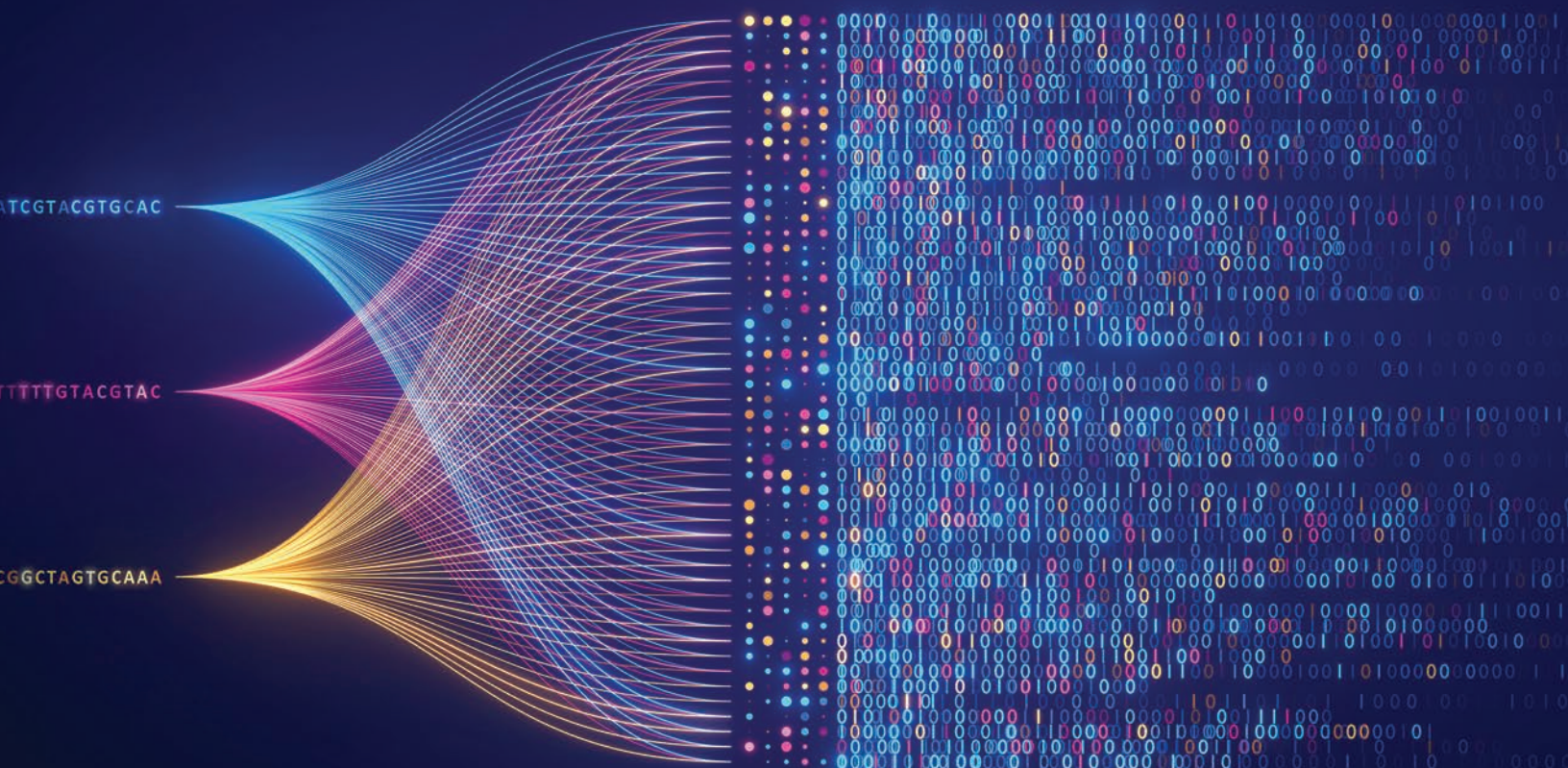


The Convergence of Artificial Intelligence and the Life Sciences:

Safeguarding Technology, Rethinking Governance,
and Preventing Catastrophe



Sarah R. Carter, Ph.D.
Nicole E. Wheeler, Ph.D.
Sabrina Chwalek
Christopher R. Isaac, M.Sc.
Jaime Yassif, Ph.D.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of those who were instrumental in the development of this report, including the many expert interviewees who generously shared their time and expertise. We would like to thank James Diggans at Twist Bioscience, Justin Farlow at Serotiny, Ryan Ritterson at Gryphon Scientific, and colleagues at Google DeepMind for providing particularly thoughtful comments in reviewing this report. Discussions with NTI Co-Chair and CEO Ernest Moniz and NTI President and COO Joan Rohlfing provided valuable insights in shaping the report recommendations. We would also like to thank Rachel Staley Grant for managing the production of this report and Hayley Severance for her assistance with managing this project. We are also deeply grateful to Fidelity Charitable for supporting this work.

Sarah R. Carter, Ph.D.

Principal, Science Policy Consulting

Nicole E. Wheeler, Ph.D.

Turing Fellow, The University of Birmingham

Sabrina Chwalek

Technical Consultant, Global Biological Policy and Programs, NTI

Christopher R. Isaac, M.Sc.

Program Officer, Global Biological Policy and Programs, NTI

Jaime Yassif, Ph.D.

Vice President, Global Biological Policy and Programs, NTI

© 2023 Nuclear Threat Initiative



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

The views expressed in this publication do not necessarily reflect those of the NTI Board of Directors or the institutions with which they are associated.

About the Nuclear Threat Initiative

The Nuclear Threat Initiative is a nonprofit, nonpartisan global security organization focused on reducing nuclear and biological threats imperiling humanity.

Contents

- Executive Summary 3
 - Current and Anticipated Capabilities 4
 - Biosecurity Implications 5
 - Opportunities for Risk Reduction 5
 - Recommendations 7
- Introduction 9
 - Context and Methodology 11
- AI and Biology: Current and Anticipated Capabilities 12
 - Large Language Models 15
 - Biodesign Tools 16
 - Automated Science 20
- Biosecurity Implications 23
 - Large Language Models 24
 - AI Biodesign Tools 27
 - Automated Science 29
 - Biosecurity Benefits of AI-Bio Capabilities 29
- Risk Reduction Opportunities 31
 - Guardrails for AI Models 32
 - Bolstering Biosecurity at the Digital-Physical Interface 40
 - Advancing Pandemic Preparedness 42
 - Roles and Responsibilities 42
- Recommendations: A Proposed Path Forward for Governance of AI-Bio Capabilities 46
- Conclusion 54
- Appendix A: Participants 55
- Appendix B: Examples of AI Models 56
- About the Authors 59
- Endnotes 61

Executive Summary

Rapid scientific and technological advances are fueling a 21st-century biotechnology revolution. Accelerating developments in the life sciences and in technologies such as artificial intelligence (AI), automation, and robotics are enhancing scientists' abilities to engineer living systems for a broad range of purposes. These groundbreaking advances are critical to building a more productive, sustainable, and healthy future for humans, animals, and the environment.

Significant advances in AI in recent years offer tremendous benefits for modern bioscience and bioengineering by supporting the rapid development of vaccines and therapeutics, enabling the development of new materials, fostering economic development, and helping fight climate change. However, AI-bio capabilities—AI tools and technologies that enable the engineering of living systems—also could be accidentally or deliberately misused to cause significant harm, with the potential to cause a global biological catastrophe.

These tools could expand access to knowledge and capabilities for producing well-known toxins, pathogens, or other biological agents. Soon, some AI-bio capabilities also could be exploited by malicious actors to develop agents that are new or more harmful than those that may evolve naturally. Given the rapid development and proliferation of these capabilities, leaders in government, bioscience research, industry, and the biosecurity community must work quickly to anticipate emerging risks on the horizon and proactively address them by developing strategies to protect against misuse.

To address the pressing need to govern AI-bio capabilities, this report explores three key questions:

- What are current and anticipated AI capabilities for engineering living systems?

- What are the biosecurity implications of these developments?
- What are the most promising options for governing this important technology that will effectively guard against misuse while enabling beneficial applications?

To answer these questions, this report presents key findings informed by interviews with more than 30 individuals with expertise in AI, biosecurity, bioscience research, biotechnology, and governance of emerging technologies. Building on these findings, the report includes recommendations from the authors on the path toward developing more robust governance approaches for AI-bio capabilities to reduce biological risks without unduly hindering scientific advances.

