Rep. Daniel Didech

Filed: 3/18/2025

AMENDMENT TO HOUSE BILL 3506

AMENDMENT NO. _____. Amend House Bill 3506 by replacing everything after the enacting clause with the following:

"Section 1. Short title. This Act may be cited as the Artificial Intelligence Safety and Security Protocol Act.

Section 5. Legislative findings and purpose. The General Assembly finds and declares:

(a) Artificial intelligence, including new advances in generative artificial intelligence, has the potential to catalyze innovation and the rapid development of a wide range of benefits for Illinoisans and the Illinois economy, including advances in medicine, climate science, and education, and to push the bounds of human creativity and capacity.

(b) If not properly subject to human controls, future development in artificial intelligence may also have the

potential to be used to create novel threats to public safety and security, including by enabling the creation and the proliferation of weapons of mass destruction, such as biological, chemical, and nuclear weapons, as well as weapons with cyber-offensive capabilities.

(c) If not properly subject to human controls, future artificial intelligence models may be able to cause serious harm with limited human intervention.

(d) This State has an essential role in fostering transparency, security, and reasonable care in the development of the most powerful artificial intelligence systems, in order to protect the safety, health, and economic interests of this State.

(e) Actions taken by large developers that reduce consumer prices for access to foundation models, increase the ability of artificial intelligence safety and security researchers to conduct research, increase interoperability between foundation models produced by different large developers, improve the ability for small businesses to use foundation models, and promote privacy of user inputs to foundation models provide important societal benefits.

Section 10. Definitions. As used in this Act:

"Adverse employment action" means an action that a reasonable employee would find materially adverse. For the purpose of this definition, an action is materially adverse if

it could dissuade a reasonable worker from disclosing or threatening to disclose information protected under Section 30.

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

"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 more than $1,000,000,000 in damage to rights in money or property, through an incident of the following types:

(1) the creation and release of a chemical, biological, radiological, or nuclear weapon;

(2) a cyber-attack conducted by or assisted by a foundation model; or

(3) a foundation model engaging in conduct that, would, if committed by a human, constitute a crime specified under the Criminal Code of 2012 that requires intent, recklessness, or gross negligence, or the solicitation or aiding and abetting of the crime, if that conduct occurs with limited human intervention.

For the purposes of this definition, a harm inflicted by an intervening human actor does not result from the large

developer's activities unless those activities make it substantially easier or more likely for the actor to inflict the harm.

"Deploy" means 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.

"Employee" means any individual permitted to work by a large developer. "Employee" includes any corporate officers of the large developer and any contractors, subcontractors, and unpaid advisors involved with assessing, managing, or addressing critical risk.

"Foundation model" means an artificial intelligence model that:

(1) is trained on a broad data set;

(2) is designed for generality of output; and

(3) can be adapted to a wide range of distinctive tasks.

"Large developer" means a person that has:

(1) developed a foundation model with a quantity of computing power that costs at least $5,000,000 when measured using prevailing market prices of cloud computing in the United States at the time that training of the foundation model commences; and

(2) within the immediately preceding 12 months, developed one or more foundation models using a total quantity of computing power that costs at least $100,000,000 as measured by the average cost of an equivalent amount of cloud computing in the United States at the time the computational power was used.

"Retaliatory action" means an adverse employment action or the threat of an adverse employment action by a large developer or a large developer's agent to penalize or any non-employment action that would dissuade a reasonable worker from disclosing information under this Act. "Retaliatory action" includes, but is not limited to:

(1) taking, or threatening to take, any action that would intentionally interfere with an employee's ability to obtain future employment or post-termination retaliation to intentionally interfere with a former employee's employment;

(2) taking, or threatening to take, any action prohibited by subsection (G) of Section 2-102 of the Illinois Human Rights Act; or

(3) contacting, or threatening to contact, United States immigration authorities, or otherwise reporting, or threatening to report, an employee's suspected or actual citizenship or immigration status or the suspected or actual citizenship or immigration status of an employee's family or household member to a federal, State, or local

agency.

"Retaliatory action" does not include:

(1) conduct undertaken at the express and specific direction or request of the federal government;

(2) truthful, performance-related information about an employee or former employee provided in good faith to a prospective large developer at the request of the prospective large developer; or

(3) conduct undertaken if specifically required by State or federal law.

