PUBLIC SUBMISSION

Received: May 29, 2025
Tracking No. mb9-qckl-5rzm
Comments Due: May 28, 2025 **Submission Type:** API

Docket: NSF-2025-OGC-0001
NITRD_FRDOC_0001

Comment On: NSF-2025-OGC-0001-0001
Request for Information: Development of a 2025 National Artificial Intelligence Research and Development Strategic Plan

Document: NSF-2025-OGC-0001-DRAFT-0227
Comment on FR Doc # 2025-07332

Submitter Information

Organization: The Council on Strategic Risks

General Comment

See attachment

Attachments

OSTP_NSF RFI_2025

May 29, 2025

TO: Networking and Information Technology Research and Development (NITRD) National Coordination Office (NCO), National Science Foundation (NSF)

FROM: The Council on Strategic Risk (CSR)

SUBJECT: **Request for Information on the Development of a 2025 National Artificial Intelligence (AI) Research and Development (R&D) Strategic Plan**

—

Below are responses from the Council on Strategic Risks (CSR), a nonprofit, nonpartisan think tank dedicated to analyzing and addressing global risks. Given our technical expertise in AI and biosecurity and extensive knowledge of US biodefense efforts from our staff's former government experience, CSR is uniquely suited to offer recommendations concerning biological threats; five recommendations are included below.

A rewrite of the National Artificial Intelligence Research and Development Strategic Plan (2023 Update) should include furthering research and development (R&D) to address challenges at the intersection of Artificial Intelligence (AI) applications in bioengineering and biosecurity—termed the AIxBio nexus. To maintain our global leadership in AI infrastructure, development, and innovation, the United States should conduct research into AI standards to promote national security, bolster critical biodefense infrastructure, and increase safety in AI applications without stifling innovation or scientific discovery.

AI exacerbates both risks and benefits of biotechnology:

Positive results of the integration of biological capabilities and AI models include multiple uses in biodefense, where they can aid in surveillance for disease spread, assess the pandemic potential of mutations and toxins, and aid in troubleshooting experimentation in the laboratory, and even leverage AI-enabled lab machinery to automate scientific discovery. AI research will accelerate fundamental scientific discovery in the chemical[1] and biological disciplines, facilitating technological breakthroughs; not only are AI models already aiding in the discovery and design of new

[1] AI applications in chemistry likewise show both promise and new challenges related to chemical weapons risks.

—

Council on Strategic Risks

1025 Connecticut Ave. NW, Ste 1000
Washington, DC 20036

biotech products and treatments,[2] but many of these drugs are already moving through the FDA's clinical trial pipeline to assess their safety and efficacy.[3] U.S. companies are also developing AI-powered tools to rapidly discover how existing medical treatments on the market could be used in new ways to combat disease threats.

At the same time, AI can also lower our adversaries' barriers to entry to manufacture bioweapons or create novel, engineered bioweapons to maximize their intended damage. Non-state actors, who are more likely to be non-subject matter experts, may find using a large language model (LLM) to conduct otherwise difficult-to-perform scientific or industrial processes provides better advice than an expert scientist.[4] One study also found AI can help bad actors swiftly identify DNA synthesis companies that are unlikely to screen customers and their DNA synthesis orders, which may be misused for bioweapons potential.[5] For state and non-state actors alike, AI models specially trained on biological data capable of designing novel biological sequences, or biological design tools (BDTs), can be used to design and create new bioweapons. Although the dual-use potential of AI to develop bioweapons has not yet manifested in known incidents, many AI and biosecurity researchers fear this will not remain the case.[6] Five practical steps the U.S. can take to address this risk while maintaining an innovative edge are outlined below:

1) The U.S. interagency should support the Artificial Intelligence Safety Institute's (AISI) work conducting evaluations of public-facing large language models' (LLMs) willingness and capabilities to act in ways that challenge biosecurity, and the subsequent effectiveness of safeguards, through public-private partnerships: The federal government should facilitate the study of LLMs currently widely in use to assess their willingness to aid users in creating bioweapons in exchange for an agreement to anonymize the companies in its study. The findings should be published so long as it is determined that this would not create intolerable information hazards, and the companies should be strongly encouraged to apply the publications' findings to their LLMs' benchmarks and perhaps even their model's feedback to users. For example, in their 2025 pre-print,[7] Noever and McKee study the willingness or refusal of several LLMs to create bioweapons under varying study conditions and find that, for some LLMs, their model's chain-of-thought output was revealing potential vulnerabilities in its model

[2] Anshul Kanakia, Mark Sale, Liang Zhao, and Zhu Zhou, "*AI In Action: Redefining Drug Discovery and Development.*" Clinical and Translational Science, February 2025;18(2):e70149.
[3] Chen Fu and Qiuchen Chen, "The future of pharmaceuticals: Artificial intelligence in drug discovery and development," Journal of Pharmaceutical Analysis, 2025, 101248, ISSN 2095-1779.
[4] Justen, Lennart, "*LLMs Outperform Experts on Challenging Biology Benchmarks*" arXiv, arXiv:2505.06108, May 12, 2025.
[5] Emily H. Soice, Rafael Rocha, Kimberlee Cordova, Michael Specter, and Kevin M. Esvelt, "*Can Large Language Models Democratize Access to Dual-Use Biotechnology?*" arXiv, arXiv:2306.03809, June 6, 2023.
[6] Gladstone AI, *"Survey of AI Technologies and AI R&D Trajectories,"* November 3, 2023.
[7] David Noever, and Forrest McKee, "*Forbidden Science: Dual-Use AI Challenge Benchmark and Scientific Refusal Tests*" arXiv, arXi2502.06867, February 8, 2025.

safeguards. The development of robust safety benchmarks at the AIxBio nexus must be performed iteratively and must be done in collaboration with the creators of the major LLMs in use worldwide to ensure models are: stable and robust to repeated inquiries, compliant with the company's or regulator's stances on not aiding in bioweapon production without hampering legitimate research, and ensure their model's user interface and choice of transparency (via chain-of-thought output) isn't being misused–a massive undertaking the U.S. government is uniquely fit for given its requirement to protect national security. Under NIST, the US AISI is currently engaging in this vital work, which would greatly benefit from continued OSTP support, especially at the AIxBio nexus.

2) The NSF should create research grants to develop methods that allow for "machine unlearning"[8] for biological design tools (BDTs), together with studies of real world data provenance to assess current compliance with regulatory standards: Data provenance refers to the historical record of data's source(s) and the changes that have been made to the data throughout its lifecycle. Compared to the opaque data provenance practices used in the development of foundation LLM models, in theory, there are standardized procedures dictating provenance norms for biological material and data.[9] Yet, compliance with these standards are not well understood and must be studied. Because data provenance of these models may be poorly understood, BDTs may have been trained on a variety of sensitive biological data that could be easily misused. In a 2022 proof-of-concept study,[10] researchers found they were able to use open source data and under 6 hours of machine time to generate 40,000 substances that could be used as chemical weapons, some of which were completely novel in structure and some that matched known, banned chemical weapons. Given the ease with which researchers were able to leverage AI in that study, and how far AI development has advanced since the time of the study, academic researchers developing BDTs and companies making BDTs available for use must conduct research into methods to ensure these tools are safe. This may include developing methods for "machine unlearning" for these specialized tools,[11] followed by assessments of the effectiveness of these methods. While "machine unlearning" may be promising, a 2023 survey[12] on its techniques found current unlearning methods mostly target traditional neural networks