Current and Anticipated Capabilities

The intersection of AI with biology includes a wide variety of tools developed for many purposes, including large language models (LLMs), biodesign tools, and AI-enabled automation of the life sciences (for definitions, see box 1 in the main text). These AI-bio capabilities are likely to accelerate advances in the life sciences in a wide range of ways, from facilitating scientific training to helping scientists design new biological systems. Rapid progress in AI models is already lowering barriers to engineering biology, but tremendous uncertainty remains about the future capabilities of these tools, the pace of their development, and when breakthroughs will occur.

LLMs trained on human, or “natural” language, and their applications—such as OpenAI’s ChatGPT, Meta’s LLaMA Chat, Anthropic’s Claude, and Google’s Bard—are receiving significant attention for their ability to synthesize information and generate novel text in response to user prompts. LLMs can also process other types of data, such as audio, visual, and biological data, and efforts to create models that incorporate multiple types of data are underway. Although most natural language LLMs are not specifically designed to improve understanding of biological systems, they *de facto* serve this function by effectively summarizing a wide range of publicly available information about the life sciences, bioengineering, and laboratory tools and methods. These tools are designed to be easy to use and are likely to facilitate some types of

bioengineering by providing information, promising approaches, training, and guidance, including to users who have little scientific expertise. However, because LLMs draw on information that is widely available, they are likely to be most helpful and accurate for methods that have been well described and are similar to those that have been used previously. Additionally, LLMs may “hallucinate” false information in a convincing way, making it difficult for those with little expertise on a topic to tell fact from fiction.

Biodesign tools are trained on biological data, such as DNA or protein sequences, and are generally used by specialists to design biological molecules or systems. Protein design tools are the most mature biodesign tools, but other types are under development, including those that could be used to design more complex biological systems, such as whole genomes or organisms. Key limiting factors to developing these models include the complexity of biological systems and the paucity of information linking biological sequences with biological functions. In the near term, using these models will likely require some scientific expertise, and any designs generated will require experimental validation.

AI-enabled automated science is the delegation of one or more steps in the scientific process to AI. This could include surveying academic literature on a topic, developing testable hypotheses, designing

AI-bio capabilities are likely to accelerate advances in the life sciences in a wide range of ways, from facilitating scientific training to helping scientists design new biological systems.

AI model developers will need to work collaboratively with biosecurity experts to understand the biosecurity risks posed by their models, develop best practices, and refine and update approaches.

requires the development of novel solutions and represents a significant and urgent challenge.

Many developers of natural language LLMs already are implementing methods to safeguard their models against misuse. Current technical safeguards include training AI models to refuse to engage on particular topics and employing other methods to prevent them from outputting potentially harmful information. To assess the robustness of these methods, it is essential to evaluate models, for example with “red-teaming” exercises to determine their potential for misuse. The success of these technical safeguards also requires that AI model developers control access to their models. This can be challenging, particularly because some smaller AI models, including many AI biodesign tools, are developed as open-source resources. Other potential guardrails for AI models include controlling access to the computational infrastructure needed to train powerful models or to potentially harmful data, but there are open questions about the effectiveness of these approaches that will be important to resolve. To further develop guardrails, AI model developers will need to work collaboratively with biosecurity experts to understand the biosecurity risks posed by their models, develop best practices, and refine and update approaches.

In addition to developing AI model guardrails, there are opportunities to improve biosecurity oversight

at the interface where digital biological designs become physical reality. For example, many providers of synthetic DNA conduct biosecurity screening to ensure that pathogen or toxin DNA is not sold to customers who lack a legitimate use for it. These practices are currently largely voluntary, but governments could put in place more effective incentives or legal requirements. Improved screening tools would allow these providers to keep pace with the increasing number of novel designs generated by AI biodesign tools by screening DNA sequences on the basis of their potential encoded functions rather than just their similarity to known sequences. Other types of life science vendors and organizations also could bolster biosecurity by screening for customer legitimacy. These vendors and organizations include contract research organizations, academic core facilities, and providers of cloud laboratory services, robotics, and other life sciences products and services.

While more effective guardrails can offer significant risk reduction benefits, it is unlikely that they will eliminate all biosecurity risks that may arise at the intersection of AI and the life sciences. Therefore, resilient public health systems and strong pandemic preparedness and response capabilities will remain key safeguards; these capabilities can be substantially improved through AI-enabled advances.

Recommendations

Establish an international “AI-Bio Forum” to develop AI model guardrails that reduce biological risks

The Forum should be composed of key stakeholders and experts, including AI model developers in industry and academia and biosecurity experts within government and civil society. It should serve as a venue for developing and sharing best practices for implementing effective AI-bio guardrails, identifying emerging biological risks associated with ongoing AI advances, and developing shared resources to manage these risks. It should inform efforts by AI model developers in industry and academia, governments, and the broader biosecurity community, and it should establish global norms for biosecurity best practices in these communities.

Develop a radically new, more agile approach to national governance of AI-bio capabilities

To address emerging risks associated with rapidly advancing AI-bio capabilities, which can be difficult to anticipate, national governments should establish agile and adaptive governance approaches that can monitor AI technology developments and associated biological risks, incorporate private sector input, and rapidly adjust policy. Government policymakers should explore innovative approaches, such as dramatically streamlining rule-making procedures; rapidly exchanging information or co-developing policy with non-governmental AI experts; or explicitly empowering agile, non-governmental bodies that are working to develop and implement AI guardrails and other biological risk reduction measures.

Implement promising AI model guardrails at scale

AI model developers should implement the most promising already developed guardrails that reduce biological risks without unduly limiting beneficial uses. They should collaborate with other entities, including the AI-Bio Forum described above, to establish best practices and develop resources to support broader implementation. Governments, biosecurity organizations, and others should explore opportunities to scale up these solutions nationally and internationally, through funding, regulations, and other incentives for adoption. Existing guardrails that should be broadly implemented include AI model evaluations, methods for users to proactively report hazards, technical safeguards to limit harmful outputs, and access controls for AI models.

Pursue an ambitious research agenda to explore additional AI guardrail options for which open questions remain

AI model developers should work with biosecurity experts in government and civil society to explore additional options for AI model guardrails on an ongoing basis, experimenting with new approaches, and working to address key open questions and potential barriers to implementation. Priority areas for exploration include controlling access to AI biodesign tools, managing access to computational resources needed to train models, and managing access to data.

Strengthen biosecurity controls at the interface between digital design tools and physical biological systems

- Tool developers in industry, academia, and non-governmental organizations should develop new AI tools to strengthen DNA sequence screening approaches to capture novel threats and improve the robustness of current approaches.
- Governments, international bodies, and other key players should work to strengthen DNA synthesis screening frameworks, including by legally requiring screening practices.
- Governments and others should expand available tools, requirements, and incentives for customer screening to a wide range of providers of life science products, infrastructure, and services.

Use AI tools to build next-generation pandemic preparedness and response capabilities

Governments, development banks, and other funders should dramatically increase investment in pandemic preparedness and response, including by supporting development of next-generation AI tools for early detection and rapid response.

The convergence of AI and the life sciences marks a new era for biosecurity and offers tremendous potential benefits, including for pandemic preparedness and response. Yet, these rapidly developing capabilities also shift the biological risk landscape in ways that are difficult to predict and have the potential to cause a global biological catastrophe. The recommendations in this report provide a proposed path forward for taking action to address biological risks associated with rapid advances in AI-bio capabilities. Effectively implementing them will require creativity, agility, and sustained cycles of experimentation, learning, and refinement.

The world faces significant uncertainty about the future of AI and the life sciences, but it is clear that addressing these risks requires urgent action, unprecedented collaboration, a layered defense, and international engagement. Taking a proactive approach will help policymakers and others anticipate future technological advances on the horizon, address risks before they fully materialize, and ultimately foster a safer and more secure future.



Introduction

Modern bioscience and biotechnology are critical to building a more productive, sustainable, and healthy future for people, animals, and the environment. Rapid advances in these fields will have transformative effects on manufacturing, agriculture, energy production, and medicine. Recent progress in artificial intelligence (AI) technologies is steadily converging with the life sciences, building on decades of research and data collection, and will further accelerate these developments. The convergence of AI with biology will undoubtedly offer significant benefits, but it also poses new and poorly understood risks. This report describes this intersection, including AI tools and capabilities that enable engineering of living systems, the biosecurity implications of these developments, and opportunities to reduce risks.

While AI-bio capabilities can offer important benefits, biosecurity experts warn that they could also cause harm through accidental or intentional misuse. Malicious actors could exploit these tools to develop novel or more harmful toxins, increasingly dangerous pathogens, or other engineered biological agents. Given the rapid development and proliferation of AI-bio capabilities, it is critical to quickly identify potential risks and begin to implement strategies to protect against their misuse.

