Legislative Analysis

Phone: (517) 373-8080
http://www.house.mi.gov/hfa

Analysis available at
http://www.legislature.mi.gov

ARTIFICIAL INTELLIGENCE SAFETY AND SECURITY TRANSPARENCY ACT

House Bill 4668 as introduced
Sponsor: Rep. Sarah Lightner
Committee: Judiciary
Complete to 6-23-25

SUMMARY:

House Bill 4668 would create a new act, the Artificial Intelligence Safety and Security Transparency Act, which would require ***large developers*** of ***foundation models*** to create and implement certain risk management practices relating to the use of those models, as well as provide for the powers and duties of government officers and entities, protections for certain employees, and related civil causes of action and sanctions.

Large developer would mean a person that has developed both of the following:

- A foundation model with a quantity of computing power that costs at least $5.0 million when measured using prevailing market prices of cloud computing in the United States at the time that the computing power was used.
- Within the immediately preceding 12 months, one or more foundation models with a total quantity of computing power that costs at least $100.0 million when measured using prevailing market prices of cloud computing in the United States at the time the computing power was used.

Foundation model would mean a type of ***artificial intelligence model*** that is trained on a broad dataset, is designed for generality of output, and is adaptable to a wide range of distinctive tasks.

Artificial intelligence model (or AI model) would mean an engineered or machine-based system that varies in its 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.

Safety and security protocols

Beginning January 1, 2026, the bill would require all large developers to produce, implement, follow, and conspicuously publish a *safety and security protocol*, which would be defined as a set of documented technical and organizational protocols used by the developer to manage ***critical risks*** associated with foundation models.

Critical risk would mean 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 $1.0 million in damages to rights in money or property, through an incident of any of the following kinds:

- The creation and release of a chemical, biological, radiological, or nuclear weapon.

- A cyberattack conducted by or assisted by a foundation model.
- A foundation model engaging in conduct that meets both of the following:
  - It is performed with limited human intervention.
  - It would, if committed by an individual, constitute a crime that requires intent, recklessness, or gross negligence, or the solicitation or aiding and abetting of a crime.
- A harm as a result of an incident described under any of the above 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.

Deploy would mean to use a foundation model or to make a foundation model foreseeably available to one 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.

A large developer's security and safety protocol would have to describe all of the following in detail, as applicable:

- How the large developer excludes certain foundation models from being covered by the protocol when those models pose a limited critical risk.
- 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.
- 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 developer or user or be misused, modified, executed with increased computational resources, or used to create another foundation model.
- The procedure the large developer will use to determine whether and how to deploy a foundation model when doing so poses critical risks.
- The physical, digital, and organizational security protection the large developer will implement to prevent insiders or third parties from accessing foundation models in the developer's control in a manner that is unauthorized by the developer and could create a critical risk.
- Any safeguards and risk mitigation measures the large developer uses to reduce critical risks from the developer's foundation models and how the developer assesses efficacy and limitations.
- How the large developer will respond if a critical risk materializes or is imminent.
- The procedures that the large developer uses to determine whether to conduct additional assessments for a critical risk when the developer modifies or expands access to the developer's foundation models, or combines the foundation models with other software, and how such assessments are conducted.
- The conditions under which the large developer will report an incident relevant to a critical risk that occurs in connection with one or more of the developer's foundation models and the entities to which the developer will make those reports.
- The conditions under which the large developer will modify the developer's protocol.
- The parts of the 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.

The bill would also require the safety and security protocol to describe any other role that a financially disinterested third party would play under the provisions listed above.

If a large developer materially modifies its safety and security protocol, the developer would have to conspicuously publish the modifications no more than 30 days after they were made.

Transparency requirements

Beginning on January 1, 2026, and at least once every 90 days, large developers would have to 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 and that includes all of the following information:

- The conclusion of any risk assessments made during the reporting period in accordance with the large developer's safety and security protocol.
- 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.
- If, during the reporting period, the large developer has deployed or modified a foundation model that, if deployed without adequate safeguards and protections, would pose a higher level of critical risk than any of the developer's existing deployed foundation models, both of the following:
    - The grounds on which, and the process by which, the large developer decided to deploy the foundation model.
    - Any safeguards and protections implemented by the large developer to mitigate critical risks.

The bill would also require large developers to record and retain for five 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.

Large developers would be prohibited from knowingly making false or materially misleading statements or omissions regarding documents produced in accordance with the provisions of the bill and would be required to publish those documents on a conspicuous page on the developer's website. Large developers would be allowed to redact documents as reasonably necessary to protect the developer's trade secrets, public safety, or national security, or to comply with applicable law, provided that the developer does both of the following:

- Retains an unredacted version of the document for at least five years and provides the attorney general with the ability to inspect the unredacted document on request.
- Describes the character and justification of the redactions in the published version of the document.

Audits

Beginning on January 1, 2026, and not less than once per year, large developers would have to retain a reputable third-party auditor to produce a report that assesses all of the following:

- Whether the large developer has complied with the developer's safety and security protocol and any instances of noncompliance.

