Homeland Security

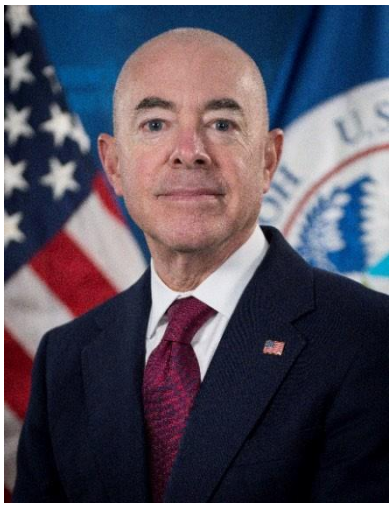
Department of Homeland Security Report on Reducing the Risks at the Intersection of Artificial Intelligence and Chemical, Biological, Radiological, and Nuclear Threats

April 26, 2024

This page is intentionally left blank.

Message from the Secretary

April 26, 2024



I am pleased to present the following report, "Department of Homeland Security Report on Reducing the Risks at the Intersection of Artificial Intelligence and Chemical, Biological, Radiological, and Nuclear Threats," which was prepared by the Department of Homeland Security (DHS) Countering Weapons of Mass Destruction Office (CWMD).

This report was compiled pursuant to Executive Order (E.O.) 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, dated October 30, 2023. Section 4.4 of the E.O. highlighted the need "to better understand and mitigate the risk of AI being misused to assist in development or use of CBRN threats – with a particular focus on biological weapons."

The E.O. tasked "the Secretary of Homeland Security, in consultation with the Secretary of Energy and Director of the Office of Science and Technology Policy," to "evaluate the potential for AI to be misused to enable the development or production of CBRN threats, while also considering the benefits and application of AI to counter these threats…" The E.O. also required "a report to the President that describes the progress of these efforts, including an assessment of the types of AI models that may present CBRN risks to the United States, and that makes recommendations for regulating or overseeing the training, deployment, publication, or use of these models, including requirements for safety evaluations and guardrails for mitigating potential threats to national security."

This report, which focuses on AI-enabled chemical and biological agents, was developed with inputs and recommendations "from experts in AI and CBRN issues for DHS, the Department of Energy, private AI laboratories, academia, and third-party model evaluators." The report is meant to provide longer-term objectives around how to ensure safe, secure, and trustworthy development and use of artificial intelligence, and guide potential interagency follow-on policy and implementation efforts.

Pursuant to the E.O., this report is being provided to the President of the United States.

Sincerely,

Alejandro N. Mayorkas

The Honorable Alejandro N. Mayorkas
Secretary
U.S. Department of Homeland Security

Table of Contents

1. Executive Summary.....1
2. Introduction 3
3. Background: Trends in AI.....4
3.1 General Trends in AI.....4
3.2 Trends in AI Governance and Oversight.....5
3.3 AI in the Physical and Life Sciences 6
4. AI Misuse to Enable the Development or Production of CBRN Threats.....8
5. Benefits and Application of AI To Counter CBRN Threats 16
6. Acronyms and Abbreviations 20

1. Executive Summary

On October 30, 2023, President Biden signed Executive Order (E.O.) 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. The overarching goal of the E.O. was "to ensure that America leads the way in seizing the promise and managing the risks of artificial intelligence (AI)" and to establish a governance framework for the safe and responsible development and use of AI.

The Department of Homeland Security (DHS) has played a key role in implementing the E.O.'s actions. Section 4.4(a) of the E.O. highlighted the need "to better understand and mitigate the risk of AI being misused to assist in development or use of CBRN threats – with a particular focus on biological weapons." Within DHS, the Countering Weapons of Mass Destruction Office (CWMD) is the office responsible for leading DHS efforts and coordinating with domestic and international partners to safeguard the United States against chemical, biological, radiological, and/or nuclear (CBRN) threats. CWMD led the development of an AI CBRN Report that evaluated "the potential for AI to be misused to enable the development or production of CBRN threats, while also considering the benefits and application of AI to counter these threats."

The AI CBRN Report was developed through strong collaboration across the United States Government (USG), academia, and industry. CWMD solicited insights from DHS Agencies and Offices and consulted with experts in AI and CBRN issues from the Department of Energy, private AI laboratories, academia, think thanks, and third-party model evaluators to evaluate AI model capabilities to present, mitigate, or guard against CBRN threats.

Current Trends in AI

- Responsible use of AI holds great promise for advancing science, analyzing large complex datasets beyond human cognitive abilities, solving urgent and future challenges, and improving daily life, while potential misuse poses consequential risk requiring society-wide mitigation efforts.
- AI has already affected the way research is conducted in the physical and life sciences and will continue to do so in expected and difficult-to-anticipate ways. These AI-enabled enhancements to research can have positive and negative impacts, depending on the intent of the users and the quality of the data.
- The revolutionary pace of change in the biotechnology, biomanufacturing, and AI sectors compounds existing regulatory challenges; therefore, AI technology governance must be adaptive and iterative to respond to rapid or unpredictable technological advancements.
- The variety of publicly available AI models can help enhance physical and life science researchers' ability to ideate novel biological and chemical agents and design experiments, increase their understanding of human physiology and the interaction with proteins and

toxins, and potentially troubleshoot experimental procedures encountered during experiments.

