HOUSE BILL NO. 4668

June 24, 2025, Introduced by Rep. Lightner and referred to Committee on Judiciary.

A bill to require large developers to implement safety and security protocols to manage critical risks of foundation models; to prescribe the duties of large developers; to provide protection for certain employees; to provide for the powers and duties of certain state and local governmental officers and entities; and to prescribe civil sanctions and provide remedies.

THE PEOPLE OF THE STATE OF MICHIGAN ENACT:

Sec. 1. This act may be cited as the "artificial intelligence safety and security transparency act".

Sec. 3. As used in this act:

(a) "Artificial intelligence model" means an engineered or machine-based system that varies in the system's level of autonomy and that can, for explicit or implicit objectives, infer from input received how to generate outputs that can influence physical or virtual environments.

(b) "Critical risk" means a foreseeable and material risk that a large 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 will result in more than $1,000,000,000.00 in damages to rights in money or property, through an incident of any of the following kinds:

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

(*ii*) A cyberattack conducted by or assisted by a foundation model.

(*iii*) A foundation model engaging in conduct that meets both of the following:

(A) Is performed with limited human intervention.

(B) Would, if the conduct was committed by an individual, constitute a crime that requires intent, recklessness, or gross negligence, or the solicitation or aiding and abetting of a crime.

(*iv*) A harm as a result of an incident described under subparagraphs (*i*) to (*iii*) that is inflicted by an intervening individual only if the large developer's activities made it substantially easier or more likely for the individual to inflict the harm.

(c) "Deploy" means to use a foundation model or to make a foundation model foreseeably available to 1 or more third parties

for use, modification, copying, or combination with other software, except as reasonably necessary for developing the foundation model or evaluating the foundation model or other foundation models.

(d) "Employee" means an individual who performs services for wages or salary under a contract of employment, express or implied, for an employer, including both of the following:

(*i*) A contractor or subcontractor and unpaid advisors involved with assessing, managing, or addressing a critical risk.

(*ii*) A corporate officer.

(e) "Foundation model" means an artificial intelligence model that meets all of the following requirements:

(*i*) Is trained on a broad data set.

(*ii*) Is designed for generality of output.

(*iii*) Is adaptable to a wide range of distinctive tasks.

(f) "Large developer" means a person that has developed both of the following:

(*i*) A foundation model with a quantity of computing power that costs not less than $5,000,000.00 when measured using prevailing market prices of cloud computing in the United States at the time that the computing power was used.

(*ii*) Within the immediately preceding 12 months, 1 or more foundation models with a total quantity of computing power that costs not less than $100,000,000.00 when measured using prevailing market prices of cloud computing in the United States at the time the computing power was used.

(g) "Safety and security protocol" means a set of documented technical and organizational protocols used by a large developer to manage critical risks that meet the requirements of section 5.

Sec. 5. A safety and security protocol must describe in detail

all of the following, as applicable:

(a) How the large developer excludes certain foundation models from being covered by the safety and security protocol when those foundation models pose a limited critical risk.

(b) The thresholds at which critical risks would be considered intolerable, any justification for the thresholds, and what the large developer will do if a threshold is surpassed.

(c) The testing and assessment procedures the large developer uses to investigate critical risks and how the tests and procedures account for the possibility that a foundation model could evade the control of the large developer or user or be misused, modified, executed with increased computational resources, or used to create another foundation model.

(d) The procedure the large developer will use to determine if and how to deploy a foundation model when doing so poses critical risks.

(e) The physical, digital, and organizational security protection the large developer will implement to prevent insiders or third parties from accessing foundation models within the large developer's control in a manner that is unauthorized by the developer and could create a critical risk.

(f) Any safeguards and risk mitigation measures the large developer uses to reduce critical risks from the large developer's foundation models and how the large developer assesses efficacy and limitations.

(g) How the large developer will respond if a critical risk materializes or is imminent.

(h) The procedures that the large developer uses to determine whether to conduct additional assessments for a critical risk when

the large developer modifies or expands access to the large developer's foundation models or combines the foundation models with other software and how such assessments are conducted.

(i) The conditions under which the large developer will report an incident relevant to a critical risk that occurs in connection with 1 or more of the large developer's foundation models and the entities to which the large developer will make those reports.

(j) The conditions under which the large developer will modify the large developer's safety and security protocol.

(k) The parts of the safety and security protocol that the large developer believes provide sufficient scientific detail to allow for the independent assessment of the methods used to generate the results, evidence, and analysis, and to which experts any unredacted versions are made available.

(*l*) Any other role a financially disinterested third party plays under subdivisions (a) to (k).

Sec. 7. (1) Beginning on January 1, 2026, a large developer shall do all of the following:

(a) Produce, implement, follow, and conspicuously publish a safety and security protocol.

(b) If materially modifying the safety and security protocol under subdivision (a), conspicuously publish the modifications not more than 30 days after the material modification was made.

(c) Not less than once every 90 days, produce and conspicuously publish a transparency report that covers the period of 120 days before the publishing of the report to 30 days before the publishing of the report that includes all of the following information:

(*i*) The conclusion of any risk assessments made during the

reporting period in accordance with the safety and security protocol under subdivision (a).

(*ii*) If different from the preceding reporting period, for each type of critical risk, an assessment of the relevant capability of the foundation model to create that critical risk of whichever of the large developer's foundation models, whether deployed or not, would pose the highest level of that critical risk if deployed without adequate safeguards and protections.

