April 16, 2024

The Honorable Chuck Schumer
Majority Leader
United States Senate

The Honorable Mike Rounds
Senator
United States Senate

The Honorable Martin Heinrich
Senator
United States Senate

The Honorable Todd Young
Senator
United States Senate

Dear Leader Schumer and Senators Rounds, Heinrich, and Young,

We appreciate your efforts over the past year to educate senators and staff on both the opportunities and risks posed by developments in artificial intelligence (AI). As the Senate’s AI Insight Forums, scientific research, and broader policy discussions have highlighted, advancements in artificial intelligence have the potential to dramatically improve and transform our way of life, but also present a broad spectrum of risks that could be harmful to the American public. Even as we focus on the tremendous benefits, experts have warned that AI could perpetuate disinformation,¹ fraud,² bias,³ and privacy concerns.⁴ Others have voiced concerns that AI could pose threats to election integrity⁵ and the future of the workforce.⁶ As you develop a framework for legislation, considering solutions to these problems will be important. However, any comprehensive framework to address risks from AI should also include measures to guard against the potential catastrophic risks with respect to biological, chemical, cyber, and nuclear weapons.

According to the U.S. government, academia, and distinguished experts, advancements in AI have the potential to be misused by bad actors. The Department of Defense,⁷ the Department of State,⁸ the U.S. Intelligence Community,⁹ and the National Security Commission on Artificial Intelligence,¹⁰ as well as senior officials at the Department of Energy,¹¹ Argonne National Laboratory,¹² the Cybersecurity and Infrastructure Security Agency,¹³ and the National Counterterrorism Center,¹⁴ have underscored that advanced AI poses risks to U.S. national security, including the development of biological, chemical, cyber, or nuclear weapons.

A September 2023 hearing titled, “Advanced Technology: Examining Threats to National Security,” held by the Senate Homeland Security and Governmental Affairs Subcommittee on Emerging Threats and Spending Oversight, heard testimony that advanced AI models could facilitate or assist in the development of extreme national security risks, and that the U.S. government may lack authorities to adequately respond to such risks posed by broadly capable, general purpose frontier AI models.¹⁵ In a worst-case scenario, these models could one day be leveraged by terrorists or adversarial nation state regimes to cause widespread harm or threaten U.S. national security.

At another September 2023 hearing before the Senate Energy and Natural Resources Committee, Dr. Rick Stevens, the Associate Laboratory Director for Computing, Environment, and Life Sciences at the Argonne National Laboratory, testified that in the future, “A small group working in secret with sufficiently powerful AI tools could develop a novel chemical, biological, or cyber
threat. We will need to transform how we manage the risks posed by bad actors using the same AI tools we are using to improve science and advance society.\(^{16}\)

The overlap between AI and biotechnology could lead to “the deliberate and incidental creation” of novel public health risks, according to the Office of Intelligence and Analysis (I&A) at the Department of Homeland Security (DHS).\(^{17}\) Researchers at Carnegie Mellon have found that large language models (LLMs) can assist in biological and chemical research but also “raise substantial concerns about the safety and potential dual use consequences, particularly in relation to the proliferation of illicit activities and security threats.”\(^{18}\) Other findings from the RAND Corporation,\(^{19}\) Gryphon Scientific,\(^{20}\) and individuals affiliated with the Massachusetts Institute of Technology, Harvard University, SecureBio, and SecureDNA\(^{21}\) highlight that certain AI models could produce outputs that could assist in the development of bioweapons or execution of a biological attack. Currently, much of this information can be found online by a dedicated party, particularly if they have domain expertise; however, the risks become clearer when we consider the implications of having this knowledge aggregated in one tool, accessible to non-experts who may be using simple prompts.\(^{22}\)

While powerful AI models may be beneficial for cybersecurity defenses, they can also be leveraged to bolster cyber offensive capabilities to assist bad actors in creating customized malware or automating cyber attacks at a larger scale and higher speed.\(^{23}\) According to DHS I&A, the proliferation of AI could help facilitate “larger-scale, faster, efficient, and more evasive cyber attacks.”\(^{24}\) FBI Director Christopher Wray likewise warned that “AI is going to enable threat actors to develop increasingly powerful, sophisticated, customizable, and scalable capabilities, and it’s not going to take them long to do it.”\(^{25}\)

One alarming study found that red-teaming efforts produced instructions from an LLM on how to build a dirty bomb. The author notes that the results of their initial efforts contain “information that is broadly available online … however, additional questions yielded more precise estimations and recommendations … A would-be terrorist might not know where to find detailed and accurate instructions for building weapons of mass destruction, but could potentially circumvent that crucial barrier by simply tricking a publicly available AI model.”\(^{26}\)

U.S. allies have also identified risks posed by advanced AI models. The U.K. Department for Science, Innovation & Technology released a report which found that “[f]rontier AI may help bad actors to perform cyberattacks, run disinformation campaigns and design biological or chemical weapons. Frontier AI will almost certainly continue to lower the barriers to entry for less sophisticated threat actors.”\(^{27}\)

President Biden’s Executive Order 14110, released this past October, echoed the concern over catastrophic risk through its focus on chemical, biological, radiological, nuclear (CBRN) risks and cyber risks. The E.O. requires the National Institute of Standards and Technology (NIST) to establish guidance for the evaluation of AI-enabled cyber and biological harms to assist in the development of safe and secure AI models. The Department of Energy must also develop tools to assess whether AI model outputs could lead to CBRN, cyber, and related security threats.\(^{28}\)
The E.O. also sets reporting requirements for advanced AI developers to inform the Department of Commerce on the development of the most advanced frontier models, initially defined as models trained on a quantity of computing power greater than $10^{26}$ operations. Entities that acquire, develop, or possess large-scale computing clusters are also subject to reporting requirements. Additionally, cloud service providers must report on training runs for the most advanced frontier models when they involve transactions with foreign persons.

