part
11 of the decision to deploy a frontier model or use it
12 extensively internally;

13 (D) uses third parties to assess the potential for
14 catastrophic risks and the effectiveness of
15 mitigations of catastrophic risks;

16 (E) implements cybersecurity practices to secure
17 unreleased frontier model weights from unauthorized
18 modification or transfer by internal or external
19 parties; and

20 (F) assesses and manages catastrophic risk
21 resulting from the internal use of its frontier
22 models, including risks resulting from a frontier
23 model circumventing oversight mechanisms or being used
24 for artificial intelligence research and development
25 in a manner that could materially increase
26 catastrophic risk.

1 (2) For a large chatbot provider only, how the large
2 chatbot provider:

3 (A) assesses potential for child safety risks;

4 (B) applies mitigations to address the potential
5 for child safety risks based on the results of the
6 assessments undertaken in accordance with subparagraph
7 (A); and

8 (C) uses third parties to assess the potential for
9 child safety risks and the effectiveness of
10 mitigations of child safety risks.

11 (3) For both large frontier developers and large
12 chatbot providers, how the large frontier developer or
13 large chatbot provider:

14 (A) incorporates national standards, international
15 standards, and industry-consensus best practices into
16 its public safety and child protection plan;

17 (B) revisits and updates the public safety and
18 child protection plan, including any criteria that
19 trigger updates and how the large frontier developer
20 determines when its foundation models or frontier
21 models are substantially modified enough to require
22 disclosures in accordance with subsection (d) or
23 subsection (e);

24 (C) identifies and responds to safety incidents;
25 and

26 (D) institutes internal governance practices to

1 ensure implementation of these processes.

2 (b) A large frontier developer shall write its public
3 safety and child protection plan so that, if successfully
4 implemented, it would prevent unreasonable catastrophic risk.

5 (c) If a large frontier developer or large chatbot
6 provider makes a material modification to its public safety
7 and child protection plan, the large frontier developer or
8 large chatbot provider shall clearly and conspicuously publish
9 the modified public safety and child protection plan and a
10 justification for that modification within 30 days after the
11 modification is made.

12 (d) Before, or concurrently with, integrating a new
13 foundation model, or a version of an existing foundation model
14 that has been substantially modified, into a covered chatbot
15 operated by a large chatbot provider, a large chatbot provider
16 shall conspicuously publish on its website summaries of the
17 following:

18 (1) all assessments of child safety risks conducted in
19 accordance with the large chatbot provider's public safety
20 and child protection plan;

21 (2) the results of those assessments;

22 (3) the extent to which third-party evaluators were
23 involved; and

24 (4) other steps taken to fulfill the requirements of
25 the public safety and child protection plan with respect
26 to child safety risks.

1 (e) Before, or concurrently with, deploying a new frontier
2 model or a version of an existing frontier model that a large
3 frontier developer has substantially modified, a large
4 frontier developer shall implement appropriate safeguards to
5 prevent unreasonable catastrophic risk and conspicuously
6 publish on its website summaries of the following:

7 (1) all assessments of catastrophic risks from the
8 frontier model conducted in accordance with the large
9 frontier developer's public safety and child protection
10 plan;

11 (2) the results of those assessments;

12 (3) the extent to which third-party evaluators were
13 involved; and

14 (4) other steps taken to fulfill the requirements of
15 the public safety and child protection plan with respect
16 to catastrophic risks from the frontier model.

17 A large frontier developer that publishes the information
18 described in this subsection as part of a larger document,
19 including a system card or model card, shall be deemed in
20 compliance with this subsection.

21 (f) A large frontier developer shall not use or deploy a
22 frontier model if doing so would pose unreasonable
23 catastrophic risk.

24 (g) A large frontier developer or large chatbot provider
25 shall not make a materially false or misleading statement or
26 omission about covered risks from its activities or its

1 management of covered risks.

2 A large frontier developer or large chatbot provider shall
3 not make a materially false or misleading statement or
4 omission about its implementation of, or compliance with, its
5 public safety and child protection plan.

6 This subsection does not apply to a statement that was
7 made in good faith and was reasonable under the circumstances.

8 (h) When a large frontier developer or large chatbot
9 provider publishes documents to comply with this Section, the
10 large frontier developer or large chatbot provider may make
11 redactions to those documents that are necessary to protect
12 the large frontier developer's or large chatbot provider's
13 trade secrets, the large frontier developer's or large chatbot
14 provider's cybersecurity, public safety, or the national
15 security of the United States or to comply with any State or
16 federal law. If a large frontier developer or large chatbot
17 provider redacts information in a document under this
18 subsection, the large frontier developer or large chatbot
19 provider shall describe the character and justification of the
20 redaction in any published version of the document to the
21 extent permitted by the concerns that justify redaction and
22 shall retain the unredacted information for 5 years.