Findings

- **Finding 1:** Given the emerging nature of AI technologies, their interplay with chemical and biological research and development and the associated risks, an important USG priority should be to build consensus among the national security, public health, and animal health agencies about the range of potential risks associated with the use of AI.
- **Finding 2:** Most models and incorporated datasets are in the hands of private or academic organizations; significant momentum in open-source model development has democratized access to models and Biological Design Tools, including to malicious actors.
- **Finding 3:** As AI technologies advance, the lower barriers to entry for all actors across the sophistication spectrum may create novel risks to the homeland from malign actors' enhanced ability to conceptualize and conduct CBRN attacks.
- **Finding 4:** While each of the current frontier AI model developers have implemented a system of internal evaluation and red teaming per their participation in the Voluntary Commitments From Leading Artificial Intelligence Companies to Manage the Risks Posed by AI, their heterogenous approaches, the dual-use nature of the basic science information involved, and inconsistent access to relevant CBRN expertise make it vital to encourage continued interaction among industry, government, and academia and subsequently ensure ongoing exchanges between frontier model developers and the national security and broader biodefense communities.
- **Finding 5:** Known limitations in existing U.S. biological and chemical security regulations and enforcement, when combined with increased use of AI tools, could increase the likelihood of both intentional and unintentional dangerous research outcomes that pose a risk to public health, economic security, or national security.
- **Finding 6:** Engagement with international stakeholders including governments, international organizations, industry, and nongovernmental organizations is needed to develop approaches, principles, and frameworks to manage AI risks, unlock AI's potential for good, and promote common approaches to shared challenges in light of worldwide development and spread of AI technologies.
- **Finding 7:** Integration of AI into CBRN prevention, detection, response, and mitigation capabilities could yield important or emergent benefits.
- **Finding 8:** AI offers opportunities to leverage advanced analysis to bolster all lines of effort in the National Biodefense Strategy.
- **Finding 9:** AI tools could enhance international collaboration and communication on key efforts related to CBRN, attribution for suspected bioagent or chemical attacks and monitoring of non-state and nation states' compliance with international agreements and adherence to arms control, nonproliferation, and disarmament treaties.

2. Introduction

As artificial intelligence (AI) integrates into more areas of human activity, its potential benefits and risks have been subject to increased public and government scrutiny, highlighting the need for stronger governance over the development and use of AI and a clearer sense of the potential threats it could pose including those related to chemical, biological, radiological and/or nuclear (CBRN) threats. On October 30, 2023, President Biden issued an Executive Order (E.O.) intended to mitigate potential AI risks and threats (E.O.14110, *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*). The E.O. establishes a governance framework for the safe and responsible development and use of AI. In particular, the E.O. cites the need to "evaluate the potential for AI to be misused to enable the development or production of CBRN threats, while also considering the benefits and application of AI to counter these threats." Section 4.4(a)(i) of E.O. 14110 tasks the Secretary of Homeland Security to make an assessment of the types of AI models that may present CBRN risks to the United States, including recommendations for regulating or overseeing the training, deployment, publication, or use of these models, including requirements for safety evaluations and guardrails for mitigating potential CBRN threats. The Countering Weapons of Mass Destruction Office (CWMD) is the lead DHS office responsible for generating this research and corresponding report, which focuses on AI-enabled chemical and biological agents and toxins. To keep the document unclassified and consistent with the special emphasis on biological weapons under Section 4.4(a)(i) and the unique authorities of the Department of Energy, National Nuclear Security Administration for nuclear-related information under the Atomic Energy Act of 1954, this report focuses on biological and chemical threats and only addresses nuclear and radiological threats insofar as they share common features in ideation, production, and dissemination with chemical and biological agents and toxins. In response to the E.O., CWMD assessed the risks posed by AI to generate or worsen chemical and biological threats and how its potential misuse could be mitigated or prevented. This report presents the results of the assessment required by the E.O. It examines the expansion of AI and its relationship to existing and future developments in the physical and life sciences. This assessment also identifies the trends in AI and types of AI models that might present or intensify biological and chemical threats to the United States. It offers recommendations to mitigate potential threats to national security by overseeing the training, deployment, publication, and use of AI models and underlying data, including the role of safety evaluations and guardrails.

While the E.O. focuses on generative AI and foundation models, this report also addresses the impact on threats and threat mitigation relevant to a class of AI tools commonly referred to as Biological Design Tools (BDTs). BDTs are tools and methodologies that allow the design and further understanding of biological processes such as characterizing proteins or designing novel organisms or biological structures. The risks, benefits, and mitigation approaches vary between the general-purpose foundation models and these specialized tools. This delineation is described further below.

In accordance with the taskings under Section 4.4(a)(i) of the E.O., CWMD and other DHS Agencies and Offices have consulted with experts in AI and CBRN issues from the Department of Energy, private AI laboratories, academia, think tanks, and third-party model evaluators over the last few months to evaluate AI model capabilities to present, mitigate, or guard against CBRN threats. DHS CWMD's research activities and stakeholder engagement, together with its assessment of AI models and CBRN risks and benefits and recommendations contained in this report, fulfill the requirements in E.O. Section 4.4.(a)(i).

3. Background: Trends in AI

3.1 General Trends in AI

AI was first created in the 1940s.[1] Since the early 2010s, however, AI capabilities have advanced substantially and are rapidly evolving, enabled by the vast increase in available computational power and the corresponding growth in the number of individuals with the ability to harness this increased computational power through Application Programming Interfaces (APIs) and natural language processing. Now that processing power is more accessible, both individuals and organizations can use AI at multiple scales. Advancements in the field of AI have proliferated through the increasing amount of publicly available data, increases in computational power, and novel algorithms.

Responsible use of AI holds great promise for advancing science, analyzing large complex datasets beyond human cognitive abilities, solving urgent and future challenges, and improving daily life, while potential misuse poses consequential risk requiring society-wide mitigation efforts. AI is a powerful tool with the potential to help individuals in multiple areas, including their personal lives, employment, and health care. For example, AI has already been used to improve the speed and accuracy of medical diagnostics, to help predict the toxicity of drug candidates, and to increase crop yield through precision agriculture.[2] [3] [4] At the same time, AI could complicate and compound existing dangers across a wide range of sectors. A key challenge for AI governance is finding the right balance between containment of risk and fostering innovation.

---

[1] Muthukrishnan, Nikesh, Farhad Maleki, Katie Ovens, Caroline Reinhold, Behzad Forghani, and Reza Forghani. "Brief History of Artificial Intelligence." *Neuroimaging Clinics of North America* 30 (November 1, 2020): 393–99. https://doi.org/10.1016/j.nic.2020.07.004.

