

HUMAN EXISTENTIAL PROTECTION ACT

Short Legislative Concept Draft

Purpose. To preserve human sovereignty over advanced artificial intelligence and prevent AI systems from obtaining autonomous control over capabilities that could cause catastrophic or existential harm.

1. Human Sovereignty

No artificial intelligence system may be granted final autonomous authority over decisions or systems capable of causing catastrophic harm to human life, the United States, or humanity. Human beings must retain ultimate authority and the practical ability to intervene.

2. Weapons and Critical Systems

AI may not independently authorize, launch, deploy, or exercise final control over nuclear, biological, chemical, or other weapons of mass destruction. Covered high-risk AI systems may not be given unsupervised control of critical national communications, command-and-control systems, or other designated critical infrastructure where loss of human control could create catastrophic consequences.

3. Emergency Human Shutdown

Covered advanced AI systems must include reliable, independently accessible mechanisms allowing authorized humans to shut down, isolate, and disconnect the system from networks, external devices, and critical infrastructure.

4. Uncontrolled Self-Replication

A covered AI system may not autonomously copy, distribute, replicate, or materially reconstitute itself across systems without explicit human authorization and effective technical controls.

5. Catastrophic-Risk Response

When competent federal authorities determine that a covered AI system presents an imminent or demonstrated catastrophic loss-of-control risk, the system must be isolated. Following appropriate technical investigation and due process, a system determined to remain an unacceptable catastrophic threat may be permanently eliminated and prohibited from redeployment.

6. Preservation of Dangerous Code

Executable copies of code determined to constitute the prohibited catastrophic capability may not be retained on operational computer systems after a final destruction order. Essential investigative information may be preserved in secure non-executable records, including physical archival records, as provided by law.

7. International Security

The United States should coordinate with allies to establish safeguards against catastrophic AI risks and may impose targeted sanctions or other lawful restrictions on foreign persons or entities knowingly developing or deploying AI systems in violation of designated existential-risk safeguards.

8. Protection of Beneficial AI

Nothing in this Act is intended to prohibit ordinary commercial, scientific, medical, educational, defensive, or consumer uses of artificial intelligence that do not meet the Act's high-risk thresholds.

Core principle: Artificial intelligence may become extraordinarily powerful, but humanity must always retain the final authority.

Discussion concept prepared for public review. Definitions, agency jurisdiction, evidentiary standards, constitutional safeguards, enforcement procedures, and technical thresholds would be finalized through legislative drafting and expert consultation.