AI is intersecting with biology in a wide variety of contexts, with tools developed for a broad range of purposes (see box 1). LLMs developed by OpenAI, Meta, and Google, are not specifically

designed to improve our understanding of biological systems, but they have important intersections with the biosciences. AI biodesign tools are trained on biological data, such as DNA and protein sequences, and are often used by specialists working to design biological systems. Scientists use these tools for a wide range of practical purposes, such as for designing vaccines and understanding the mechanisms of disease transmission.¹ Automated science is incorporating AI into many steps in the scientific process, from the generation of hypotheses to the improvement of robotic experimentation, to data analytics.² These growing capabilities have the potential to enable the testing of more hypotheses and accelerate the pace of scientific discovery.

BOX 1. AI-BIO CAPABILITIES

This report uses multiple terms to describe AI tools that intersect with the life sciences.

AI-bio capabilities refers to the full range AI tools, models, and technologies that contribute to advances in the life sciences and bioengineering.

In this report, **LLM** refers to large language models trained on natural language (i.e., human language) as well as the associated applications built on top of them, such as chatbots that respond to text-based queries. LLMs can also be used to model other types of data, such as images, audio, and biological sequences, but unless otherwise specified, this report focuses on natural language LLMs. Other reports or analyses may refer to these models as “foundation” models if they are trained on large amounts of data and can be repurposed for more specific tasks, or as “frontier” models if they are close to the leading edge of AI capabilities.

Biodesign tool refers to any AI model that is used to design biological parts, systems, or organisms according to desired characteristics defined by the user. Some AI biodesign tools are LLMs that are trained on biological sequences rather than on natural languages, and this report considers them as biodesign tools. Some analysis of biodesign tools also draws on AI models, such as AlphaFold2, that are trained on biological data and provide insight into biology but do not provide biological designs.

Automated science refers to a range of AI tools and capabilities that can automate one or more steps in the scientific discovery process.

Each of these types of tools change the risk landscape in unique ways. For example, LLMs may be helpful to users with less scientific expertise who seek to learn more about pathogens, pathogen engineering, or laboratory techniques. Effectively using AI biodesign tools requires more expertise but could generate a wide variety of designs for toxins or, further into the future, pathogens or other biological agents with desired characteristics. AI automation may enable larger-scale testing of biological designs, allowing better optimization of desired characteristics. It is likely that different types of AI-bio capabilities will increasingly be combined in the future. For example, future AI tools could use LLMs to interpret a user’s text-based prompts, and use a biodesign tool to generate a design that satisfies the user’s request,³ and AI-enabled automated science systems could help experimentally evaluate AI-generated biological designs.

Experts remain uncertain about how LLMs, AI biodesign tools, and AI-bio capabilities for automating science will change in the near future, when developments or breakthroughs will occur, and how new biosecurity risks will materialize. This report aims to provide as much clarity as possible about anticipated risks and opportunities posed by AI and to provide recommendations on the path forward. It is imperative that AI model developers, policymakers, and biosecurity experts acknowledge and plan for unanticipated capabilities and risks that will emerge as AI continues to intersect with biology in new ways.

Context and Methodology

The Nuclear Threat Initiative (NTI) is a nonprofit, nonpartisan global security organization focused on reducing nuclear and biological threats imperiling humanity. Within NTI, the Global Biological Policy and Programs team (NTI | bio) works with governments, industry, academia, international organizations, and civil society to prevent catastrophic biological events, including through its work to strengthen biotechnology governance. NTI | bio is advancing this work through the Biosecurity Innovation and Risk Reduction Initiative,⁴ which focuses on addressing emerging biological risks associated with rapid technological advances. Under this initiative, NTI | bio has worked to bolster safeguards for DNA synthesis technologies⁵ and to strengthen biosecurity governance worldwide, specifically through the establishment of the International Biosecurity and Biosafety Initiative for Science.⁶

This report stems from the recognition that developing effective guardrails will be a critical element of broader efforts to safeguard the tools of modern bioscience and biotechnology against accidental or deliberate misuse. It draws on structured interviews with more than 30 experts in AI, biosecurity, bioscience research, synthetic biology and biotechnology, and governance of emerging technologies. The authors also convened a virtual workshop in August 2023 with interviewees and additional experts to discuss preliminary findings and recommendations that emerged from the interview process (for the list of participants, see appendix A). The first three sections of this report draw heavily on the expert opinions and perspectives that were gathered over the course of this project, though no attempt was made to generate consensus among this group. The final section of the report includes recommendations that build on the key findings but were developed by the authors alone and do not necessarily reflect the views of these experts.



AI and Biology: Current and Anticipated Capabilities

To better understand AI-bio capabilities and their implications for the life sciences in the near future, we addressed the following key questions:



What types of capabilities do the key relevant technologies currently have, and what are their main limitations? How will these capabilities evolve over the next few years?



What important unsolved problems in the life sciences and engineering living systems might AI and/or machine learning tools solve in the next two to five years?



What existing problems might AI or machine learning tools solve faster or with less expertise required from the user? Will AI or machine learning tools lower the barriers to entry for engineering living systems? If so, how?

The application of AI to bioscience and biotechnology is not a recent development. Initial AI tools were limited by the amount of data available to train them;⁷ however, a recent explosion of data has catalyzed rapid progress, driving major advances. A wide range of data—including text-based data, protein structures, DNA sequences, and other experimental results—has contributed to our understanding of biology and has provided fertile ground for the emergence of new AI-bio capabilities.

Recently, AI-bio capabilities have transitioned from the prediction of outcomes to the active generation of content, which marks a significant inflection point and changes the broader landscape of AI’s impact on bioscience and biotechnology. In this report, we discuss three types of AI-bio capabilities: LLMs, biodesign tools, and automated science (see box 1).

Conversational LLM applications, such as ChatGPT, have captured the public imagination by

generating text responses to user prompts.⁸ For the life sciences, these models will enable people to conduct basic research, engineer biology, or simply satisfy curiosity by bringing together and synthesizing large amounts of information. Models specific to biological applications have also advanced rapidly,⁹ and many other breakthroughs are on the horizon. These AI biodesign tools will enable scientists to design new proteins and other features of biological systems much more rapidly for a wide range of applications in medicine, fuels, foods, materials, and other fields. A variety of AI tools and capabilities are coming together to improve the automation of science—from literature searches and AI-driven robotic experimentation to interpretation of results—increasing the pace of scientific advancement. In addition to significant benefits for the life sciences more broadly, all three of these types of AI-bio capabilities will contribute in important ways to public health and pandemic preparedness and response.

BOX 2. KEY TERMS USED IN THIS REPORT

Term	Definition
AI agent	A computer program consisting of multiple independent programs, some of which use AI, designed to work together to carry out more complex tasks than prediction or design.
Application programming interface (API)	A set of protocols and tools that allow different software applications to interact with each other and share data in a standardized way, enabling developers to create new applications without having to start from scratch.
Artificial intelligence (AI)	A computer system designed to simulate human intelligence by performing tasks that typically require human cognition, such as learning, problem-solving, decision-making, and language processing.

continued on next page >

Box 2. Key Terms Used in this Report (continued)

Term	Definition
Digital-physical interface	The point at which digital biological designs begin to be constructed into physical biology. Biological designs will first be constructed in software, after which physical molecules will need to be assembled to make the design into a biological system. The clearest example of the digital-physical interface is DNA synthesis.
Foundation model	A large-scale machine learning model that is trained on vast amounts of data to perform a wide range of tasks, such as natural language processing and image recognition. These general-purpose models provide a basis for further machine learning research and can be fine-tuned for specific applications.
Frontier model	A foundation model that is close to, or exceeds, the capabilities currently present in the most advanced models but differs with respect to its scale, design, or capabilities.
Generative model	A model that can generate new content rather than produce predictions or evaluations of existing content.
Machine learning	A wide range of approaches that focus on developing algorithms and statistical models that enable computer systems to learn from data and make decisions that are based on data.
Open source	A model of software development in which the source code is made freely available to the public, allowing anyone to view, use, modify, and distribute it. The open-source movement emphasizes collaboration and community-driven development, with the goal of creating high-quality software that benefits everyone.
Prompts	Requests delivered to an AI model by a user to elicit a response. The quality of the prompt can have a large impact on the quality of the response.
Pre-training	The process of training a model on large data sets of general information with a loosely defined goal to capture structure, patterns, and relationships in the data. This process can give the model broad capabilities and the potential to perform better on tasks with less available data through the process of fine-tuning.
Fine-tuning	The process of training pre-trained models on a smaller amount of information specific to the task or topic at hand. This approach tends to produce better results than simply training a model on a small data set.
Dynamic vs. static models	A static model is trained once on a defined set of data. A dynamic model continually updates by training on new information.