[8] Jie Xu, Zihan Wu, Cong Wang, and Xiaohua Jia. "*Machine Unlearning: Solutions and Challenges*" arXiv, arxiv:2308.07061, August 14, 2023.
[9] International Organization for Standardization. (2023). "*Biotechnology — Provenance information model for biological material and data*" (ISO/TS 23494-1:2023).
[10] Fabio Urbina, Filippa Lentzos, Cédric Invernizzi, and Sean Ekins. "*Dual use of artificial-intelligence-powered drug discovery*." *Nature machine intelligence* 4, no. 3 (2022): 189-191.
[11] Jaspreet Pannu, Doni Bloomfield, Robert MacKnight, Moritz S. Hanke, Alex Zhu, Gabe Gomes, Anita Cicero, Thomas V. Inglesby, "*Dual-use capabilities of concern of biological AI models.*" PLoS Comput Biol. 2025 May 8;21(5):e1012975.
[12] Thanh Tam Nguyen, Thanh Trung Huynh, Zhao Ren, Phi Le Nguyen, Alan Wee-Chung Liew, Hongzhi Yin, Quoc Viet Hung Nguyen, "*A survey of machine unlearning*." *arXiv preprint arXiv:2209.02299* (2022).

—

(some but not all BDTs rely on this type of AI model). Though many are interested in and are actively researching "machine unlearning" methods, there is an absence of literature in applying machine unlearning specifically to BDTs. Through R&D funding to the national laboratories and/or academic research institutions with a proven track record of ML experience and the university resources to support this work, the NSF could address this gap in the literature and, as a result, inform policy best practices on methods to make existing and future BDTs safer.

3) NSF should invest in Technical Education, Training, and Human Capacity Building at the AIxBio nexus as well as strengthen dialogue between the intelligence and research communities: To maintain US global leadership in AI, NSF should educate both young scientists and industry partners about the potential risks and opportunities at the AIxBio nexus. Many training programs for young scientists are highly specialized, meaning AI researchers may not receive any education in molecular biology and bioengineering and the converse may be true. To do this, NSF should fund interdisciplinary education for researchers whose tools could be used to design bioweapons or aid in their synthesis. Similarly, AI companies, as a condition of receiving federal funding, must be educated in best practices if their product has the potential for AIxBio dual-use. Additionally, among scientists there exists a culture of openness, data sharing, even the sharing of work widely in online archives before it is published; in fact, collaboration is common among scientists of different nations. Thus, many scientists are not trained to consider the misuse of the data they generate. To address this, consideration of dual-use must be integrated into these' scientists' education, which is a challenge that the NSF can accomplish. Additionally, given the threat of cybertheft and the interest of hostile nations in American R&D at the AIxBio nexus, the intelligence community needs to create a means of dialogue with the AIxBio research community, including both AI companies and academics. By creating incentive structures for AI companies and academics' use, agencies such as the Federal Bureau of Investigation can continue their outreach to relevant entities and encourage researchers to inform them of any misuse indicators, theft, or attempted theft of American scientific work. As an example of alternative models, the US could look to the UK's Secure Innovation campaign.[13]

4) As BDTs can be used to evade traditional sequence-based biosecurity screening methods, research needs to be conducted to advance screening techniques and a public-private partnership established to share findings and best practices: In their 2024 paper,[14] Wittmann, et al. detail how AI enabled BDTs can be used to design

[13] https://www.npsa.gov.uk/specialised-guidance/secure-innovation

[14] Bruce J. Wittmann, Tessa Alexanian, Craig Bartling, Jacob Beal, Adam Clore, James Diggans, Kevin Flyangolts, Bryan T. Gemler, Tom Mitchell, Steven T. Murphy, Nicole E. Wheeler, and Eric Horvitz, "*Toward AI-Resilient Screening of Nucleic Acid Synthesis Orders: Process, Results, and Recommendations*." *bioRxiv, December 4,* 2024.