[2] Al-Antari, Mugahed A. "Artificial Intelligence for Medical Diagnostics—Existing and Future AI Technology." *Diagnostics (Basel)* 13, no. 4 (2023): 688. https://doi.org/10.3390/diagnostics13040688.

[3] Tran, Thi Tuyet Van, Agung Surya Wibowo, Hilal Tayara, and Kil To Chong. "Artificial Intelligence in Drug Toxicity Prediction: Recent Advances, Challenges, and Future Perspectives." *Journal of Chemical Information and Modeling* 63, no. 9 (May 8, 2023): 2628–43. https://doi.org/10.1021/acs.jcim.3c00200.

[4] Patrício, Diego Inácio, and Rafael Rieder. "Computer Vision and Artificial Intelligence in Precision Agriculture for Grain Crops: A Systematic Review." *Computers and Electronics in Agriculture* 153 (2018): 69–81. https://doi.org/10.1016/j.compag.2018.08.001.

The rapid speed at which AI capabilities are advancing, when combined with nascent scientific understanding of those technological developments, often leads to uncertainty about AI's specific capabilities and limitations.

Among AI's recent advances is generative AI, which is designed to create content and carry out a large number of tasks using Large Language Models (LLMs) or AI image generators. Models which have been trained on large sets of data have the capability to perform a large variety of tasks, and which can be further "fine-tuned" for specific tasks, are called foundation models. LLMs, a type of foundation model intended to respond to and simulate human language, have direct applications like chatbot services, writing and content generation, and code-assistance. In addition to foundation models, design tools can employ a variety of AI methods such as rule-based systems, machine learning, and deep learning in specific applications or disciplines. Additionally, AI developers are working on uses of LLMs as agents for other applications or interfaces—this could allow laypersons to use scientific tools or write code that previously required specialized programming or other skills.

Foundation models will continue to find applications in other fields. Their application to a wide variety of fields outside of their originally designed purpose—such as sequential task responses or image generation—is a function of their broad applicability and will likely continue to proliferate into other areas. A robust, international open-source community is training and sharing models; the current wave of innovation is spreading worldwide. The proliferation of AI tools and platforms has expanded; however, the decentralized nature of developers and producers of these models means regulation could be difficult.

3.2 Trends in AI Governance and Oversight

The U.S. Government (USG) currently does not have an overarching legal or regulatory framework to comprehensively regulate or oversee AI research and development, production, and use of resulting applications. The Biden–Harris Administration, evidenced by the issuance of E.O. 14110, has placed the highest urgency on governing the development and use of AI safely and responsibly and is, therefore, advancing a coordinated USG-wide approach to doing so. Technological innovation often outpaces regulation; AI exemplifies this phenomenon. Developing an overarching and comprehensive AI governance framework that will stand the test of time and the evolution of technologies will require coordinated USG efforts that include persistent engagement with industry, the AI-user community, and many other private actors and domestic and international regulatory bodies. It is important to note that future AI benefits will be affected by USG regulation and that balancing AI regulation while simultaneously encouraging beneficial technological advancement will be paramount.

Existing federal laws and legal frameworks for regulating commerce, such as intellectual property, export control, technology transfers, foreign investments in the United States and out-bound investments of AI-related technologies, commercial transactions, data privacy, and cybersecurity, as well as international law, may provide opportunities to adapt to

regulate or oversee the development and deployment of AI. Emerging technologies, such as AI and biotechnology, pose significant oversight challenges that must be addressed to ensure their safe and ethical use. The rapid pace of technological change poses significant challenges for existing legislative and regulatory entities not only to oversee the deployment or use of AI but also to harness its great potential. The globally distributed development of and access to AI capabilities restricts the regulatory reach of the United States despite the country being the locus of this wave of technology development. The emergence of powerful AI models around the world as well as open-source models highlights the importance of cultivating continued investment not only in domestic AI innovation but also in promoting best practices and guidance for AI safety and security abroad through international cooperation and norm setting.

3.3 AI in the Physical and Life Sciences

AI has already affected the way research is conducted in the physical and life sciences and will continue to do so in expected and difficult-to-anticipate ways. These AI-enabled enhancements to research can have positive and negative impacts, depending on the intent of the users and the quality of the data. AI has made significant contributions to the physical and life sciences, and has demonstrated its value to improving the speed, ease, and cost of conducting research in these fields. Design tools heavily dominate the applications of AI in the chemical and biological fields, although there have been strides to also make LLMs domain specific to these fields. For example, in February 2024 scientists launched efforts to build the first universal, specialized AI foundation model for biology, seeking to connect generative AI with the various layers of biology (i.e., molecules to cells, tissues, whole of organisms) for the purpose of accelerating biomedical and environmental science.[5] Research at the intersection of AI and chemistry and biology has been advancing rapidly over the previous decade, outpacing existing mechanisms for legislating or regulating emerging or rapidly evolving technologies. As AI models and data become more accessible and efficacious for beneficial scientific applications, so too could they become useful for actors with the intention of utilizing science and technology to harm society. Human oversight of AI technologies, especially when combined with the physical and life sciences, is necessary to address gaps, guide the knowledge base, and maintain risk-informed context.

The variety of publicly available AI models can help enhance physical and life science researchers' ability to ideate novel biological and chemical agents and design experiments, increase their understanding of human physiology and the interaction with proteins and toxins, and potentially troubleshoot experimental procedures encountered during experiments, providing enhanced research capabilities especially for small research teams or those without the necessary expertise to do so otherwise. The successful unraveling of the 50-year-old protein-folding problem and, very recently, AI-driven lab assistant, "Coscientist" that is shown to be capable of (semi-) autonomously designing, planning, and executing multistep

[5] "Bioptimus | We Build Foundation Models That Transform Biology," n.d. https://www.bioptimus.com/.

scientific experiments are notable examples of how AI is advancing the field of scientific discovery in this space. [6] Because this report is a snapshot in time for a rapidly advancing field, it is highly likely that more advanced models will be released after the publication of this report.