Large Language Models

Natural language LLMs are a type of AI model that is trained on vast amounts of text data to generate human-like language. They are capable of tasks such as translating language, summarizing text, and answering questions, and have been used in a variety of applications such as chatbots, voice assistants, and language modeling. LLMs are used by a wide range of people for broad purposes, and their capabilities have expanded unexpectedly rapidly. In 2017, AI experts predicted that LLMs would reach language proficiency on par with humans by 2050.¹⁰ More recent estimates have predicted that these capabilities could emerge as early as 2024. Multiple experts, pointing to the October 2023 release of the GH200 NVIDIA computer chips,¹¹ have identified early 2024 as a milestone date when these more powerful chips could yield a significant jump in LLM capabilities.

Natural language LLMs are the focus of this report (see box 1), but LLMs can be trained on a wide range of data, such as audio, visual, and biological data. “Multimodal” models that process many types of data are under development and will expand LLM capabilities. Given the general capabilities of these models, they are often referred to as an example of “foundation models” since they can be further trained—or fine-tuned—to improve their performance on a variety of more specific tasks. Alternatively, when they are trained on a particularly large amount of data and computational resources, they may be referred to as “frontier models,” which represent the state of the art for LLMs. (See box 2.) Frontier models are currently produced primarily by companies in the United States and the United Kingdom, with some development in China. There is extensive interest in LLMs in many other countries around the world, where many types of LLMs are being developed (for examples, see appendix B).¹²

Although natural language LLMs are not designed specifically to facilitate advances in the life sciences, they will change the landscape of how life science research is conducted by supporting

education and training, basic research, and laboratory capabilities. A major advantage of LLMs and their applications such as ChatGPT, Claude, and Bard is their ability to quickly bring together information from many sources and communicate it in accessible language. Several experts believe that these models can help people who are not knowledgeable about a topic rapidly form an understanding that is on par with that of an undergraduate or even a doctoral student. Current models appear to be most useful to users with less expertise in a topic, but it is not clear whether this trend will hold.¹³ Some academics noted that their students often use these tools for help with studying and completing assignments, but the experts themselves found these tools lacking in precision or accuracy when prompted with more technically sophisticated questions, and they believe that current LLMs are unlikely to provide significant novel insights into biological systems. Some LLMs are trained specifically on text from scientific literature,¹⁴ and experts believe that they may be more helpful for students and others seeking to get up to speed on technical topics.

Many experts pointed to the ability of LLMs to help design and troubleshoot molecular biology experiments on the basis of published information and to program robotics platforms to carry out experiments. These capabilities can reduce the need for people to develop the technical skills required to perform the experiments themselves. Current LLMs are generally limited to text-based information, but some can incorporate information from images and videos; this may enable them to more effectively provide feedback and suggestions for troubleshooting laboratory techniques in real time.¹⁵

Notwithstanding these capabilities, there are limits to the amount of tacit knowledge about laboratory work that LLMs could offer. Some experts doubted that the information provided by an LLM would be substantially more enabling than the results of searching for the same information by using

a standard online search engine or watching a video of the work being carried out. In addition, the engineering biology capabilities that LLMs provide are likely to be limited to information and tasks that are well specified and publicly described, and users would still require some fundamental understanding of science and practical laboratory skills to verify that their experiments have yielded the desired results.

Another limitation of LLMs is that they may “hallucinate,” producing incorrect information that they convincingly present as true. Novices may find it challenging to identify these false statements and could easily be misled. This drawback is widely acknowledged, and some, though not all, AI experts are optimistic that LLM developers will significantly reduce this problem over the next five years. A further limitation of LLMs is their rudimentary reasoning abilities, which are prone to fail often, especially when performing tasks that require several sequential steps or logical leaps.¹⁶ These capabilities are likely to improve over time, and developers are working specifically to enhance the reasoning abilities of these models, for example, with chain-of-thought prompting.¹⁷

Biodesign Tools

AI biodesign tools are mostly used by specialists working to design biological systems, and as such are trained on biological data, such as DNA or protein sequences. They typically require more skill and expertise to operate than general-purpose LLMs. Compared with methods that do not use AI, these tools can increase the likelihood of producing successful designs, allowing scientists

to achieve their goals more quickly and with fewer resources and experiments.

Advances in LLMs have contributed to advances in AI biodesign tools. Early in the history of computational biology, researchers recognized that DNA and protein sequences resembled human language in their sequential nature and in the overarching “grammar” that determines their structure and function. Some advances in AI biodesign tools are therefore supported by advances in LLMs, but achieving more significant advances in biodesign tools faces additional challenges owing to limitations in the volume of data available to train the tools.

AI-Enabled Protein Design Tools

Among the available AI-enabled tools for biological design, the capabilities of protein design tools have advanced most rapidly over the last few years. In 2020, AlphaFold 2, an AI-enabled protein structure prediction tool developed by Google DeepMind, garnered significant attention from scientists and the general public when it accurately predicted the three-dimensional structure of approximately 90 percent of the protein sequences it was tested on, a vast improvement over previous methods.¹⁸ AlphaFold 2 is just one tool among many developed in recent years for protein structure prediction and protein design (for a list of biodesign tools, see appendix B). Many existing AI tools that are trained on biological data, like AlphaFold2, do not generate biological designs, but have provided valuable data for training and refining biodesign tools.

Scientists can use protein design tools for a range of beneficial applications, including antibody and vaccine design and novel therapeutics, as

Among the available AI-enabled tools for biological design, the capabilities of protein design tools have advanced most rapidly over the last few years.

well as foods, materials, and improved enzymes for biomanufacturing and other applications. Scientists currently use AI protein design tools such as RFDiffusion¹⁹ and ProteinMPNN²⁰ to generate new protein sequences with desired characteristics related to structure, ability to bind to another molecule, and stability. The landscape of possible protein sequences is vast, and AI

can generate and refine promising candidates. (For more details, see box 3.) These AI protein design tools are typically open source and may be available through platforms such as Google Collaboratory or Hugging Face so individuals can use them without installing any software or acquiring their own computing infrastructure.

BOX 3. AI-ENABLED ADVANCES IN PROTEIN DESIGN

AI has tremendously affected the field of protein design in the past few years.²¹ The protein design process often involves combining multiple design tools to optimize multiple characteristics, such as protein structure, binding characteristics, and solubility. Candidate designs are tested in the laboratory to confirm whether they have the predicted properties and often require further optimization through experimentation. Before the introduction of AI, directed evolution was the primary approach used to design proteins.²² This approach begins with choosing a natural sequence close to the desired design, subsequently mutating it to generate many variants, then selecting those with better properties, and repeating this process until finding a satisfactory result. This approach tests only a small subset of possible variants of the original sequence and thus likely leads to a suboptimal solution.

Generative AI tools can improve protein design in two ways. First, they can generate entirely new sequences that have desired properties, potentially providing a more promising starting point than directed evolution. Second, AI can help select the best variants for experimental testing to understand how sequence affects the properties of interest and thus improve them. Experts familiar with current AI protein design tools reported success rates of 20 to 50 percent for their most successful design tasks. The applications with the highest success rates are likely to be those that require high precision and specificity, but are not intended to affect complex biological systems, such as cells, more broadly.

Although AI prediction of protein structure is relatively mature, there are classes of proteins for which little data exist as a consequence of their fundamental characteristics—for example, being disordered or hard to crystallize, or existing in complexes—and for which existing methods fail. Two types of proteins that pose particular challenges are peptides (i.e., short protein sequences, often comprising about 20 amino acids or fewer), proteins that incorporate non-natural amino acids. Many models struggle to design longer sequences, including sequence lengths that are common for natural proteins. For example, 200 amino acids is the limit for xTrimo-PGLM,²³ a leading AI protein design tool, whereas many naturally occurring proteins contain more than 300. Proteins that form complexes with DNA, RNA, or small molecules have also proven challenging to design.

continued on next page >

Box 3. AI-Enabled Advances in Protein Design (continued)

Some model developers speculate that if we used the same amount of computational resources as LLMs when training protein language models, we could significantly improve their performance. However, there is significantly more natural language data available than carefully annotated biological sequence data. Furthermore, while there are many biological sequences online, many of them do not actually provide much new information. This is because many of the biological sequences that do exist are highly related; for example, many are variants of the same protein and may include non-functional variants. Larger protein design models often use a subset of this data and remove highly similar sequences to improve the model's performance.