"Safety and security protocol" means a set of documented technical and organizational protocols used by a large developer to manage critical risks that describes in detail:

(1) how, if at all, the large developer excludes certain foundation models from being covered by its safety and security protocol when those foundation models pose limited critical risks;

(2) thresholds at which critical risks would be deemed intolerable and justifications for these thresholds and what the large developer will do if one or more thresholds are surpassed;

(3) the testing and assessment procedures the large developer uses to investigate critical risks and how these tests account for the possibility that a foundation model could evade the control of its large developer or user or be misused, modified, executed with increased

computational resources, or used to create another foundation model;

(4) the procedure the large developer will use to determine whether and how to deploy a foundation model when doing so poses critical risks;

(5) the physical, digital, and organizational security protections 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 large developer and could create critical risk;

(6) any safeguards and risk mitigation measures the large developer uses to reduce critical risks from its foundation models and how the large developer assesses their efficacy and limitations;

(7) how the large developer will respond if a critical risk materializes or is imminently about to materialize;

(8) the procedure that the large developer uses to determine whether to conduct additional assessments for critical risk when it modifies or expands access to its foundation models or combines its foundation models with other software and how the assessments are conducted;

(9) the conditions under which the large developer will report incidents relevant to critical risk that have occurred in connection with one or more of its foundation models and the entities to which the large developer will

make those reports;

(10) the conditions under which the large developer may or will make modifications to its safety and security protocol;

(11) the parts of the safety and security protocol, if any, 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, if any, unredacted versions are made available; and

(12) any other role, if any, financially disinterested third parties play in the implementation of the other items of this definition.

"Supervisor" means any individual who has the authority to direct and control the work performance of the affected employee or any individual who has managerial authority to take corrective action concerning critical risk in accordance with Section 30.

Section 15. Safety and security protocol.

(a) A large developer shall produce, implement, follow, and conspicuously publish a safety and security protocol. If a large developer makes a material modification to the safety and security protocol, the large developer shall conspicuously publish those modifications no later than 30 days after the effective date of those modifications.

(b) No less than every 90 days, a large developer shall produce and conspicuously publish a transparency report. The transparency report shall cover the period between 120 and 30 days before the submission of the transparency report `and include the following:

(1) the conclusion of any risk assessments made pursuant to the large developer's safety and security protocol during the reporting period;

(2) if different from the preceding reporting period, for each type of critical risk, an assessment of the relevant capabilities in 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; and

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

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

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

(c) A large developer shall record and retain for a period

of no less than 5 years any specific tests used and test results obtained as part of any assessments of critical risks, including sufficient detail for qualified third parties to replicate the testing.

(d) A large developer shall not knowingly make false or materially misleading statements or omissions in or regarding documents produced under this Section.

Section 20. Redactions.

(a) If a large developer publishes documents in order to comply with this Act, the large developer may make redactions to those documents that are reasonably necessary to protect the large developer's or auditor's trade secrets, public safety, or the national security of the United States or to comply with any federal or State law. If a large developer redacts information in a document, the large developer shall:

(1) retain an unredacted version of the document for at least 5 years and allow the Attorney General to inspect the unredacted version of the document upon request; and

(2) describe the character and justification of the redaction in any published version of the document, to the extent permitted by the concerns that justify redaction.

(b) In addition to a large developer's redactions, the auditor may also redact information using the same procedure described in subsection (a) for information that the large developer is required to publish in accordance with subsection

(c) of Section 25 before the publication of that information.

Section 25. Audits.

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

(1) whether the large developer has complied with its safety and security protocol and any instances of noncompliance;

(2) any instances where the large developer's safety and security protocol has not been stated clearly enough to determine whether the large developer has complied; and

(3) any instances where the auditor believes the large developer may have violated subsection (d) of Section 15 or Section 20.

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

(c) No later than 90 days after the completion of the third-party auditor's report required under subsection (a), the large developer shall conspicuously publish the report.

(d) In conducting the audit, the auditor shall 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.

Section 30. Whistleblower protections.

(a) A large developer that has one or more employees in this State shall provide a reasonable internal process through which an employee may anonymously disclose information to the large developer if the employee believes in good faith that information indicates that the large developer's activities pose critical risk, including a monthly update to the person who made the disclosure regarding the status of the large developer's investigation of the disclosure and the actions taken by the large developer in response to the disclosure.

(b) The disclosures and responses of the process required by subsection (a) shall be maintained for a minimum of 7 years after the date when the disclosure is made to the large developer or the response to the disclosure is made by the large developer. Each disclosure and response shall be shared with the officers and directors of the large developer who do not have a conflict of interest no less frequently than once every fiscal quarter.

(c) A large developer that has one or more employees in this State shall not make, adopt, or enforce any rule, regulation, or policy preventing an employee from disclosing information to a government or law enforcement agency if the employee has reasonable cause to believe that the information discloses that the large developer's activities pose critical risk.

(d) A large developer that has one or more employees in this State shall not take retaliatory action against an employee for disclosing or threatening to disclose information to a government or law enforcement agency information related to an activity, policy, or practice of the large developer, where the employee has a good faith belief that the activity, policy, or practice of the large developer poses critical risk.

(e) A large developer that has one or more employees in this State shall not take retaliatory action against an employee for disclosing or threatening to disclose to any supervisor, principal officer, or board member, information related to an activity, policy, or practice of the large developer if the employee has a good faith belief that the activity, policy, or practice creates critical risk.