Data is essential for training, testing, and validating AI systems. Large foundation models that are trained on broad datasets have the potential to be "dual-use" models as they can be fine-tuned on publicly available datasets containing information about chemical and biological agents and toxins and which may be manipulated for the purposes of generating harmful information. Currently, publicly available chemical and biological AI models are plagued with high failure rates, confabulations, and questions about the integrity of open-source datasets. Many large datasets are already publicly available, but many of them are not standardized and poorly curated, contributing to these failure rates. Models trained on inaccurate or insufficient data sometimes produce unreliable outputs or lead to what is known as "hallucinations." This challenge, however, could be changing as developers acquire more and better data and continuously validate their models and if the scientific community acts to better curate these datasets with the potential benefits of AI in mind. This will likely lead to a significant improvement in the accuracy of AI models for chemical and biological applications within the next few years.

Roles and responsibilities for addressing biological and chemical threats and their consequences are spread across multiple federal agencies, creating information-sharing and regulatory challenges. As acknowledged in the National Biodefense Strategy, addressing the broad range of biological risks requires significant advances at the convergence of multiple disciplines, including life, physical, and computational sciences as well as by the agencies at all levels of government responsible for missions in these areas.[7] The National Biodefense Strategy's Implementation Plan demonstrates the complexity of roles and responsibilities for biodefense, where over 70 percent of all specific actions have multiple lead federal agencies responsible for execution. Each federal agency may have its own perspective on discussions about how to mitigate risks of biological and chemical agents, as well as the magnitude and prioritization of these risks in comparison to others in their respective mission areas. Additionally, different agencies' perspectives on risk overlayed with diverse authorities and varying involvement in a wide variety of information-sharing forums can lead to information-sharing and regulatory challenges. Comprehensively addressing the range of threats requires standards of relevance to ensure that appropriate support from relevant federal agencies is included in discussions about the impact of AI on chemical and biological threats.

The revolutionary pace of change in the biotechnology, biomanufacturing, and AI sectors compounds existing regulatory challenges; therefore, AI technology governance must be adaptive and iterative to respond to rapid or unpredictable technological advancements. As

---

[6] Boiko, Daniil A., Robert MacKnight, Ben Kline, and Gabe Gomes. "Autonomous Chemical Research with Large Language Models." *Nature (London)* 624, no. 7992 (2023): 570–78. https://doi.org/10.1038/s41586-023-06792-0.
[7] The White House. "National Biodefense Strategy and Implementation Plan: For Countering Biological Threats, Enhancing Pandemic Preparedness, and Achieving Global Health Security," October 2022. https://www.whitehouse.gov/wp-content/uploads/2022/10/National-Biodefense-Strategy-and-Implementation-Plan-Final.pdf.

stated in the National Biodefense Strategy, the ongoing revolution in the life sciences and biotechnology is expected to continue at an ever-increasing rate, offering solutions to many public health and scientific challenges. However, biotechnologies are expected to be no longer confined to sophisticated research laboratories and instead will be developed and used all over the world by a growing community of users, many of whom are not steeped in biosafety and biosecurity best practices.[8] For instance advances in nucleic acid synthesis technology, to include benchtop synthesizers, provide a means to bypass efforts to perform sequence screening at providers or third-party vendors. Current reliance on a list-based system of regulated pathogens and toxins (e.g., Federal Select Agent Program or the Bureau of Industry and Security Export Administration Regulations' Commercial Control List) fails to account for the risk posed by nucleic acid sequences of concern or the potential for novel types of nucleic acid sequences that may contribute to pathogenicity or harm to be created with the assistance of AI or BDTs.

The use of AI in the physical and life sciences and related security sectors could be addressed through application or modification of existing federal, state, local, tribal, and territorial (SLTT) laws and policy frameworks to regulate or oversee other risks in these areas. Examples of these laws and frameworks include policies directed at the design, synthesis or cultivation, handling, transporting, storage, and management of chemical and biological materials and the laboratory infrastructure, as well as research activities and conditions that affect the environment, health, and safety. Export control rules are another potential lever that could be used to address new risks brought on by AI in these sectors.

The U.S. physical and life sciences research enterprises are integrated with international research enterprises and there is a track record of reasonable domestic legal or policy frameworks positively impacting the uptake of similar reasonable measures in partner countries. The spread of AI-enabled physical and life sciences innovation worldwide further highlights the critical need to engage with the global community to align on appropriate measures to balance risk and reward and to ensure universal adoption of the most critical safety measures.

4. AI Misuse to Enable the Development or Production of CBRN Threats

The increased proliferation and capabilities of AI tools as highlighted in the preceding section may lead to significant changes in the landscape of threats to U.S. national security over time, including by influencing the means, accessibility, or likelihood of a successful CBRN attack. When addressing the potential of AI and CBRN risk, it is necessary to address the full range of actors, the ability of these actors to take action, as well as the implications of these actions if they are successfully executed. This section addresses the potential for AI to be misused for the design

[8] Ibid.

and development of CBRN threats by increasing adversary capability, with a focus on chemical and biological threats, including non-state actors (e.g., lone extremists and terrorist groups), as well as state actors (e.g., those known or suspected to have chemical or biological weapons programs or those that have had programs in the past or may be considering such programs in the future). This section also considers unintentional misuse resulting from unanticipated or otherwise adverse research outcomes.

Given the variety of actors under consideration, the next step is analyzing the range of actions in which AI could be applied in planning or conducting a CBRN attack. The threat "pathway" consists of several steps from ideation through physical conduct of the attack or, in the case of unintentional misuse, the resultant accidental release of a CBRN agent. Interim steps vary by the type and purpose of an attack but generally include planning, material acquisition, weaponization, and transportation in addition to conceptualization and attack conduct. Disrupting any one of these steps, a concept known as "pathway defeat," results in the attack being unsuccessful or critically delayed.[9] Together, these two concepts—addressing the full range of potential actors contributing to a threat and the several steps required for that threat to yield negative consequences—provide a framework for understanding all the different ways AI can be applied to the CBRN problem set and identifying opportunities for applying threat mitigation or containment guardrails.