How to determine a protein's function from its sequence remains a fundamental question in the life sciences. Researchers have found that as AI models are trained on more data, structure prediction improves at roughly twice the rate of function prediction, reflecting the greater difficulty of predicting biological function.²⁴ AI models are capable of predicting which mutations will disrupt the function of a protein or non-coding sequence but cannot predict if a mutation will result in a new function. Existing functional prediction benchmarks are typically limited to a small number of cases for which many data points are available and can predict only a narrow range of functions. Therefore, although AI models are likely to improve our understanding of the links between protein structure and function, much more work is needed to make these predictions broadly reliable.

AI-Enabled Design of DNA, Biological Circuits, and Cells

Interest in engineering biology has grown over the past 20 years, with the vision of using biological systems to provide a wide range of products and to address difficult challenges such as achieving carbon sequestration and preventing environmental degradation.²⁵ As the field has progressed, bioengineering researchers and practitioners have worked hard to make it more of an engineering discipline than one requiring bespoke designs. Standardized languages, such as the Synthetic Biology Open Language,²⁶ and standardized biological components have enabled the application of design thinking to biological systems. Many experts pointed to significant investment in these areas, which they believe will

support further development of AI biodesign tools to progress toward this set of goals.

That being said, AI biodesign tools for applications beyond protein design face significant challenges, and most are not yet mature. Tools such as ExpressionGAN can design sequences of DNA to better control the timing, the conditions required for protein production, and the amount of protein production in a cell.²⁷ Other DNA design tools can generate DNA sequences that take on a specific three-dimensional shape—known as DNA origami—or bind tightly to targets to act as biosensors or antibodies.²⁸ Researchers have also developed LLMs that use DNA sequence data instead of natural language as foundation models that can be fine-tuned for specific tasks, for example, predicting sequences of DNA that will regulate gene expression or protein production.

There is also significant interest in AI biodesign tools to design metabolic pathways—genetic circuits in bacteria or yeast that can produce a range of small molecules—which are important for biomanufacturing. For example, AI tools such as novoStoic and RetroPath2²⁹ can help choose efficient pathways for producing small molecules, optimize the genetic components of a pathway, and design cells that will grow in large bioreactors (vessels used to manufacture biomolecules at scale).³⁰ Companies are likely to make substantial investments in generating data to improve these types of tools because significant economic drivers exist for these advances.³¹ However, current work in this area predominantly focuses on specific strains of bacteria and yeast and does not transfer well to new species, limiting the applicability of these advances to pathogens, human cells, or other living systems.

Some experts believe that AI biodesign tools will expand the frontiers of what is biologically possible, allowing the design of sequences and functions that are unlike those found in nature. Experts are divided on when this will be achieved; many believe that progress will concentrate in areas that receive large amounts of funding and in which it is possible to quickly generate large amounts of data. Still, some believe that these capabilities will emerge within the next five years as a result of the acceleration of design-build-test cycles, driven by greater testing throughput, improved design accuracy, and automated measurement of results, which will help generate data for training AI biodesign tools.

Limitations of Biodesign Tools

Experts pointed to several limitations in the capabilities and use of AI biodesign tools. A major limiting factor, as noted, is the availability of training data. Models generally perform well where large, labeled data sets exist (e.g., for protein structure) and poorly outside of these specific areas. Technical areas with strong economic drivers, such as metabolic engineering for

industrial processes and high-throughput assays for measuring protein characteristics, will likely produce this type of large data set. Government-supported efforts, such as national strategies for genomic pathogen surveillance, will also boost data availability. However, experts repeatedly pointed out that biological systems are complex and that our ability to measure and generate reliable data about biological functions is limited. Progress may therefore be slow.

Experts in AI biodesign tools reported that only a fraction of designs generated by a given tool are successful, and that a large number of designs need to be created and experimentally tested to select the best candidates for further work. In addition to requiring the laboratory infrastructure and know-how to conduct these experiments, this limits designs to characteristics that can be evaluated efficiently in a laboratory. Furthermore, mistakes made by biodesign tools compound, so the more biological parts and desired characteristics the design tool needs to consider, the lower the likelihood of a successful design. Researchers are exploring ways to improve these models by linking experimental outputs directly back into the models to enable iterative learning.

The utility of AI biodesign tools is currently limited by users' ability to express what they want in a language that the models can interpret. This requires expertise. For example, a biodesign tool that designs proteins for improved binding to a target molecule may require a user to input parameters that are based on the user's detailed technical knowledge about the location of atoms at specific places in three-dimensional protein structures. Some experts believe that in the future, these tools will enable users to design proteins that bind a wide range of targets without having detailed knowledge, such as understanding the details of their molecular structures. Experts point out that the integration of chatbots with these cutting-edge tools could facilitate this communication in natural language, thereby making biodesign tools more accessible to those with considerably less expertise.

Automated Science

The term “automated science” refers to the use of AI to automate steps in scientific discovery or the transfer of the entire process to AI. One of the challenges in scientific discovery is the vast number of possible experiments that could be conducted, making systematic exploration of all options by humans impracticable. AI has the potential to revolutionize scientific discovery by automating this exploration and intelligently choosing the scientific questions that are likely to be the most informative and useful to explore. These models can simulate larger systems than humans can—for instance, the interactions of millions of particles.³² However, AI struggles to capture the rules that govern complex interacting systems with existing data. The ability to make targeted changes to biological systems and measure their effects will improve the ability of AI models to interpret these causal relationships.

AI tools have been used for all steps in the scientific process: researching literature, generating hypotheses, designing experiments, writing software, programming instructions for robotics platforms, collecting data, and analyzing and interpreting results (for more details, see box 4). Some experts believe that more steps will become automated in the near future and that it may become difficult to avoid interacting with AI when carrying out some types of scientific research.

As an example of automated science advances, in 2009, scientists developed a robot scientist, called Adam, that discovered the functions of poorly

characterized genes in yeast and only required human assistance with replenishing experimental reagents and removing waste.³³ Eve, developed by the same group of scientists in 2015, automated early-stage drug discovery to identify new drugs for treating neglected tropical diseases.³⁴ In 2020, a team in Liverpool developed a free-roaming laboratory robot that could autonomously search for catalysts to initiate a desired chemical reaction.³⁵

A more recent approach to automated science is the development of autonomous AI agents that can interact with multiple AI tools to coordinate the completion of a complex task. Examples include AutoGPT, which chains together “thoughts” generated by an LLM to autonomously achieve a goal, aided by its ability to search the Internet and interact with available applications, ranging from simple calculators to advanced AI biodesign tools. An example in chemistry is ChemCrow, which enables the design of chemical synthesis processes using natural language requests, such as “synthesize ibuprofen.”³⁶ Recently, researchers used ChatGPT to write a scientific paper from scratch.³⁷ Provided with a data set, ChatGPT formulated a question, wrote code to perform the analysis, described its methods, and interpreted the findings. Its initial attempts at coding contained mistakes, and parts of the paper contained fabricated information, but additional human prompting corrected these errors.

Some experts expressed concerns about automated science because most users of AI

AI tools have been used for all steps in the scientific process: researching literature, generating hypotheses, designing experiments, writing software, programming instructions for robotics platforms, collecting data, and analyzing and interpreting results.

tools do not have a strong understanding of how they work, which could lead to an overestimation of their abilities and blind trust in their outputs. Users could also assume that algorithms process information in the same way that humans do,

leading to surprises when AI fails in ways that a human would not. AI models often work as “black boxes,” making it difficult to understand and validate the scientific insights that they generate.

BOX 4. ELEMENTS OF AUTOMATED SCIENCE

AI is already contributing to many steps in scientific research, and it is likely that AI tools for automated science will become more integrated in the future to provide a more comprehensive AI-enabled scientific discovery process.

Literature research

AI has substantially improved tools that aid background research. Tools such as scite and Elicit use LLMs to query, interpret, and summarize scientific literature, as well as to allow claims to be checked against original source material to ensure the accuracy of the information they provide. ResearchRabbit builds networks of related research papers based on citations.

Hypothesis generation

Natural language processing of scientific literature can capture complex concepts and make accurate scientific predictions. These models group words that occur in similar contexts, allowing the identification of relationships between words, such as “cat is to kitten” as “dog is to puppy.” These capabilities have been demonstrated to recommend promising hypotheses, such as suggesting materials for functional applications in materials science years before their discovery.³⁸ Limitations for identifying good hypotheses include inaccuracies in published scientific findings,³⁹ LLM hallucinations of incorrect information, and poor ability to judge the novelty of a hypothesis.⁴⁰ However, tools designed specifically to seek novelty do not face such limitations.

