13 January 2025

# The Policymaker's Guide to Artificial Intelligence *in the 119th Congress*

Contact: Hamza Chaudhry, **hamza@futureoflife.org** | View online: futureoflife.org/ai-guide

## What is Artificial Intelligence (AI)?

AI is a branch of computer science devoted to developing data processing systems that perform functions normally associated with human intelligence, such as reasoning, learning, and self-improvement.

**Tool AI / Narrow AI:** Systems designed to perform specific, well-defined tasks within a limited domain. These systems have driven revolutionary scientific progress. For example, DeepMind's AlphaFold quickly and accurately predicts protein structures—a task that once took scientists years of research.

**Dual-Use / General-Purpose AI (GPAI):** Powerful systems capable of solving a wide array of tasks across domains without being specifically designed for them. An example is GPT-4o, OpenAI's "high-intelligence flagship model for complex, multi-step tasks." Advanced GPAIs with emergent capabilities are often called "frontier" systems.

**Artificial General Intelligence (AGI):** An AI system capable of outperforming humans at virtually all cognitive tasks. Creating AGI is the stated goal of corporations like OpenAI, DeepMind, Meta, and x.AI.

| Key Terms and Acronyms | |
|---|---|
| AISI | The Artificial Intelligence Safety Institute, housed within the National Institute of Standards and Technology (NIST), researches AI safety standards, evaluates models, and develops guidelines. |
| Agents | AI systems that can perceive their environment, make decisions, and take actions to achieve specific goals. |
| Benchmarks | Standardized tests or evaluation criteria used to measure and compare the performance, capabilities, and safety of AI systems. |
| Compute security | Protecting the systems that power AI, including hardware and software, from hacking, misuse, and unauthorized access. Hardware security falls under the umbrella of compute security. |
| DL | Deep learning is a subset of machine learning that uses artificial neural networks, much like the brain, to process data through layered architectures, enabling tasks like image recognition. |
| Fine Tuning | Adapting a pre-trained AI model to a specific task or to a new dataset. |
| Hardware security | Ensuring that physical devices powering AI systems, such as GPUs, TPUs, and other specialized chips, are protected against tampering, hacking, and unauthorized access. |
| LLM | Large Language Models are AI systems trained on massive datasets to understand and generate human-like text. |
| Model Weights | The numerical parameters in a machine learning model that are adjusted during training to decide how important different pieces of information are when making a prediction or decision. |
| OS | Open source refers to software or projects where the source code is freely available for anyone to view, modify, and distribute. |
| Red Teaming | Evaluating a model by testing it with challenging or adversarial inputs to identify vulnerabilities. |

## State of Play

### Within Congress: 120+ AI-related bills were introduced last year.

- **Deepfakes Bills:** Several proposals sought to introduce measures to hold accountable users who generate harmful deepfakes. However, these proposals stopped short of suggesting liability for developers.
- **Blumenthal-Hawley Framework:** Called for legal accountability and for defending national security by holding AI companies liable for privacy breaches, civil rights violations, and harmful content while leveraging export controls and sanctions to restrict AI advancements from adversaries.
- **AI Research, Innovation, & Accountability Act:** This bill called for a self-certification framework aiming to foster AI innovation while establishing measures like risk management for high-risk AI applications.
- **Romney Framework:** Proposes a federal oversight mechanism to mitigate extreme risks posed by frontier models. This would include oversight of computing hardware, implementing safeguards, and licensing based on the model's risk level.
- **Bipartisan Senate AI Working Group Roadmap:** This document recognized the unpredictability and risk of developing increasingly advanced GPAIs and emphasized the need to "hold AI developers and deployers accountable if their products or actions cause harm to consumers".

### Beyond Congress:

- **White House National Security Memo (NSM):** Prohibits AI applications that violate laws, mandates cataloging and transparency through annual inventories of high-impact AI use, establishes Chief AI Officers and Governance Boards for oversight and accountability, and calls for whistleblower protections to report misuse. It also mandates the DoD to conduct classified evaluations of advanced AI models.
- **NIST Risk Management Framework (RMF):** Provides a voluntary framework to manage risks and promote trustworthiness through safety measures such as corporate governance, internal monitoring, and transparency throughout the AI lifecycle.
- **Voluntary Commitments:** Several leading AI companies have committed to safety measures including internal and external security testing, allowing third-party scrutiny and reporting of vulnerabilities. They have also committed to publicly disclosing their systems' capabilities.
- **Executive Order #14110:** Mandates measures including NIST developing standards, tools, and tests to ensure AI systems are safe; establishing new standards for biological synthesis screening; and requiring that developers of the most powerful AI systems share their test results with the government.

## Contextual Risks & Considerations

**Cybersecurity:** Enabling adversaries to launch more sophisticated, deadly attacks on the United States. These attacks could involve highly targeted phishing schemes, automating hacking processes, or developing malware that evades detection.

**Loss of Control:** A superintelligent AI system acting unpredictably or against human intentions, potentially causing harm by pursuing objectives misaligned with its original purpose or exceeding intended limits.

**Chemical, Biological, Radiological, and Nuclear (CBRN):** Accidentally causing the development or release of toxic chemicals, infectious diseases, harmful radiation, or nuclear weapons. In one experiment, researchers were able to use an AI system to identify 40,000 chemical weapon compounds.

**Deepfakes:** Realistic AI-generated content used for sexual exploitation, fraud, or electoral manipulation.

**Concentration of Power:** Allowing AI to centralize financial, political, and informational control, leaving power in the hands of a few elites, corporations, or even AI systems themselves.

**Amplification of Bias:** Learning from skewed or unrepresentative training data and producing politically or demographically biased outputs.

**Job Loss:** Replacing human workers in both blue-collar and graduate jobs.