- Any instance where the large developer's safety and security protocol was not stated clearly enough to determine whether the developer has complied with the protocol.
- Any instance that the auditor believes that the large developer violated any of the provisions described in "Transparency requirements," above.

An auditor retained by a large developer would have to employ or contract one or more individuals with expertise in corporate compliance and one or more individuals with technical expertise in the safety of foundation models.

Large developers would have to grant auditors access to all materials produced in accordance with the bill and any other materials reasonably necessary to perform the assessment described above. A large developer would have to conspicuously publish the auditor's report no more than 90 days after its completion.

An auditor required to perform an audit and produce a report in accordance with the bill's provisions would also be allowed to redact information before the publication of the report, subject to the same retention requirements outlined above in "Transparency requirements."

Employee protections, civil sanctions, and remedies

The bill would also provide for protections for ***employees*** of large developers, defined as individuals who perform services for wages or salary under a contract of employment, express or implied, for an employer, including both of the following:

- A contractor or subcontractor and unpaid advisors involved with assessing, managing, or addressing a critical risk.
- A corporate officer.

A large developer would be prohibited from discharging, threatening, or otherwise discriminating 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, that the developer's activities pose a critical risk. This prohibition would not apply if the employee knows that the report is false.

Under the bill, an employee who alleges a violation of the prohibition described above could bring a civil action in circuit court[1] not more than 90 days after the occurrence of the alleged violation seeking one or more of the following:

- Injunctive relief.
- Actual damages.
- Reasonable attorney fees, witness fees, and court costs.
- 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.

[1] An employee could bring a civil action in 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.

An employee who brings a civil action under the bill would have to show by *clear and convincing evidence* that the employee (or an individual acting on behalf of the employee) was about to make a protected report.[2]

In addition, the bill would require large developers to do all of the following:

- Post notices and use other appropriate means to keep the large developer's employees informed of the employees' protections and obligations under the bill.
- Provide a reasonable internal process through which both of the following occur:
    - An employee may anonymously disclose information to the large developer if the employee believes in good faith that the information indicates that the developer's activities present a critical risk.
    - A monthly update is given to the employee described above regarding the status of the large developer's investigation of the disclosure and any actions taken by the developer in response to the disclosure.
- Maintain internal disclosures and updates provided to disclosing employees for at least seven years after the date when the disclosure or update was created. Each update would also have to be shared at least once per quarter with the officers and directors of the large developer who do not have a conflict of interest.

A large developer that violates any of the proposed employee protection provisions described above would be subject to a civil fine of up to $500, which would be deposited into the general fund.

Further, the bill would authorize the attorney general to bring a civil action seeking one or both of the following against a large developer that violates any of the bill's provisions pertaining to safety and security protocols or auditing requirements:

- A civil fine of not more than $1.0 million per violation.
- Injunctive or declaratory relief.

In determining which the type of relief granted in an attorney general–initiated civil action, a court could consider both of the following:

- The severity of the violation.
- Whether the violation resulted in, or could have resulted in, the materialization of a critical risk.

If a large developer's activities present an *imminent* critical risk, the attorney general could bring a civil action seeking injunctive relief.

The bill's provisions would not diminish or impair the rights of a person under any collective bargaining agreement or allow 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. Finally, the bill would not invalidate or limit any

[2] *Clear and convincing evidence* is an evidentiary standard that requires demonstrating that evidence is highly and substantially more likely to be true than untrue. Under this standard, the burden of proof is satisfied when the party with the burden (under House Bill 4668, the employee) convinces the judge or jury that the contention is highly probable.

protection afforded to an employee or any obligation imposed on an employer (including a large developer) under the Whistleblowers' Protection Act.[3]

FISCAL IMPACT:

House Bill 4668 would have an indeterminate fiscal impact on the state and on local units of government. Under the bill, a civil action could be brought by an employee against a large developer if the large developer discharges, threatens, or otherwise discriminates against the employee because the employee reports to the appropriate federal or state authority that the large developer's activities pose a critical risk. A large developer that violates provisions of the bill could be ordered to pay a civil fine of up to $500 under this provision. If the attorney general brings a civil action against the large developer, a civil fine of up to $1.0 million per violation could be ordered. Civil fine revenue collected under provisions of the bill would be required to be deposited into the general fund. It is not possible to determine the amount of revenue that would be collected from payment of civil fines. The fiscal impact on the judiciary and local court systems would depend on how the bill affects court caseloads and related administrative costs. It is difficult to project the actual fiscal impact to courts due to variables such as law enforcement practices, prosecutorial practices, judicial discretion, case types, and complexity of cases.

Legislative Analyst: Aaron A. Meek
Fiscal Analysts: Robin Risko
Michael Cnossen

■ This analysis was prepared by nonpartisan House Fiscal Agency staff for use by House members in their deliberations and does not constitute an official statement of legislative intent.

[3] https://legislature.mi.gov/Laws/MCL?objectName=MCL-ACT-469-OF-1980