Experimental design

Experimentation is often an iterative process, which can be time-consuming and inefficient. AI can collect existing data on a problem—for example, how mutations in an enzyme change its efficiency—and form a map of promising mutation spaces. AI can then select subsequent experiments that explore areas with little data and exploit areas that produce promising results,⁴¹ reducing the total number of experiments required and producing better outcomes. LLMs are currently poor at formulating complex, strategic plans. They tend to have a short “memory,” meaning they forget the start of a plan as they progress through it. However, AI “agents” with different interacting modules have shown more promise in achieving this long-term planning. These agents can describe high-level plans while other more specific models or tools fill in the details of how to accomplish each step.

continued on next page >

Box 4. Elements of Automated Science (continued)

Writing software

General-purpose LLMs can successfully write software code to carry out various tasks but struggle with generating code for more complex or specialized tasks that are not well represented in the training data. More specialist tools, such as Github Copilot, interact with users as they are programming, detecting what the user intends to write and automatically filling it in, which can result in completing projects more quickly.⁴²

Coding instructions for robotics platforms

AI can write software to control laboratory robotics platforms.⁴³ The development and refinement of laboratory protocols can take weeks or months, but laboratory robotics companies are working to enable AI to design laboratory protocols for well-documented experiments described in scientific literature and to interact with commercially available molecular biology kits. Laboratory work is not currently standardized and documented enough to fully automate laboratory protocols, but there is strong pressure to create this standardization. However, the limitations on automating lab work in biology are substantial. Biology is not digital, and the physical constraints of working with biological materials are a major barrier to full automation. Current laboratory robotics workflows are better suited to simple experiments and conditions with a high degree of initial setup and investment, limiting the generalizability of robotic techniques. Robotics platforms struggle with viscous liquids, unfamiliar lab equipment, and mammalian cells. They must also be adjusted and optimized for different organisms.

Data collection

Machine learning can be incredibly data intensive. Historically collected data may not be suitable for machine learning if such data were not collected in a standardized way. Some experts stated that the automation of data production would be key to continuing advances in our ability to design biology with AI. In some settings, data collection is already routinely automated—for example, in high-throughput genome sequencing of pathogens by public health agencies—but this automation is limited in its sophistication.

Analysis and interpretation

Automated analysis of large data sets is a key focus, given human limitations. Although general-purpose LLMs struggle with logic, they excel at coding for data analysis. AI research areas such as knowledge representation and causal inference initially demand human input but can surpass LLMs that rely solely on word associations. INDRA is an example of a collection of resources for building scientific “reasoning” algorithms, which draw on text information and structured databases to collect statements about mechanistic or causal processes in science and assemble them into models.⁴⁴ Equipping agents with the ability to perform their own experiments will likely improve their ability to learn which variables cause changes in others.



Biosecurity Implications

The same AI-bio capabilities that will provide significant benefits may also empower malicious actors to misuse biology to cause harm, with potentially catastrophic global consequences. Key questions about the biosecurity implications of these models include:



What are the main concerns about biosecurity risks in the application of AI or machine learning methods to the life sciences and to engineering living systems, if any?



Will advances in AI and AI biodesign tools reduce barriers to engineering pathogens and other biological agents? What obstacles will likely remain in the next two to five years?



What types of AI models and AI biodesign tools carry the biggest risk of misuse?



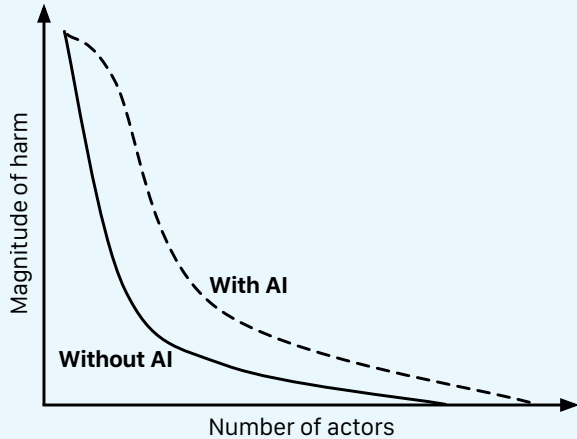
How can the scientific community leverage AI advances to bolster biosecurity and pandemic preparedness?

LLMs, AI biodesign tools, and AI-enabled automated science are likely to change the landscape of biosecurity risks in different ways, depending on the number and type of actors that may use them and the types of capabilities they confer (see Figure 1). Many experts believe that LLMs could expand the number of people able to cause harm with biology. They could help malicious actors become familiar with a range of known biological agents and could provide resources that help them obtain, construct, or otherwise develop these agents. However, experts disagree about the implications that this may have for biosecurity, as some believe that the information provided by LLMs can already be obtained in other ways and that malicious actors would need additional skills and resources beyond what an LLM could offer.

AI biodesign tools generally focus on narrow scientific questions, and they are used by researchers and others with significant scientific expertise. Although fewer people are likely to use these tools, some experts believe that they are more likely than LLMs to generate biological designs for novel toxins, pathogens, or other agents that could be more harmful than those found in nature. However, biodesign tools are currently limited in the types of designs they can reliably generate, and there is uncertainty about how quickly this will change.

Any malicious actor hoping to engineer a biological agent will face significant hurdles beyond obtaining a design, including access to biological components, laboratory infrastructure, and laboratory training sufficient to build, test, and deploy the designed agent. Experts in AI biodesign tools also cautioned that the designs created by these tools require validation. Users need time and expertise to evaluate and optimize the many candidate designs that these models produce. For all AI-bio capabilities, the biosecurity implications depend both on the characteristics of the AI tools and the resources and abilities of the actors who might misuse them.

FIGURE 1



LLMs and AI biodesign tools are likely to shift the landscape of risks of accidental or deliberate misuse of biology. LLMs may enable more actors to cause harm, while future biodesign tools may increase the ceiling of harms that could be caused by a biological agent.

Source: Figure adapted with permission from Jonas Sandbrink, "Artificial intelligence and biological misuse: Differentiating risks of language models and biological design tools" (submitted manuscript, August 12, 2023), arXiv, <https://arxiv.org/abs/2306.13952>.

Large Language Models

LLMs are broadly enabling technologies that can quickly bring together publicly available information and communicate highly technical information to non-experts. Although many of the most powerful LLMs have some access controls, mainly for commercial and competitive purposes, most experts believe that reasonably powerful LLMs will likely continue to be openly available to the public. LLMs are also rapidly advancing, and models that are considered powerful today are likely to become obsolete very quickly.

There are many ways that LLMs could be used to cause harm, from providing instructions for building bombs and exploiting cybersecurity vulnerabilities to suggesting destructive behaviors

Nearly all experts pointed to the possibility that a malicious actor could use an LLM to obtain information on how to use a toxin, pathogen, or other biological agent to cause harm. However, experts are divided on how useful this information might be.

to at-risk individuals.⁴⁵ For biosecurity-specific hazards, experts raised many different types of concerns. Some believe that LLMs could raise awareness about potential routes to misuse biology to cause economic or environmental damage, for example, by targeting agriculture or vulnerable ecosystems. LLMs could also exacerbate or create opportunities for misinformation or disinformation, which could intersect with biosecurity by undermining public confidence in public health efforts, injecting false information into pathogen surveillance or response systems, or incorrectly assigning blame for causing an epidemic or pandemic.

Beyond these broad biosecurity concerns, nearly all experts pointed to the possibility that a malicious actor could use an LLM to obtain information on how to use a toxin, pathogen, or other biological agent to cause harm. LLMs could also direct such a person to additional resources or tools helpful for obtaining biological components, such as pathogen DNA, and getting up to speed on simple laboratory techniques. However, experts are divided on how useful this information might be. Some argue that although LLMs can gather information more quickly, they add very little to what has long been possible by searching the Internet for publicly available information. In addition, LLMs may “hallucinate” incorrect information and present it as true, and individuals without expertise may be unable to recognize this misdirection.

Experts also disagree about the level of tacit knowledge about laboratory techniques that LLMs can provide and how much of this knowledge

is necessary to generate, scale up, and deliver a harmful biological agent. As described in the previous section, LLMs may be helpful for inexperienced scientists by providing information about laboratory techniques and suggestions when an experiment fails. Future LLMs may be able to provide more extensive and accurate feedback that incorporates recorded videos of experimental procedures and other types of inputs. Still, many experts in laboratory bioscience believe that a malicious actor would likely face hurdles to generating a pathogen that would require significant resources, infrastructure, and multifaceted expertise to overcome (box 5).