Congress should consider a permanent framework to mitigate extreme risks. This framework should also serve as the basis for international coordination to mitigate extreme risks posed by AI. This letter is an attempt to start a dialogue about the need for such a framework, which would be in addition to, not at the exclusion of, proposals focused on other risks presented by developments in AI.

Under this potential framework, the most advanced model developers in the future would be required to safeguard against four extreme risks – the development of biological, chemical, cyber, or nuclear weapons. An agency or federal coordinating body would be tasked to oversee the implementation of these proposed requirements, which would apply to only the very largest and most advanced models. Such requirements would be reevaluated on a recurring basis as we gain a better understanding of the threat landscape and the technology.

The American private sector is the engine that makes our economy the envy of the world. Whatever Congress does to address the risks of AI, we must ensure that our domestic AI industry is able to develop and maintain an advantage over foreign adversaries. We also must ensure that any new requirements placed on industry do not bar new entrants, who will help drive innovation and discovery. We hope this letter generates engagement and feedback from experts, industry, policymakers, and other stakeholders in the weeks to come, which will be necessary for us to create a framework that can become law.

We look forward to working with you on these ideas and other matters related to AI this year.

Sincerely,

Mitt Romney  
United States Senator

Jack Reed  
United States Senator

Jerry Moran  
United States Senator

Angus S. King, Jr.  
United States Senator
Framework to Mitigate AI-Enabled Extreme Risks

The following proposal establishes a framework for federal oversight of frontier model hardware, development, and deployment to mitigate AI-enabled extreme risks, including biological, chemical, cyber, and nuclear threats.

Frontier Models:
Frontier models – those covered under this framework – would be only the most advanced AI models developed in the future – those that are both: (1) trained on an enormous amount of computing power – greater than $10^{26}$ operations, and that (2) are either broadly-capable, general purpose, and able to complete a variety of downstream tasks, or are intended to be used for bioengineering, chemical engineering, cybersecurity, or nuclear development. The $10^{26}$ operations compute threshold is the standard identified by Executive Order 14110, and it represents a metric which would be reevaluated on a regular basis to ensure it remains appropriate as technological advancements occur.

Oversight of Frontier Models:

I. Hardware

Training a frontier model would require tremendous computing resources. Entities that sell or rent the use of a large amount of computing hardware, potentially set at the level specified by E.O. 14110, for AI development would report large acquisitions or usage of such computing resources to the oversight entity and exercise due diligence to ensure that customers are known and vetted, particularly with respect to foreign persons.

II. Development of Frontier Models

Developers would notify the oversight entity when developing a frontier model and prior to initiating training runs. Developers would be required to incorporate safeguards against the four extreme risks identified above, and adhere to cybersecurity standards to ensure models are not leaked prematurely or stolen.

Frontier model developers could be required to report to the oversight entity on steps taken to mitigate the four identified risks and implement cybersecurity standards.

III. Deployment of Frontier Models

Frontier model developers would undergo evaluation and obtain a license from the oversight entity prior to release. This evaluation would only consider whether the frontier model has incorporated sufficient safeguards against the four identified risks.

A tiered licensing structure would be utilized to determine how widely the frontier model could be shared. For instance, frontier models with low risk could be licensed for open-source deployment, whereas models with higher risks could be licensed for deployment with vetted customers or limited public use.
Oversight Entity:

Congress could give these oversight authorities to a new interagency coordinating body, a preexisting federal agency, or a new agency. Four potential options for this oversight entity are:

A. **Interagency Coordinating Body.** A new, interagency body could be created to facilitate cross-agency regulatory oversight. This body could be modeled on the Committee on Foreign Investment in the United States (CFIUS). It would be organized in a way to leverage domain-specific subject matter expertise while ensuring coordination and communication among key federal stakeholders.

B. **Department of Commerce.** Commerce could leverage the National Institute for Standards and Technology (NIST) and the Bureau of Industry and Security to carry out these responsibilities.

C. **Department of Energy (DoE).** DoE has expertise in high-performance computing and oversees the U.S. National Laboratories. Additionally, DoE has deep experience in handling restricted data, classified information, and national security issues.

D. **New Agency.** Since frontier models pose novel risks that do not fit neatly within existing agency jurisdictions, Congress could task a new agency with these responsibilities.

Regardless of where these authorities reside, the oversight entity should be comprised of: (1) subject matter experts, who could be detailed from relevant federal entities that have experience handling issues such as biosecurity, chemical security, cybersecurity, and nuclear security, and (2) skilled AI scientists and engineers.

The oversight entity would also be tasked to study and report to Congress on unforeseen challenges and new risks to ensure that this framework remains appropriate as technology advances.


Homeland Threat Assessment 2024 at v.


