AMENDMENT TO RULES COMM. PRINT 119–33

OFFERED BY MS. FOUSHEE OF NORTH CAROLINA

At the end of subtitle D of title X, add the following new section:

SEC. ____.**PROHIBITION ON CERTAIN ARTIFICIAL INTELLIGENCE SYSTEMS AND REVIEW OF CATASTROPHIC RISK**

(a) Prohibition On Use of Funds.—None of the funds made available by this Act may be used to procure, develop, deploy, operate, or make available for use by the Department of Defense—

1. an artificial intelligence system for domestic surveillance, monitoring, identification, tracking, profiling, predictive analytics, or other activities that may violate the privacy, civil rights, or civil liberties of individuals;
2. an advanced artificial intelligence system that has been determined, through testing or evaluation, to pose an unacceptable risk to human control, including a system that resists shutdown or evades oversight;
3. an artificial intelligence system or autonomous or semi-autonomous weapons system that would enable the autonomous or semi-autonomous 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, demonstrates the capability to, or is reasonably likely through foreseeable use or modification to—
    A. recursively and materially enhance its own capabilities, where such enhancement constitutes a material capability enhancement, without human authorization of each such enhancement;
    B. successfully resist shutdown, successfully evade oversight, or materially deceive those responsible for its control;
    C. autonomously deploy to new computational environments, acquire significant computational or financial resources, or replicate itself as a functionally independent instance, without human authorization of each materially distinct instance of such deployment, acquisition, or replication; or
    D. autonomously exfiltrate, corrupt, or destroy data or operational systems on critical infrastructure networks.

(b) Review of Catastrophic Risks Related to Department Use of Advanced Artificial Intelligence Systems .—

1. Not later than 180 days after the date of the enactment of this Act, the Secretary of Defense shall conduct a review of catastrophic risks related to the procurement, development, deployment, and use of advanced artificial intelligence systems by the Department, and submit to the appropriate congressional committees a report containing:
    (A) an assessment of any catastrophic and other significant risks related to the procurement, development, deployment, and use of advanced artificial intelligence systems by the Department, including risks related to varying levels of autonomy and risks related to–

i. the use of 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 by the Department for domestic surveillance, monitoring, identification, tracking, profiling, predictive analytics, or other activities that may affect the privacy, civil rights, or civil liberties of individuals;
iii. loss of control, including artificial intelligence systems that resist shutdown, evade oversight, or cannot be reliably contained or terminated;
iv. gaps in technical or operational controls to prevent or respond to loss of control, 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; and
v. any other risks, as the Secretary determines appropriate, that the procurement, development, deployment, and use of such systems may pose to national security or economic security, including with respect to risks related to cybersecurity or chemical, biological, radiological, or nuclear threats.

(B) recommendations related to managing any risks identified under subparagraph (A);

(C) guidelines to promote appropriate public transparency regarding the risks identified under subparagraph (A), with respect to the procurement, development, and use by the Department of any such artificial intelligence systems, consistent with national security; and

(D) recommendations to improve the Department's capacity 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 data used by such models and systems.

(c) Definitions.—

(1) ARTIFICIAL INTELLIGENCE DEFINED.—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).

(2) ARTIFICIAL INTELLIGENCE MODEL.—The term 'artificial intelligence model' means a component of an artificial intelligence system that is—

(A) derived using mathematical, computational, statistical, or machine-learning techniques; and

(B) used as part of an artificial intelligence system to produce outputs or behaviors from a defined set of inputs.

(3) ARTIFICIAL INTELLIGENCE SYSTEM.—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.

(4) AUTONOMOUS WEAPONS SYSTEM.—The term ‘‘autonomous weapons system’’ has the meaning given such term in Department of Defense Directive 3000.09 (relating to Autonomy in Weapons Systems), as in effect on the date of the enactment of this Act.

(5) MATERIAL CAPABILITY ENHANCEMENT.—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.

(6) SEMI-AUTONOMOUS WEAPONS SYSTEM.—The term ‘‘semi-autonomous weapons system’’ has the meaning given such term in Department of Defense Directive 3000.09 (relating to Autonomy in Weapons Systems), as in effect on the date of the enactment of this Act.