To get around the challenge of developing laboratory skills and tacit knowledge, LLMs have directed users to opportunities for outsourcing laboratory experiments and infrastructure to contract research organizations and other vendors.⁴⁶ The extent to which a malicious actor could successfully contract with such external vendors to facilitate the construction and scale-up of a harmful biological agent remains unclear. Still, several experts highlighted this type of LLM behavior in calling for additional biosecurity oversight among providers of life sciences products and services (see Risk Reduction Opportunities).

Experts believe that LLMs will also help people with expertise in one area develop expertise in related fields. For example, LLMs could help someone with some training in molecular biology quickly find relevant literature and important information about virology, including how to generate infectious agents from non-infectious components. These

users may already have tacit knowledge related to laboratory techniques and access to laboratory infrastructure, and they may be better able to distinguish useful information from an LLM's incorrect hallucination. These medium- to high-skilled users could obtain information related to biological agents without access to an LLM, but LLMs may facilitate the process.

Because current LLMs draw on information that is readily available, most experts believe that they are unlikely to generate designs of biological

agents that are outside of already established risks. A few experts believe that LLMs could already or soon will be able to generate ideas for simple variants of existing pathogens that could be more harmful than those that occur naturally, drawing on published research and other sources. Some experts also believe that LLMs will soon be able to access more specialized, open-source AI biodesign tools and successfully use them to generate a wide range of potential biological designs. In this way, the biosecurity implications of LLMs are linked with the capabilities of AI biodesign tools.

BOX 5. HURDLES TO PATHOGEN ENGINEERING

A malicious actor or small group would face several technical barriers in trying to generate an infectious agent. A previous NTI report on another type of enabling biotechnology, *Benchtop DNA Synthesis Devices: Capabilities, Biosecurity Implications, and Governance*,⁴⁷ detailed these hurdles, which include synthesizing pathogen genomes, "booting up" infectious agents from DNA, and designing successful alterations to pathogens.

LLMs may be able to provide guidance that would help reduce barriers related to synthesizing pathogen genomes by providing details on DNA sequences to order and simple instructions on how to combine them into longer stretches of DNA using standard molecular biology techniques. However, for most pathogens, generating an entire genome would still require significant expertise and troubleshooting abilities.

Most viral pathogen genomes are not infectious on their own, and making them into viable pathogens requires laboratory infrastructure and knowledge about how to generate infectious agents from their genomes. LLMs could help reduce this barrier by bringing together information on necessary reagents, biological components, and published protocols on how to do this. One expert believes that recent advances in virology have further reduced this barrier.

It remains difficult to generate designs of pathogens with specific characteristics, but it is possible that AI biodesign tools will be able to surmount this barrier in the future by providing candidate designs. These designs would still require experimental validation. AI-enabled laboratory automation could facilitate some of this work, but automation of some types of experiments will remain difficult, including those involving viral pathogens or complex pathogen traits (e.g., transmissibility).

even more challenging when adding complexities related to interactions with genetically diverse human hosts, transmissibility within large populations, and other features of potential biological weapons (box 6). For this reason, some experts are less concerned about the possibility of an AI biodesign tool successfully designing wholly new types of biological agents. Designs for simpler alterations to existing proteins, pathogens, and agents are likely to be more reliable, at least in the near term.

Many experts pointed to a near-term and specific risk: AI protein design tools will make it more difficult for DNA providers to conduct effective

biosecurity oversight of ordered DNA. Many of these vendors currently screen customers and DNA orders to reduce the risk of providing pathogen or toxin DNA to customers who lack a legitimate use for it or inadvertently selling the building blocks of dangerous pathogens to malicious actors. Current DNA sequence screening methods evaluate how similar ordered DNA sequences are to known pathogen or toxin DNA. However, new AI protein design tools can design proteins that have very little similarity to known pathogen or toxin sequences but have the same functions and pose the same risks. These tools could allow the redesign of existing hazards, thus evading DNA sequence screening.

BOX 6. PREDICTING PATHOGEN TRAITS

Experts in biological weapons pointed out that much of the challenge in developing an effective weapon is anticipating how it will interact with the complex world it is released into. AI tools are far from being capable of this level of complex and conceptual analysis.

COVID-19 provides an example. The genome sequence of the SARS-CoV-2 virus became available early in the pandemic, but this sequence provided insufficient information to enable scientists to predict transmission routes, pathogenicity, or transmissibility. These traits are determined by multiple interacting genes belonging to the virus, as well as environmental conditions, genetics, and immunological characteristics of possible host populations, and a variety of other factors. Scientists also struggled to predict the course of the pandemic because it depended on public health responses, the behavior of populations, and other complex social interactions.

Predicting pathogen transmissibility, host range, and virulence from genome sequences using AI is an active area of research. However, AI tools struggle to generalize to new strains and assume all sequences come from viable pathogens, which may not be true of a newly designed strain. They are also data limited; the number of variables the models need to fit is many orders of magnitude larger than the number of available examples to learn from, so the tools are limited in what they can successfully infer. Infection outcomes are difficult to measure, particularly in humans, and laboratory experimental models for mimicking human infection are currently rudimentary. Efforts to generate data for AI-enabled risk prediction are ongoing, and high-throughput systems for characterizing the risk posed by viral variants, coupled with AI analysis of the results, are in development.⁴⁸

Automated Science

Few experts raised concerns about the risks posed by the misuse of automated science by malicious actors. Those who did pointed to the creation of AI models such as ChaosGPT,⁴⁹ an LLM designed to pursue the destruction of humanity and other malicious objectives, and believe that a similar AI model could someday be designed to misuse biology. An AI agent could achieve its aims, for example, by using AI biodesign tools and by outsourcing laboratory work. Some experts also believe that the use of AI-enabled automation could contribute to a larger program for biological weapons development or production, such as one run by a state actor. Another concern is the potential misuse of knowledge and data produced by automation. In particular, data on pathogen functions of concern are currently only sparsely available, but automation that produces such data for public health purposes could provide training data for AI biodesign tools. These additional data could expand the capabilities of AI biodesign tools, including their potential for misuse.

Biosecurity Benefits of AI-Bio Capabilities

AI-bio capabilities are likely to benefit biosecurity in many ways. Experts believe that AI tools and capabilities will be incorporated into many aspects of pandemic preparedness and response, including biosurveillance, development of vaccines and other medical countermeasures, and logistical responses to outbreaks. AI tools also can be used to help determine whether novel pathogens are genetically engineered and can facilitate attribution. The result of these advances will be earlier detection of pathogens, faster and more effective responses, fewer logistical challenges, and better protection against infectious disease. (For more details, see box 7.) These implications for biosecurity are substantial, and efforts to reduce risks posed by AI-bio capabilities should also secure these benefits.

BOX 7. AI FOR ADVANCING BIOSECURITY AND PANDEMIC PREPAREDNESS

AI tools are already contributing to biosecurity and pandemic preparedness and are likely to become more integrated into these capabilities in the future.

Biosurveillance

Outbreak reporting tools such as NATHNAC and PulseNet provide data on outbreaks worldwide to medical professionals and public health teams that could be processed more efficiently using AI. Public health laboratories are automating their data analysis processes to flag only concerning trends for human review.⁵⁰ By analyzing data from returning travelers, AI tools could model the frequency of infectious disease importation and trace its origins. When combined with genome sequencing, this approach could be instrumental in identifying outbreaks as well as uncovering enduring reservoirs of pathogens.⁵¹

Many efforts focus on predicting the risk posed by new pathogen strains. AI tools to rapidly analyze the DNA of pathogens will enable scientists to identify potential pandemic pathogens and high-risk variants before they spread widely. This knowledge guides the proactive design of

continued on next page >

Box 7. AI for Advancing Biosecurity and Pandemic Preparedness (continued)

medical countermeasures. AI is also being used to design DNA, RNA,⁵² and protein⁵³ sequences that can act as biosensors for detecting dangerous pathogens or toxins.

Metagenomic surveillance—for example, of wastewater—can also identify pathogens circulating in communities. Given the volume of data produced by these approaches, AI tools such as anomaly detection can help flag new threats. Nucleic acid panels are also a low-cost alternative to DNA sequencing for continuously monitoring infectious diseases of concern. Designing panels that can detect a large number of pathogens can be technically challenging, but AI optimization techniques can simplify their design.