23 Section 20. Reporting of safety incidents.

24 (a) The Attorney General shall establish a mechanism to
25 be used by a frontier developer, a large chatbot provider, or a

1 member of the public to report a safety incident that includes
2 the following:

3 (1) the date of the safety incident;

4 (2) the reasons the incident qualifies as a safety
5 incident; and

6 (3) a short and plain statement describing the safety
7 incident.

8 (b) A frontier developer shall report any critical safety
9 incident pertaining to one of its frontier models to the
10 Attorney General within 15 days after discovering the critical
11 safety incident.

12 (c) If a frontier developer discovers that a critical
13 safety incident poses an imminent risk of death or serious
14 physical injury, the frontier developer shall disclose that
15 incident within 24 hours after discovering the critical safety
16 incident to an authority, including any law enforcement agency
17 or public safety agency with jurisdiction, that is appropriate
18 based on the nature of that incident and as required by law.

19 (d) A large chatbot provider shall report any child safety
20 incident pertaining to one of its covered chatbots to the
21 Attorney General within 15 days after discovering the child
22 safety incident.

23 (e) The Attorney General shall establish a mechanism to be
24 used by a large frontier developer to confidentially submit
25 summaries of any assessments of the potential for catastrophic
26 risk resulting from internal use of its frontier models.

1 (f) A large frontier developer shall transmit to the
2 Attorney General a summary of any assessment of catastrophic
3 risk resulting from internal use of its frontier models no
4 less frequently than every 3 months.

5 (g) The Attorney General may transmit reports of safety
6 incidents, summaries of assessments of the potential for
7 catastrophic risk from internal use, and reports from
8 employees to the General Assembly, the Governor, the federal
9 government, or an appropriate State agency. The Attorney
10 General shall consider any risks related to trade secrets,
11 public safety, cybersecurity, or national security when
12 transmitting reports.

13 Section 25. Rulemaking; definitions.

14 (a) On or before January 1, 2028, and annually thereafter,
15 the Attorney General shall assess recent evidence and
16 developments relevant to the purposes of this Act and may
17 adopt rules to update the following definitions for the
18 purposes of this Act to ensure that they accurately reflect
19 technological developments, scientific literature, and widely
20 accepted national and international standards:

21 (1) "Frontier model" so that it applies to foundation
22 models at the frontier of artificial intelligence
23 development.

24 (2) "Frontier developer" so that it applies to
25 developers of frontier models who are themselves at the

1 frontier of artificial intelligence development.

2 (3) "Large frontier developer" so that it applies to
3 well-resourced frontier developers.

4 (4) "Large chatbot provider" so that it applies to
5 well-resourced companies developing covered chatbots that
6 may pose child safety risks.

7 (b) In adopting rules under this Section, the Attorney
8 General shall take into account the following:

9 (1) similar thresholds used in international standards
10 or federal law, regulations, or guidance documents for the
11 management of catastrophic risks or child safety risks;

12 (2) input from stakeholders, including academics,
13 industry, the open-source community, and governmental
14 entities;

15 (3) the extent to which a person will be able to
16 determine, before beginning to train or deploy a
17 foundation model, whether that person will be subject to
18 this Act as a frontier developer or as a large frontier
19 developer with a focus toward allowing earlier
20 determinations if possible;

21 (4) the complexity of determining whether a person or
22 foundation model is covered, with a focus toward allowing
23 simpler determinations if possible;

24 (5) the external verifiability of determining whether
25 a person or foundation model is covered, with a focus
26 toward definitions that are verifiable by parties other

1 than the frontier developer; and

2 (6) thresholds used by other states in similar law.

3 (c) The Attorney General shall align any rules adopted
4 under this Section with a definition adopted in a federal law
5 or regulation, to the extent that it is consistent with the
6 purposes of this Act.

7 Section 30. Protection of whistleblowers.

8 (a) A frontier developer or large chatbot provider shall
9 not make, adopt, enforce, or enter into a rule, regulation,
10 policy, or contract that prevents an employee from making a
11 disclosure protected under the Whistleblower Act.

12 (b) A large frontier developer shall provide a reasonable
13 internal process through which an employee may anonymously
14 disclose information to the large frontier developer if the
15 employee has a good faith belief that the information
16 discloses a substantial and specific danger to employees,
17 public health, or safety or a violation of this Act, including
18 a monthly update to the person who made the disclosure
19 regarding the status of the large frontier developer's
20 investigation of the disclosure and the actions taken by the
21 large frontier developer in response to the disclosure.

