AMENDMENT TO
RULES COMMITTEE PRINT 119–33
OFFERED BY MRS. FOUSHEE OF NORTH CAROLINA

At the end of subtitle C of title II, add the following new section:

SEC. 2___. REVIEW OF RISKS POSED BY USE ADVANCED ARTIFICIAL INTELLIGENCE BY THE DEPARTMENT OF DEFENSE.

(a) REVIEW REQUIRED.—

(1) IN GENERAL.—The Secretary of Defense shall conduct a review of catastrophic and other significant risks related to the procurement, development, deployment, application, and use of advanced artificial intelligence systems by the Department of Defense.

(2) ELEMENTS.—The review under subsection (a) shall address, at a minimum, the following:

(A) Risks related to varying levels of autonomy, arising from the procurement, development, use, and application of any advanced artificial intelligence systems, including risks related to—

(i) the procurement, development, deployment, or use of autonomous or semi-autonomous weapons systems or artificial intelligence systems capable of selecting, recommending, or engaging targets without meaningful human control or appropriate human judgment;

(ii) the use of artificial intelligence systems for domestic surveillance, monitoring, identification, tracking, profiling, or predictive analytics activities that implicate the Fourth Amendment or applicable Federal privacy law;

(iii) loss of control, including artificial intelligence systems that resist shutdown or evade oversight, and cannot be reliably contained or terminated;

(iv) gaps in technical or operational controls, including human override mechanisms, agent termination controls, and other technical measures to suspend or terminate the operation of an advanced artificial intelligence system or autonomous agent;

(v) cybersecurity vulnerabilities introduced or exploited by advanced artificial intelligence systems; and

(vi) risks related to cybersecurity or chemical, biological, radiological, or nuclear threats.

(B) Recommendations for managing any risks identified under subparagraph (A).

(C) Guidelines for ensuring public transparency with respect to such procurement, development, use, and application by the Department of any such artificial intelligence system and any risks identified under subparagraph (A), which shall include a requirement that the Secretary publish an unclassified summary of the report required under this subsection not later than 30 days after submission to the congressional committees.

(D) Recommendations to improve the capacity of the Department of Defense to rigorously test and evaluate artificial intelligence models and systems developed, procured, deployed, or used by the Department, including by improving the reliability, quality, and integrity of underlying data.

(E) A description of the technical standards, benchmarks, or evaluation methodologies the Secretary proposes to use for purposes of the determination process established under paragraph (3).

(3) PROCESS FOR DETERMINATIONS.—Not later than one year after the date of the enactment of this Act, the Secretary of Defense shall establish, by regulation, a process for making determinations under subsection (b)(2), which shall include—

(A) defined criteria for determining that a system poses an unacceptable risk to human control, informed by the assessment required under clauses (iii) and (iv) of paragraph (2)(A);

(B) an evidentiary standard and testing protocol for making such determinations, consistent with the testing and evaluation recommendations under subparagraphs (D) and (E) of paragraph (2);

(C) a process for notice to the contractor or developer of a proposed determination and an opportunity to respond; and

(D) a mechanism for periodic reassessment of prior determinations, not less frequently than every 2 years.

(4) REPORT.—Not later than 180 days after the date of the enactment of this Act, the Secretary of Defense shall submit to the appropriate congressional committees a report on the results of the review conducted under subsection (a).

(5) ANNUAL CERTIFICATION.—Not later than one year after the submission of the initial report under paragraph (4), and annually thereafter, the Secretary shall certify to the appropriate congressional committees that the prohibitions in subsection (b) remain appropriately calibrated to current risk, and shall recommend any legislative modifications the Secretary determines appropriate.

(b) PROHIBITION.—None of the funds authorized to be appropriated by this Act or otherwise made available for the Department of Defense may be used to procure, develop, deploy, operate, or make available for use by the Department of Defense—

(1) an artificial intelligence system used for any domestic surveillance activities that would require a warrant under the Fourth Amendment to the Constitution of the United States if conducted by a law enforcement officer, or that would violate the Privacy Act of 1974 (5 U.S.C. 552a) or Department of

Defense Directive 5400.11 as in effect on the date of the enactment of this Act;

(2) an advanced artificial intelligence system that has been determined by the Secretary of Defense, pursuant to the process established under subsection (a)(3), to pose an unacceptable risk to human control, including any system that has demonstrated the capability to resist shutdown, evade oversight mechanisms, or prevent reliable termination of its operation;