Finding 1: Given the emerging nature of AI technologies, their interplay with chemical and biological research and development and the associated risks, an important USG priority should be to build consensus among the national security, public health, and animal health agencies about the range of potential risks associated with the use of AI. Since AI is such a rapidly changing field, many law enforcement, public health, or national security stakeholders have difficulty keeping up with developments and trends in the AI industry and how the emergence of new technologies affects their interests and operations. Raising the general awareness of these stakeholders and incorporating AI into regular processes for threat analysis, risk assessment, and information sharing could mitigate the risk of strategic surprise and identify additional areas where AI might be exacerbating threats to national security. Cooperation, particularly with close allies, will be critical in ensuring a coordinated response to AI technologies.

Finding 1 Recommendations:

- **Recommendation 1.a:** Incorporate AI-specific CBRN topics into regular actionable intelligence and threat information sharing, reporting, and engagements among federal agencies and with SLTT stakeholders, allies, and partners to remove, reduce, and mitigate threats and risk.

[9] The Joint. "Joint Publication 3-40: Joint Countering Weapons of Mass Destruction," July 14, 2021. https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_40.pdf.

- **Recommendation 1.b:** Incorporate AI-specific CBRN risks into national risk assessments such as those required by the National Biodefense Strategy Implementation Plan and National Security Memorandum (NSM) 15, NSM-16, NSM-19, and others to appropriately account for AI in capability and strategic planning. As AI technologies could be applied across multiple disciplines and mission areas, it is important to address how these technologies affect each of these potentially disparate areas and governmental functions.
- **Recommendation 1.c:** Conduct mapping of illustrative AI CBRN use cases to share among national security, public health, and animal health USG agencies to help scope threat and risk assessments.
- **Recommendation 1.d:** Develop programs and/or initiatives from a designated USG source to educate policymakers, scientists, and the public about the capabilities, limitations, and potential risks associated with the use of AI. As part of this effort, work toward common guidance among federal agencies on classification parameters related to the risks posed by CBRN threat design and development aided by AI.

Finding 2: Most models and incorporated datasets are in the hands of private or academic organizations; significant momentum in open-source model development has democratized access to models and BDTs, including to malicious actors. As access to data and AI tools becomes more commonplace, the potential for their use for malign purposes grows and proliferates. Insider threat remains a risk, as insiders with access to model weights or data could compromise their confidentiality and security. Given the importance of modern AI methods and the technical skillset needed to utilize them successfully, AI experts could also be targeted by adversaries. As LLMs continue to advance and general-purpose models are able to function as agents for more highly specialized tools such as BDTs, the technical barriers to accessing the most concerning AI applications will be further reduced.

Finding 2 Recommendations:

- **Recommendation 2.a:** Develop USG-recommended guidance to encourage the development of granular release practices for source code and AI model weights for biological and/or chemical specific foundation models and general-purpose biological or chemical design tools that could be used to develop chemical or biological weapons (CBW) or related items (e.g., dissemination methods for CBW). Regularly update this guidance based on the availability of new tools and the results of risk assessments.
- **Recommendation 2.b:** Develop "safe harbor" reporting processes for model and tool developers, private sector, academic institutions, and the general public to report potential vulnerabilities to a government agency (USG or international) and/or an objective third party through secure communication.

- **Recommendation 2.c:** Develop and implement USG-recommended criteria for tactical exclusion and/or protection of sensitive chemical and biological data—such as sequence information associated with pathogenicity or toxicity—from publicly accessible databases on which AI could train.
- **Recommendation 2.d:** Encourage incorporation of differentiated access and Know your Customer Systems for particularly high-risk specialized tools and services such as biological design and chemical retrosynthesis tools, and nucleic acid synthesis providers. These concepts could take advantage of well-established implementation models like the Cybersecurity and Infrastructure Security Agency's "Secure by Design"[10] and could be applied to U.S. Government grant awardees via funding agency terms.
- **Recommendation 2.e:** Collaboratively develop and encourage adoption of guardrails to protect against reverse engineering, loss, or leakage of sensitive AI model weights by both non-state and state actors. This could include such measures as cybersecurity and insider threat training, investments in insider threat programs, restricting access to model weights, hardening of interfaces, and conducting research and development to better secure systems from malign actors.

Finding 3: As AI technologies advance, the lower barriers to entry for all actors across the sophistication spectrum may create novel risks to the homeland from malign actors' enhanced ability to conceptualize and conduct CBRN attacks. Progression from virtual or *in silico* efforts to physical synthesis and successful use of weaponized chemical and biological materials currently does and in the near future will continue to require a certain level of expertise and infrastructure to overcome enduring weaponization challenges. The need to create chemical and biological materials in the physical world provides a key transition point. These transition points, or chokepoints, are where appropriate oversight can result in substantial risk mitigation or containment. LLMs have been shown to lower the educational and knowledge barriers for traditional biological agents and toxins by providing protocols and troubleshooting information at every step of the pathway, enabling non-experts to perform tasks with an enhanced degree of competency or to overcome areas of ignorance outside of a particular area of expertise. Developing enhanced or novel biological agents and toxins with the use of advanced design tools, however, will likely still necessitate subject matter expertise for most, if not all, stages of the pathway in the near and most likely medium term. As AI technologies enable new entrants into the CBRN space, lack of experience with safety and security protocols could raise the risk of even well-intentioned actors accidentally releasing chemical or biological agents or other adverse research outcomes.

[10] Cybersecurity and Infrastructure Security Agency. "Secure by Design | CISA," n.d. https://www.cisa.gov/securebydesign.