(*iii*) If, during the reporting period, the large developer has deployed or modified a foundation model that would pose a higher level of critical risk than any of the large developer's existing deployed foundation models if deployed without adequate safeguards and protections, both of the following:

(A) The grounds on which and the process by which the large developer decided to deploy the foundation model.

(B) Any safeguards and protections implemented by the large developer to mitigate critical risks.

(d) Record and retain for 5 years any specific tests used and results obtained as a part of an assessment of critical risk with sufficient detail for qualified third parties to replicate the testing.

(2) A large developer shall not knowingly make false or materially misleading statements or omissions in or regarding documents produced in accordance with this section.

(3) If a large developer publishes a document in accordance with the requirements of this act, the large developer shall publish the information on a conspicuous page on the large developer's website. The large developer may redact the document as reasonably necessary to protect the large developer's trade

secrets, public safety, or national security, or to comply with applicable law. An auditor required to perform an audit and produce a report under section 9 may redact information from the report using the same procedure described in this subsection before the publication of that report under section 9(3).

(4) If a large developer or auditor makes a redaction under subsection (3), the large developer or auditor shall do both of the following:

(a) Retain an unredacted version of the document for not less than 5 years and provide the attorney general with the ability to inspect the unredacted document on request.

(b) Describe the character and justification of the redactions in the published version of the document.

Sec. 9. (1) Beginning on January 1, 2026, not less than once per year, a large developer shall retain a reputable third-party auditor to produce a report that assesses all of the following:

(a) If the large developer has complied with the large developer's safety and security protocol and any instances of noncompliance.

(b) Any instance where the large developer's safety and security protocol was not stated clearly enough to determine if the large developer has complied with the safety and security protocol.

(c) Any instance that the auditor believes the large developer violated section 7(2), (3), or (4).

(2) A large developer shall grant the auditor access to all materials produced to comply with this act and any other materials reasonably necessary to perform the assessment under subsection (1).

(3) Not more than 90 days after the completion of the

auditor's report under subsection (1), a large developer shall conspicuously publish that report.

(4) In conducting an audit under this section, an auditor shall employ or contract 1 or more individuals with expertise in corporate compliance and 1 or more individuals with technical expertise in the safety of foundation models.

Sec. 11. (1) A large developer shall not discharge, threaten, or otherwise discriminate against an employee regarding the employee's compensation, terms, conditions, location, or privileges of employment because the employee, or an individual acting on behalf of the employee, reports or is about to report to an appropriate federal or state authority, verbally or in writing, information that indicates that the large developer's activities pose a critical risk, unless the employee knows that the report is false.

(2) An employee who alleges a violation of subsection (1) may bring a civil action not more than 90 days after the occurrence of the alleged violation seeking 1 or more of the following:

(a) Injunctive relief.

(b) Actual damages.

(c) Reasonable attorney fees, witness fees, and court costs.

(d) Any other relief the court considers appropriate, including the reinstatement of the employee, the payment of back wages, and full reinstatement of fringe benefits and seniority rights.

(3) An employee who brings a civil action under subsection (2) must show by clear and convincing evidence that the employee, or an individual acting on behalf of the employee, was about to make a report protected by subsection (1).

(4) A civil action commenced under subsection (2) may be brought in the circuit court for the county where the alleged violation occurred, the county where the complainant resides, or the county where the person against whom the civil complaint is filed resides or has the person's principal place of business.

(5) A large developer shall do both of the following:

(a) Post notices and use other appropriate means to keep the large developer's employees informed of the employees' protections and obligations under this section.

(b) Provide a reasonable internal process through which both of the following occur:

(*i*) An employee may anonymously disclose information to the large developer if the employee believes in good faith that the information indicates the large developer's activities present a critical risk.

(*ii*) A monthly update is given to the employee under subparagraph (*i*) regarding the status of the large developer's investigation of the disclosure and any actions taken by the large developer in response to the disclosure.

(6) A large developer shall maintain the disclosures and updates provided under subsection (5)(b) for not less than 7 years after the date when the disclosure or update was created. Each disclosure and update must be shared with the officers and directors of the large developer who do not have a conflict of interest not less than once per quarter.

(7) A large developer that violates this section is subject to a civil fine of not more than $500.00. A civil fine under this subsection must be deposited into the general fund.

(8) This section does not diminish or impair the rights of a

person under any collective bargaining agreement or permit disclosures that would diminish or impair the rights of any person to the continued protection of confidentiality of communications where statute or common law provides such protection.

(9) This section does not invalidate or limit any protection afforded to an employee or any obligation imposed on an employer, including an employer that is a large developer, under the whistleblowers' protection act, 1980 PA 469, MCL 15.361 to 15.369.

Sec. 13. (1) If a large developer violates section 7 or 9, the attorney general may bring a civil action seeking 1 or both of the following:

(a) A civil fine of not more than $1,000,000.00 per violation.

(b) Injunctive or declaratory relief.

(2) In determining the relief granted under subsection (1), the court may consider both of the following:

(a) The severity of the violation.

(b) If the violation resulted in, or could have resulted in, the materialization of a critical risk.

(3) If a large developer's activities present an imminent critical risk, the attorney general may bring a civil action seeking injunctive relief.