22 (c) Except as provided in subsection (d), the disclosures
23 and responses of the process required by this Section shall be
24 shared with officers and directors of the large frontier
25 developer at least once each quarter.

1 (d) If an employee has alleged wrongdoing by an officer or
2 director of the large frontier developer in a disclosure or
3 response, subsection (c) shall not apply with respect to that
4 officer or director.

5 Section 35. Third-party audits.

6 (a) At least once every calendar year, a large frontier
7 developer shall retain a reputable third-party auditor to
8 produce a report assessing the following:

9 (1) whether the large frontier developer has complied
10 with its public safety plan and any instances of
11 noncompliance;

12 (2) any instances where the large frontier developer's
13 public safety plan has not been stated clearly enough to
14 determine whether the large frontier developer has
15 complied; and

16 (3) whether redactions made by the large frontier
17 developer in documents published in accordance with this
18 Act are reasonable and whether any statements made by the
19 large frontier developer may be false or misleading.

20 (b) A large frontier developer shall allow the third-party
21 auditor access to all materials produced to comply with this
22 Act and any other materials reasonably necessary to perform
23 the assessment required under subsection (a).

24 (c) The large frontier developer shall retain the
25 auditor's report for 5 years and allow the Attorney General to

1 inspect the unredacted version of the report upon request.

2 (d) In conducting the audit, the auditor shall employ or
3 contract one or more individuals with expertise in corporate
4 compliance and one or more individuals with technical
5 expertise in the safety of foundation models.

6 Section 40. Civil penalties.

7 (a) A large frontier developer that violates this Act
8 shall be subject to a civil penalty in an amount dependent upon
9 the severity of the violation that does not exceed \$1,000,000
10 per violation.

11 (b) A large chatbot provider that violates this Act shall
12 be subject to a civil penalty in an amount dependent upon the
13 severity of the violation that does not exceed \$50,000 per
14 violation.

15 (c) A civil penalty described in this Section shall be
16 recovered in a civil action brought by the Attorney General.

17 Section 45. Loss of equity. The loss of value of equity
18 does not count as damage to or loss of property for the
19 purposes of this Act.

20 Section 50. Compliance with other laws.

21 (a) The Attorney General may adopt rules creating
22 alternative compliance pathways for frontier developers or
23 large chatbot providers that comply with a federal law,

1 regulation, or guidance document or a law of another state of
2 the United States.

3 (b) A rule adopted under this Section shall:

4 (1) Specify the provisions of this Act for which the
5 alternative compliance pathway is being established.

6 (2) Specify the federal law, regulation, or guidance
7 document, or the law of another state, compliance with
8 which shall serve as the alternative compliance pathway
9 for the provisions specified under paragraph (1). The
10 federal law, regulation or guidance document or the law of
11 another state shall be substantially equivalent to, or
12 more protective against catastrophic risk than, the
13 provisions of this Act specified under paragraph (1).

14 (c) If a rule adopted under this Section identifies, as
15 described in paragraph (1) of subsection (b), a provision of
16 this Act that requires reporting to the State and if the
17 alternative compliance pathway requires reporting to the
18 federal government, the rule may, but need not, continue to
19 require reporting to the State. The rule shall not consider
20 reporting to another state to be sufficient for compliance
21 with the relevant provision of this Act.

22 (d) A rule adopted under this Section may establish steps
23 frontier developers or large chatbot providers must take to
24 demonstrate their compliance with the alternative compliance
25 pathway if it would otherwise be challenging for the State to
26 verify compliance, such as the submission of documentation to

1 the State.

2 (e) A frontier developer or large chatbot provider that
3 intends to make use of an alternative compliance pathway
4 created by rule under this Section shall declare its intent to
5 do so to the Attorney General.

6 After declaring its intent, a frontier developer or large
7 chatbot provider shall be deemed in compliance with the
8 provision of this Act identified by the rule under paragraph
9 (1) of subsection (b) to the extent that the frontier
10 developer or large chatbot provider complies with the
11 requirements of the rule and meets the standards of, or
12 complies with the requirements imposed or stated by, the
13 federal law, regulation, or guidance document or law of
14 another state identified by the rule under paragraph (2) of
15 subsection (b) until the frontier developer or large chatbot
16 provider declares the revocation of that intent to the
17 Attorney General or the Attorney General revokes the rule in
18 accordance with subsection (f).

19 (f) The Attorney General shall revoke a rule adopted under
20 this Section if the conditions specified by this Section no
21 longer apply.

22 Section 99. Effective date. This Act takes effect January
23 1, 2027.