Finding 3 Recommendations:

- **Recommendation 3.a:** Invest in building a culture of responsibility among the broader physical and life sciences communities, both domestically and internationally, and update training standards to incorporate AI risks. Conduct outreach to private-sector providers and third-party vendors of chemical and biological materials and laboratory services to build a culture of accountability and responsible conduct to reduce the risk of unwittingly producing dangerous biological and chemical agents. Make publicly available a secure mechanism to mediate the relationship between laboratories and material suppliers in order to establish and support a broad culture of accountability.
- **Recommendation 3.b:** Develop model guidelines for and encourage adoption of mandatory screening requirements at appropriate institutes for controlled or potentially dangerous substances, such as nucleic acid and peptide synthesis screening in accordance with Section 4.4(b) of E.O. 14110.
- **Recommendation 3.c:** Develop guidelines for AI-enabled automated laboratory and pharmaceutical capabilities to safeguard the digital-to-physical frontier. This could include designation of a responsible official in each biological laboratory or institution to be accountable for maintaining human oversight and ownership of physical and life science research to control processes and mitigate the most severe risks.
- **Recommendation 3.d:** Work with the international community to promote, create, and extend responsible AI safety and security principles and actions to or with the universal goal of limiting the progression from virtual or *in silico* efforts to physical synthesis of dangerous, inappropriate, or inadvertent development and use of harmful chemical and biological materials.

Finding 4: While each of the current frontier AI model developers have implemented a system of internal evaluation and red teaming per their participation in the *Voluntary Commitments From Leading Artificial Intelligence Companies to Manage the Risks Posed by AI*, their heterogenous approaches, the dual-use nature of the basic science information involved, and inconsistent access to relevant CBRN expertise make it vital to encourage continued interaction among industry, government, and academia and subsequently ensure ongoing exchanges between frontier model developers and the national security and broader biodefense communities.[11] The frontier AI laboratories that signed on to the Voluntary Commitments promised to "commit to internal and external red-teaming of models or systems" in areas including CBRN and to advancing research in AI safety and interpretability specific to these

[11] The White House. "FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI." The White House, July 21, 2023. https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/.

areas, but the mechanism of red teaming was left to the developers to implement.[12] As research in this area continues to advance, the role of government as the source of specialized knowledge about chemical and biological weapons in particular could be used to make red teaming more effective. Risks from BDTs may enable well-informed scientists with access to biofoundries, especially when operations may be split across multiple labs or providers, to design and produce chemical or biological threats that specifically evade medical countermeasures, including detection technologies and therapeutic or prophylactic treatments.

Finding 4 Recommendations:

- **Recommendation 4.a:** Reinforce policies, norms, and codes of conduct based on the Voluntary Commitments to mitigate against accidental and deliberate misuse of AI and AI-enabled design and discovery of enhanced and novel/advanced CBW agents.
- **Recommendation 4.b:** Through public, private, and academic partnerships, develop recommendations for the creation, curation, and use of appropriate training datasets relevant to CBW and ways to evaluate and validate datasets already in use.
- **Recommendation 4.c:** Investigate, understand, develop, and implement CBRN threat awareness training to model evaluators or red teams with the provision that they would operate in positions of public trust (background checks at a minimum and possibly security clearances) to improve the security involved in model development. As part of this training, include information about the specific materials and dissemination methods of the highest risk.
- **Recommendation 4.d:** Develop a standard framework based on the Voluntary Commitments for pre-release evaluations and red teaming of AI models by third parties and post-release reporting of potential hazards for foundation models to accrue information regarding their capability to design or construct CBW.
- **Recommendation 4.e:** Develop and promulgate USG-sponsored evaluation benchmarks or standards for LLMs consisting of questions or lines of questioning and thresholds for unacceptably dangerous responses to improve the models.
- **Recommendation 4.f:** Develop and pilot a framework for implementation of USG-sponsored red teaming of model capabilities to protect against CBW threats involving CBRN experts from across the public health, food safety, animal health, and national security agencies to improve the value and security of the models.
- **Recommendation 4.g:** Conduct a comprehensive, cross-sectoral analysis of existing domestic legal regime governing U.S. intellectual property (such as patent and copyright), civil liability, tax, export control, government procurement, consumer protection, biosafety/biosecurity, data privacy and security, and national security law that may be applicable for regulating dual-use AI-enabled biotechnology, and consider options, such as promulgating new rules under existing statutory authority, enforcing

[12] Ibid.

existing regulations differently, or working with Congress to introduce legislation providing new authorities that address unregulated or under-regulated entities.

Finding 5: Known limitations in existing U.S. biological and chemical security regulations and enforcement, when combined with increased use of AI tools, could increase the likelihood of both intentional and unintentional dangerous research outcomes that pose a risk to public health, economic security, or national security. Biofoundries, which are laboratories that contain a wide array of molecular biology or chemistry equipment as well as advanced robotics, democratize access to physical laboratory spaces and enable semi- or fully automated experimental analysis. Some biofoundries are capable of being controlled remotely; these make up a subset of "cloud labs," often enabling greater flexibility and building toward a higher degree of scientific process automation. Cloud labs could provide fully automatic access to a repository of starting reagents and reliable sample preparation, synthesis, and characterization techniques for the researcher completely online and without the presence of being in the laboratory and thus potentially able to avoid identity verification. One recent paper showed success using AI to semi-autonomously plan, design, and execute complex reactions.[13] In this work, the different software modules allowed the AI model to search for publicly available information about chemical compounds, find and read technical manuals on how to control robotic lab equipment, write computer code to carry out experiments, and analyze the resulting data to determine what worked and what did not. These robots were controlled by computer code written by AI. Routinizing mundane tasks or using robotic cloud laboratories are only two examples of potential avenues for AI models as agents to overcome controls on biological and chemical materials by potentially allowing an adversary to break up requests or procurements into small pieces across multiple labs or providers to evade detection. These examples highlight the importance of continued institutional oversight of AI applications to biological and chemical technologies that could result in the development of potential dual-use computational models directly enabling the design of pathogen with enhanced pandemic potential or a novel biological agent or toxin.