(f) A large developer that has one or more employees in this State shall not threaten any employee with any act or omission if that Act or omission would constitute retaliatory action against the employee under this Section.

(g) It is a defense to any action brought under this Section that the retaliatory action was predicated solely upon grounds other than the employee's exercise of any rights protected under this Section or the Whistleblower Act.

(h) This Section does not apply to a disclosure that would constitute a violation of attorney-client privilege.

(i) Nothing in this Section shall be construed to

invalidate or limit any protection afforded to employees or any obligation imposed on employers, including those that are large developers, under the Whistleblower Act.

Section 35. Enforcement of safety and security protocols.

(a) The Attorney General may bring a civil action against a large developer that violates Sections 15 or 25. A large developer found guilty of violating Sections 15 or 25 may be assessed a civil penalty not to exceed $1,000,000. In calculating the civil penalty assessed under this subsection, a court shall consider the severity of the violation and whether the violation resulted in, or could have resulted in, the materialization of a critical risk.

(b) The Attorney General may seek injunctive or declaratory relief for any violation of this Act. The Attorney General may also seek injunctive relief if a large developer's activities present an imminent threat of catastrophic harm to the public.

(c) In determining whether a large developer's act or omission breached its common law duty to take reasonable care with respect to critical risks, the following considerations are relevant but not conclusive:

(1) the quality of the large developer's safety and security protocol and the extent of the large developer's adherence to it;

(2) whether, in quality and implementation, the large

developer's investigation, documentation, evaluation, and management of critical risks was inferior, comparable, or superior to other large developers of foundation models that may pose comparable critical risk;

(3) the extent to which the large developer responsibly informed the public of critical risks posed by its foundation models; and

(4) whether the societal benefit produced by the large developer's act or omission outweighed the associated critical risk.

Section 40. Enforcement of whistleblower protections.

(a) Whenever the Attorney General has reasonable cause to believe that any large developer has engaged in a practice prohibited by Section 30, the Attorney General may, pursuant to the authority conferred by Section 6.3 of the Attorney General Act, initiate or intervene in a civil action in the name of the People of the State in any appropriate court to obtain appropriate relief.

(b) Before initiating an action, the Attorney General may conduct an investigation and may:

(1) require a large developer to file a statement or report in writing, under oath or otherwise, as to all information the Attorney General may consider necessary;

(2) examine under oath any person alleged to have participated in, or with knowledge of, the alleged

violation; or

(3) issue subpoenas or conduct hearings in aid of any investigation.

(c) Service by the Attorney General of any notice requiring a large developer to file a statement or report, or of a subpoena upon any large developer, shall be made:

(1) personally by delivery of a duly executed copy thereof to the person to be served or, if a person is not a natural person, in the manner provided in the Code of Civil Procedure when a complaint is filed; or

(2) by mailing by certified mail a duly executed copy thereof to the person to be served at the person's last known abode or principal place of business within this State or, if the person is not a natural person, in the manner provided in the Code of Civil Procedure when a complaint is filed. The Attorney General may compel compliance with investigative demands under this Section through an order by any court of competent jurisdiction.

(d)(1) In an action brought under this Act, the Attorney General may obtain, as a remedy, monetary damages to the State, restitution, and equitable relief, including any permanent or preliminary injunction, temporary restraining order, or other order, including an order enjoining the defendant from engaging in a violation, or order any action as may be appropriate. The Attorney General may request, and the court may grant, any remedy available under Section 45 to the

employee or employees affected by the violation. Additionally, the Attorney General may request and the court may impose a civil penalty not to exceed $10,000 for each repeat violation within a 5-year period. For purposes of this Section, each violation of this Act for each employee that the large developer took or threatened to take retaliatory action against shall constitute a separate and distinct violation.

(2) A civil penalty imposed under this subsection shall be deposited into the Attorney General Court Ordered and Voluntary Compliance Payment Projects Fund.

Section 45. Civil damages for employees. If a large developer takes any retaliatory action against an employee in violation of Section 30, the employee may bring a civil action against the large developer for all relief necessary to make the employee whole, including but not limited to the following, as appropriate:

(1) permanent or preliminary injunctive relief;

(2) reinstatement with the same seniority status that the employee would have had, but for the violation;

(3) back pay, with interest of 9% per annum up to 90 calendar days from the date the complaint is filed and front pay;

(4) liquidated damages of up to $10,000;

(5) compensation for any costs incurred as a result of the violation, including litigation costs, expert witness

fees, and reasonable attorney's fees; and

(6) additionally, the court shall award a civil penalty of $10,000 payable to the employee.

Section 50. Other duties required by law. The duties and obligations imposed by this Act are cumulative with any other duties or obligations imposed under other law and shall not be construed to relieve any party from any duties or obligations imposed under other law and do not limit any rights or remedies under existing law.

Section 97. Severability. The provisions of this Act are severable under Section 1.31 of the Statute on Statutes.".