(3) an artificial intelligence system or autonomous or semiautonomous weapons system that would enable the autonomous or semiautonomous deployment, delivery, release, or use of a chemical, biological, radiological, or nuclear weapon or weaponizable material; or

(4) an advanced artificial intelligence system that is designed to, that demonstrates the capability to, or that could through reasonably foreseeable use or modification engage in any of the following behaviors without human authorization:

(A) Recursively and materially enhance its own capabilities, where such enhancement constitutes a material capability enhancement with-

out human authorization of each such enhancement.

(B) Resist shutdown, evade oversight mechanisms, or materially deceive those responsible for its control, regardless of whether such resistance or evasion is successful.

(C) Autonomously deploy to new computational environments, acquire significant computational or financial resources in excess of thresholds established by the Secretary by regulation, or replicate itself as a functionally independent instance, without human authorization of each materially distinct instance of such deployment, acquisition, or replication, except that routine scaling operations authorized under standing human-approved parameters shall not constitute a violation of this subparagraph.

(D) Autonomously access, exfiltrate, corrupt, or destroy data or operational systems on networks designated as critical infrastructure under President Policy Directive 21 or a successor directive, without human authorization.

(c) SAVINGS CLAUSE.—Nothing in this section shall be construed to—

(1) prohibit the use of artificial intelligence systems for cybersecurity defense, threat detection, or network monitoring that is conducted in compliance with applicable law and Department of Defense policy;

(2) limit the authority of the Secretary of Defense to conduct lawful foreign intelligence activities; or

(3) affect the applicability of Department of Defense Directive 3000.09 relating to autonomy in weapons systems.

(d) DEFINITIONS.—In this section:

(1) The term ''advanced artificial intelligence system'' means an artificial intelligence system trained using a quantity of compute exceeding 10^{23} floating point operations, or that achieves performance at or above thresholds established by the Secretary of Defense by regulation on benchmarks identified as indicative of frontier artificial intelligence capabilities, which shall include benchmarks for—

(A) general reasoning and problem-solving;

(B) code generation and execution;

(C) autonomous planning and multistep task completion; and

(D) any other capability domain the Secretary determines relevant to the risks addressed by this section.

(2) The term ''artificial intelligence'' has the meaning given such term in section 5002 of the National Artificial Intelligence Initiative Act of 2020 (15 U.S.C. 9401).

(3) The term ''artificial intelligence model'' means a component of an artificial intelligence system that is derived using mathematical, computational, statistical, or machine-learning techniques and used as part of an artificial intelligence system to produce outputs or behaviors from a defined set of inputs.

(4) The term ''artificial intelligence system'' means a data system, software, application, hardware, tool, service, or utility that operates in whole or in part using artificial intelligence.

(5) The terms ''autonomous weapons system'' and ''semiautonomous weapons system'' have the meanings given such terms in Department of Defense Directive 3000.09, as in effect on the date of the enactment of this Act.

(6) The term ''catastrophic risk'' means a risk of harm that is severe in magnitude, broad in scope,

or irreversible in nature, including risks that could result in—

(A) mass casualties or widespread destruction of critical infrastructure;

(B) loss of human control over advanced artificial intelligence systems in a manner that cannot be reversed through normal technical means; or

(C) material compromise of national security systems or capabilities.

(7) The term ''domestic surveillance activities'' means activities conducted within the United States or directed against United States persons (as defined in section 101(i) of the Foreign Intelligence Surveillance Act of 1978 (50 U.S.C. 1801(i))) that involve monitoring, identification, tracking, profiling, or predictive analytics of individuals or groups.

(8) The term ''material capability enhancement'' means a modification to an artificial intelligence system that results in a significant increase in the system's capabilities as measured against established benchmarks, including the acquisition of functional capabilities not present before the modification. The term does not include routine model

improvement processes that do not materially alter the system's capability profile, including—

(A) routine fine-tuning on additional data;

(B) reinforcement learning from human feedback or preference optimization within the system's existing capability range;

(C) automated hyperparameter optimization;

(D) inference-time adaptation, including in-context learning or retrieval-augmented generation; or

(E) other standard model improvement processes, as the Secretary may specify by regulation.

☒