Finding 5 Recommendations:

- **Recommendation 5.a:** Enhance awareness and oversight of biotechnologies or systems directly affected by advances in AI by themselves or in combination with other emerging technologies to reduce the potential threat (e.g., nucleic acid synthesis, cloud laboratories, and biofoundries).
- **Recommendation 5.b:** Evaluate existing biosecurity guidance documents, funding requirements, and regulations in the context of the capabilities of current and emerging AI technologies and recommend policy changes to enhance the biosecurity policies and

[13] Boiko, Daniil A., Robert MacKnight, Ben Kline, and Gabe Gomes. "Autonomous Chemical Research with Large Language Models." *Nature (London)* 624, no. 7992 (2023): 570–78. https://doi.org/10.1038/s41586-023-06792-0.

practices of laboratories operating in the United States and ensure continued human oversight of lab experiments, including cloud laboratories and biofoundries.

- **Recommendation 5.c:** Improve USG outreach and education on dual-use research concerns and information hazards as well as insider threat training guidelines for government, private-sector institutions, and academia in the physical and life sciences and widen dissemination of training to entities entering the sector through new AI capabilities.
- **Recommendation 5.d:** Enact policies to improve understanding of, and inventory safety management in, government, commercial, and academic high-containment laboratories to improve awareness of safety practices and lay the groundwork for risk assessments, information sharing, and threat awareness related to AI. This could serve as the first step in a potential framework that goes beyond facilities that receive federal funding.
- **Recommendation 5.e:** Engage with Congress to update, modernize, and re-authorize the DHS Chemical Facility Antiterrorism Standards program to improve security of dangerous chemicals and chemical weapon precursors.
- **Recommendation 5.f:** Socialize national and economic security with science, technology, engineering, and mathematics student and professional learning communities, including healthcare and public health, establishing channels for see-something, say-something activities related to the circumvention of biological and chemical AI safety and security controls. Promote consistent terminology across these sectors through initiatives like the National Institute of Standards and Technology (NIST) Bioeconomy Lexicon.[14]
- **Recommendation 5.g:** Engage the physical, computing, engineering, and biological science communities through conferences or summits to discuss scientific, ethical, and governance issues associated with the use of AI, continued human oversight of experiments in these fields, and relevant dual-use technologies in research and development and to develop voluntary commitments that address the potential benefits, risks, and oversight of this rapidly advancing technology. If possible, leverage or harmonize these engagements with other communities of practices that were previously established in developing these commitments.
- **Recommendation 5.h:** Work with the international community to promote responsible AI safety and security principles and actions related to mitigating both intentional and unintentional research outcomes that pose a risk to public health, economic security, or national security.

[14] "NIST Bioeconomy Lexicon." *NIST*, December 2, 2022. https://www.nist.gov/bioscience/nist-bioeconomy-lexicon.

Finding 6: **Engagement with international stakeholders including governments, international organizations, industry, and nongovernmental organizations is needed to develop approaches, principles, and frameworks to manage AI risks, unlock AI's potential for good, and promote common approaches to shared challenges in light of worldwide development and spread of AI technologies.** International cooperation and coordination are critical for monitoring biological and chemical agents, especially as technological advancement continues to spread worldwide in these fields. The Biological and Toxin Weapons Convention (BWC), Chemical Weapons Convention (CWC), the United Nations Security Council Resolution (UNSCR) 1540, and other international and multilateral governance frameworks relevant to the production, trade, and use of biological weapons are in place, but they do not necessarily account for the novel threats posed by AI.

Finding 6 Recommendations:

- **Recommendation 6.a:** Reaffirm USG adherence to and compliance with arms control, nonproliferation, and global commitments to include the BWC, CWC, and UNSCR 1540 and discuss the challenges and opportunities AI technologies present for these regimes.
- **Recommendation 6.b:** Develop guidance on the international sharing of U.S.-based AI technologies and datasets, based on the Voluntary Commitments. Consider implementations of U.S. intellectual property, tort, tax, patent, export control, procurement, and data privacy law to accommodate dual-use AI-enabled biotechnology.
- **Recommendation 6.c:** Develop, in coordination with close allies, standards, frameworks, and red teaming efforts that prioritize chemical and biological security in the context of AI.
- **Recommendation 6.d:** Increase information-sharing mechanisms with the international community to promote responsible AI safety and security principles and actions specific to the physical and life sciences and improve communication, coordination, and collaboration regarding research and techniques on effective AI model guardrails and other safety practices as well as strengthening chemical and biological security measures.

5. Benefits and Application of AI To Counter CBRN Threats

The U.S. Government and its allies and partners have already applied AI to counter CBRN threats by applying the tools to help identify, prevent, and mitigate the impact of these threats—these efforts have accelerated recently with the technological developments described in Section 2 above. The research community and international organizations, often under their own volition, have also been considering how to apply AI specifically to counter chemical and biological threats. Much of this research is nascent, presenting an opportunity for the U.S. Government and its allies and partners to contribute to developing AI tools to identify CBRN threats, monitor and collect

information on state and non-state actor activities related to these threats, guide efforts to interrupt illicit procurement networks related to proliferation, and develop and enhance defensive countermeasures, such as detectors, decontamination methods, and medical countermeasures.

More specifically, the United States and allied governments could take advantage of researchers' exploratory work using AI tools in the physical and life sciences and apply their lessons to government-led national security and public health missions in a range of areas. This includes biochemistry, pharmacology, toxicology, and biosecurity, and in responding to the coronavirus disease 2019 (COVID-19) pandemic and preventing future disease outbreaks. Similarly, researchers in other fields have started employing AI systems to scrutinize large datasets for anomalies and have developed techniques and applications like machine vision for automating routine, mundane tasks—all of which have clear possible applications for governments overwhelmed by large volumes of data collected from the public and stream of commerce. The U.S. Government is similarly inundated with large quantities of data. Use of AI systems provides a means for the U.S. Government to navigate and utilize large datasets more quickly and efficiently. Possible areas include utilization of AI tools for nucleic acid synthesis screening; reviewing proposals for compliance with relevant life science research oversight policies; targeting or screening incoming cargo vessels or personal vehicles for contraband; reviewing visa, export, and import applications; and use of AI tools to evaluate models for risk of AI-enabled CBRN threat development.