—

proteins homologous to “proteins of concern,” and that likely retain the same function but that have different nucleotide sequences. This study was the first large-scale, coordinated effort between nucleic acid synthesis companies and providers of biosecurity screening tools. It proved that BDTs can be used to easily bypass sequence-based screening methods. The authors note that AI red-teaming processes like the adversarial generation of novel “proteins of concern” candidates to stress test the screening software followed by iteratively updating the software with these candidates was successful in increasing its performance. Red-teaming related to bioweapons potential is inherently dangerous—and could risk being misconstrued as violating the Biological Weapons Convention under some circumstances—and would benefit from the involvement of the federal government to ensure appropriate security standards are maintained. Additionally, research into how novel biosecurity screening methods can be developed to leverage machine learning and known and newly uncovered homologies to improve detection of potential bioweapons must be conducted and the resulting knowledge and best practices shared openly between industry competitors. In addition to federal best practices in how BDTs can be leveraged to increase biosafety, the federal government should also guarantee the establishment of a Know Your Customer (KYC) database, whether operated directly by the federal government or statutorily-mandated but operated as a private entity. This kind of collaboration and knowledge sharing must be facilitated by the federal government to prevent security gaps and ensure that leading US companies aren’t disadvantaged by the costs of taking on this security measure independently. CSR has recommended that this administration or Congress establish an independent task force mandated to determine the best structure and other details for a KYC entity with a 9-month deadline to ensure this issue is addressed quickly.[15]

5) In preparation for a future in which Artificial General Intelligence (AGI) interfaces directly with BDTs and bioengineering machinery, the US needs to invest in research around AI model alignment and cybersecurity best practices for autonomous and cloud-based laboratories, aiming to ensure these labs are secure and compliant with synthesis regulations in place while preserving privacy and IP: As biomanufacturing leverages AI and robotics to increase in scale and capability, the potential methods for attack and the risk associated with those attacks also increases. This will necessitate “increasingly strong deployment and security protections.”[16] In an attempt to address increased risk in the context of bioweapons, the AI company Anthropic has developed a Responsible Scaling Policy (RSP) with a three-part approach: making their AI system more difficult to jailbreak; detecting jailbreaks when they do occur; and iteratively

[15] Christine Parthemore and Dan Regan, *“Biosecurity in the Next Administration,”* CSR Blog, January 31, 2025, https://councilonstrategicrisks.org/2025/01/31/biosecurity-in-the-next-administration/.
[16] Anthropic. *“Activating AI Safety Level 3 Protections,”* May 22, 2025, https://www.anthropic.com/news/activating-asl3-protections.

improving their defenses. The government should research and evaluate these approaches for their effectiveness in mitigating bio risks and consider adopting binding regulations to get other industry players to the same standard. Additionally, as more labs move to automate scientific experimentation, we must ensure methods are put in place to keep these autonomous tools secure, preventing an attack, and automatically monitoring what chemical or biological material is being produced in a privacy preserving manner. The federal government needs to monitor the use of autonomous labs and their cloud security practices, allowing for incident reporting and requiring these labs undergo regular adversarial testing. The digital-to-physical boundary currently acts as a safeguard against the undetected production of bioweapons[17]–but this may not remain the case as biomanufacturing moves away from having a persistent human-in-the-loop. Because AGI loss of control in biomanufacturing could pose an extreme risk, federally funded research into best practice protocols in the case of a loss of control needs to be conducted; additionally, the NSF should fund AI alignment research (research to ensure advanced AI systems behave as intended and remain controllable), and lines of communication ought to be established between autonomous labs and federal law enforcement to quickly alert federal authorities in the event of an attack.

In summary, taking advantage of the benefits of AI for biosecurity and mitigating the risks that co-evolve with these technologies will require public-private collaboration like the United States has never seen before. The United States has a unique opportunity to secure its position as the unrivaled world leader in AI by performing R&D, strengthening AI and cybersecurity standards, investing in AI & biosecurity education, and continuing to support current AI research. In the application of AI to biotechnology and biomanufacturing, the US can act now, before AI is used to produce a biological weapon. This investment now, in prevention and preparation, will be substantially less than the cost of responding to crop failures, pandemics, or even greater catastrophes caused by such misuse of these technologies.

About CSR

The Council on Strategic Risks (CSR) is a nonprofit, non-partisan security policy institute devoted to anticipating, analyzing, and addressing core systemic risks to security in the 21st century, with a special examination of how these risks intersect and exacerbate one another.

[17] https://helena.org/projects/helena-biosecurity

—

This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in developing the 2025 National AI R&D Strategic Plan and associated documents without attribution.