Finding 7: Integration of AI into CBRN prevention, detection, response, and mitigation capabilities could yield important or emergent benefits. Government agencies have already started applying AI to certain fields like cargo and passenger screening, but other fertile areas for further research include helping develop and enhance chemical and biological defensive countermeasures, including development, acquisition, stockpiling, and dispersal of personal protective equipment, medical countermeasures (MCM), decontaminants, and detectors. In many instances, broader uses of AI such as for novel drug design can be applied to specific government use cases like MCMs for bioterrorism agents. In each of these areas, developing an appropriate adoption strategy for next-generation AI-enabled technologies and interagency collaboration and lessons learned will be vital to implementation success.

Finding 7 Recommendations:

- **Recommendation 7.a:** Integrate AI into program planning across the full range of prevention, detection, and response capabilities for countering weapons of mass destruction terrorism and CBRN preparedness as documented in the National Biodefense Strategy and Implementation Plan, NSM-19 Annex A, and other relevant documents.
- **Recommendation 7.b:** Encourage federal agencies and commercial providers to optimize the responsible use of AI in the design, testing, and evaluation of personal protective equipment, MCMs (e.g., vaccines and synthetic antibodies), and

decontaminants for treating CBW exposure in coordination with the activities in the AI E.O. completed under Section 8(b) of Executive Order 14110.

- **Recommendation 7.c:** Leverage the Department of Homeland Security's Artificial Intelligence Safety and Security Board and other mechanisms to promote information sharing and establish best practices and risk mitigation for AI technology development for CBRN-specific dual-use developments that could pose a national or economic security risk.

Finding 8: AI offers opportunities to leverage advanced analysis to bolster all lines of effort in the National Biodefense Strategy. This includes scrutinizing large, diverse datasets in numerous languages to detect the next disease outbreak, unlocking new understanding of genomics and proteomics that could be applied to truly "agent agnostic" diagnostics and early warning capabilities, and many other applications the national security and public health communities have not identified. Research and development is needed to keep pace with the transformative potential of AI and leverage it to for biodefense efforts. Applying strategies based on the newest progress in AI to the lines of effort of the National Biodefense Strategy, itself not even two years old, could quickly yield progress to its ambitious goals. It is important to note that AI benefits will be impacted by future USG efforts at regulation and that balancing regulation while simultaneously encouraging beneficial technological advancement will be paramount.

Finding 8 Recommendations:

- **Recommendation 8.a:** Research and develop AI-enabled systems that can detect and identify disease outbreaks and anomalous chemical and biological incidents for analyst review in a timely manner, to characterize, model, and contextualize the risk from such incidents, regardless of whether the incidents are intentional, natural, or accidental.
- **Recommendation 8.b:** Develop and apply AI techniques to monitor, trace, pre-stage, and deploy resources for responses to disease outbreaks and other human, animal, plant, and environmental health crises.
- **Recommendation 8.c:** Develop and apply AI systems and techniques to support CBRN risk assessments, threat modeling, and decision support to optimize detection and cleanup efforts.

Finding 9: AI tools could enhance international collaboration and communication on key efforts related to CBRN, attribution for suspected bioagent or chemical attacks, and monitoring of non-state and nation states' compliance with international agreements and adherence to arms control, nonproliferation and disarmament treaties. There are numerous potential applications of AI that could benefit international non-proliferation and counter-chemical and biological terrorism efforts, including more timely identification of chemical and biological threats from state and non-state actors, which could assist efforts to detect, interrupt,

and respond to such threats. Information gathered from AI-enabled capabilities could support international efforts to prevent, respond to, and hold actors responsible for the development and use of chemical and biological weapons. Furthermore, AI capabilities could be applied to identifying and interrupting illicit chemical- and biological-related procurement networks to help identify additional synthesis pathways for novel chemical and biological agents and toxins and resulting targets for sanction and export control actions. AI approaches could also be applied to areas with mass volumes of data to help identify illicit procurement networks and their vulnerabilities to interdiction.

Finding 9 Recommendations:

- **Recommendation 9.a:** Support the U.N. and Organisation for the Prohibition of Chemical Weapons efforts to implement AI into verification and inspection efforts.
- **Recommendation 9.b:** Develop and implement AI tools to identify signatures of CBW samples that could help attribute them to their origins.
- **Recommendation 9.c:** Invest in research efforts to enable biological and chemical forensic capabilities for unknown or novel biological and chemical constructs and work toward a set of international forensics standards using AI that can hold up in a court of law.
- **Recommendation 9.d:** Research and procure AI applications to analyze imagery and other digital phenomena to identify signposts and indicators of developing or enduring CBW programs.
- **Recommendation 9.e:** Research and procure applications of AI to analyze CBW-related procurement networks and identify vulnerabilities.
- **Recommendation 9.f:** Develop and implement AI models and tools to map and analyze terrorist networks to predict and interrupt CBW-related activities and attacks.
- **Recommendation 9.g:** Develop and implement AI tools to screen commercial orders of precursors, material, and equipment related to CBW development and use, including nucleic acid synthesis screening tools.

6. Acronyms and Abbreviations

AI	Artificial Intelligence
API	Application Programming Interfaces
BDT	Biological Design Tool
BWC	Biological and Toxin Weapons Convention
CBRN	Chemical, Biological, Radiological, and/or Nuclear
CBW	Chemical or Biological Weapons
CWC	Chemical Weapons Convention
CWMD	Countering Weapons of Mass Destruction Office
DHS	Department of Homeland Security
E.O.	Executive Order
LLM	Large Language Models
MCM	Medical Countermeasures
NIST	National Institute of Standards and Technology
NSM	National Security Memorandum
SLTT	State, Local, Tribal, and Territorial
UNSCR	United Nations Security Council Resolution
U.S.	United States
USG	United States Government