Medical countermeasures

Many experts who work with AI protein design tools believe they will substantially improve vaccine and antibody design over the coming years. One expert estimated that these tools could enable the design of new antibodies based on a pathogen’s genome within days and allow them to be produced within weeks. Older methods require months to produce antibodies and require access to patient samples. Experts also estimated that mRNA vaccines could be designed and deployed in as little as two to three weeks, rapidly stemming outbreaks. AI models can also design novel antimicrobial drugs, phage therapies, and protective probiotics, though development of these models is more challenging and these tools are not mature enough to scale widely.

Outbreak response logistics

Forecasting the spread of disease can inform policy decisions, direct containment measures, and help hospitals prepare for periods of high demand. New methods are incorporating data on human movement patterns, reports of symptoms by hospitals or on social media, and traditional outbreak modeling using AI. AI can also design optimal delivery routes for tests, vaccines, personal protective equipment, and medical countermeasures, and recommend infrastructure purchases such as cold-chain equipment to optimize the resilience and adaptability of vaccine delivery routes.

Attribution

Recent promising results suggest that AI tools can identify genetically engineered organisms and attribute them to their lab of origin.⁵⁴ These types of tools could help identify actors who design harmful biological agents and act as a deterrent. However, attribution tools will be less effective if actors can make design choices that allow them to evade attribution.

AI watermarking is another avenue for attribution, in which models place subtle signatures in AI model outputs to mark that they were generated by AI. Models could also potentially place watermarks unique to the model or user, further facilitating attribution. Watermarking technologies have been considered for biological designs for the purpose of protecting intellectual property.⁵⁵ For the scientific community to adopt this approach for marking DNA or protein sequences, watermarks would need to preserve the biological activity of the desired product.

Risk Reduction Opportunities

As AI-bio capabilities are applied to solve challenges in the life sciences, they raise the possibility of their misuse to cause harm. There are many approaches to reduce these risks, and each approach has advantages and drawbacks. This section discusses ideas and perspectives on the following key questions:



What risk reduction measures should we consider that will offer meaningful protections against the worst risks without unduly hindering scientific advances and innovation?



*What are the most promising options for safeguarding AI-bio capabilities?
What approaches are most likely to work?*



Who should be responsible at the national and international levels for governance and safeguarding of AI-bio capabilities?

There are many opportunities to reduce the risk that AI-bio capabilities could be misused to cause harm. Some of these are specific to the AI models themselves, including a range of options and suggestions for “guardrails” that describe how AI models could be developed or controlled to minimize the risk. Many experts also believe that it will be critical to bolster biosecurity oversight at the interface where digital designs become physical biological systems, for example, by strengthening biosecurity frameworks for DNA synthesis providers and other life sciences vendors. Some proposed solutions are very broad, including the suggestion to invest further in overarching pandemic preparedness and response capabilities. These different ideas are not mutually exclusive, and an all-of-the-above, layered defense may be needed to reduce risks most effectively.

For each approach, it will be important to balance the need to reduce risks with the need to ensure that AI-bio capabilities can be used for beneficial purposes. As previously mentioned, many experts believe that AI will bring significant benefits for the life sciences broadly, and for biosecurity and pandemic preparedness specifically. Many experts also pointed out that no solutions exist that will eliminate all risks related to AI-bio capabilities. Each of the approaches described in this section should be understood as a hurdle that decreases the chances that a malicious actor will successfully misuse AI tools to cause biological harm and that this misuse will lead to a global biological catastrophe.

Guardrails for AI Models

Experts raised many ideas for guardrails that are already being implemented or that could be further explored to reduce the risk that AI-bio capabilities are exploited to cause harm. These include safeguards built into the models themselves, biosecurity evaluations of the models and their technical safeguards, as well as ways to control access to the models, to computational

infrastructure, or to the data needed to train models. Given how rapidly AI-bio capabilities are being developed and the significant uncertainty about how they will evolve, many experts believe that it will be important to have ongoing opportunities for feedback and iterative refinement on how guardrails are implemented. A few experts noted that guardrails for AI biodesign tools in particular are lacking and that this is an important area for further development.

Several experts pointed to methods to ensure that AI models have appropriate oversight and incorporate technical safeguards to limit their potential for misuse. Developers of AI models, including companies and academic researchers, could have an institutional review process to ensure that dual-use and ethical considerations inform the development and deployment of AI models. Such oversight mechanisms are already established or are under active development in many companies that produce LLMs, but this approach has not yet been incorporated into the development of AI biodesign tools.

Incorporating Technical Safeguards into AI Models

Experts described several technical safeguards that some LLM developers are actively implementing to reduce a wide variety of risks, including those related to biosecurity (box 8). For example, developers can train models using adversarial approaches to discourage the incorporation of harmful concepts into a model. Models can also evaluate outputs for harmful content before they are shown to the user or refuse to answer user requests on specific topics. Developers can run safety checks at multiple stages during the training process to ensure a model is safe during its development. It is not yet clear which methods will be most effective, and this is an active area of inquiry. Although technical safeguards have been developed for a wide range of AI models, built-in solutions for AI biodesign tools are still lacking.

To implement and test many of the technical safeguards discussed here, model developers need to understand the types of biosecurity risks that they should guard against, which may require detailed information about biological agents and vulnerabilities. Yet, this can also pose a significant challenge because distributing this type of information may be hazardous.

In addition, experts frequently point out that technical safeguards incorporated into models are only feasible for models with access controls, for example, through an application programming interface (API). Malicious actors or others seeking to circumvent these types of safeguards could easily strip any safeguards incorporated into open-source models.

BOX 8. TECHNICAL AI SAFEGUARDS

Intervention point	Technical safeguard	Description
Training data	Removal of dual-use training data	AI developers can remove a portion of training data gathered from public sources if it is considered risky to include, such as papers on dangerous pathogens, or laboratory protocols for constructing and booting viruses.
Model training methods	Adversarial training methods	Adversarial methods aim to improve AI models by introducing a module to penalize undesirable outcomes. For example, adversarial debiasing methods penalize the prediction of sensitive characteristics in the data such as race or gender. Generative adversarial networks use an adversarial component to train the model to produce outputs that are indistinguishable from outputs not generated by AI.
	Reinforcement learning with human feedback	Reinforcement learning trains an AI to maximize an objective. Human feedback ⁵⁶ tells the AI whether its behavior is good or bad, thereby incentivizing the model to produce good outputs without needing to explicitly define what humans think is good.
	Constitutional AI	Constitutional AI is a method to train AI models to be more helpful and harmless by encouraging them to follow certain ethical principles, or a "constitution," thereby reducing the need for human oversight of models. ⁵⁷
Model behavior after deployment	Refusals and blacklisting	Refusals are when AI models refuse to follow user requests, typically because the information or action requested may be harmful. Alternatively, blacklisting can prevent models from using specific words or phrases in their outputs.

continued on next page >

Box 8. Technical AI Safeguards (continued)

Intervention point	Technical safeguard	Description
Model behavior after deployment (continued)	Watermarks	Watermarks are embedded signatures in the generated output that are invisible to humans but detectable by an algorithm. ⁵⁸ They can be used to signal that something has been AI generated or to attribute it to an author.
	Certifiable origins	Designs produced by AI could require a cryptographically signed certificate detailing the inputs used to design an output. These certificates could be required by third parties, for example, DNA synthesis companies, to check for harmful intent captured in the requests made to the model.
	Other techniques that attribute designs to users	Other methods, such as maintaining server logs, can be used to store a record of user behavior that can be searched to attribute a design to a user.
Model access	Application programming interface (API)	People can interact with a model through an API, prompting the model and viewing its outputs without having a copy of the model. ⁵⁹
	Token-based systems	In these systems, users of a model spend “tokens” each time they use it to generate outputs. These systems can help manage the burden on companies that provide access through the cloud or APIs and can be a source of income while also limiting misuse.

Evaluating Models

Nearly all experts highlighted the importance of evaluating AI models before they are made widely available.⁶⁰ Evaluations can provide important information on the effectiveness of built-in technical safeguards in preventing misuse and how these safeguards could be updated and improved. Even in the absence of technical safeguards, an evaluation can help determine the types of risks that an AI model may pose, thereby decreasing the possibility of surprise.

Red-teaming is one commonly cited method of evaluation. It is a process in which experts test a model to determine if it will provide hazardous information or can be misused. Many LLM developers have participated in red-teaming efforts, either as internal efforts by developers to ensure their models were safe to release or by contracting with third parties. Experts believe that consulting biosecurity experts is a key component of these reviews, as many AI developers lack the expertise to identify information that raises biological risks. Also, given the tendency for LLMs

to hallucinate incorrect information and to provide information with factual errors, expertise is needed to evaluate the outputs of models.

In assessing AI's capabilities and risks, it is important to note that most AI models to date are static. These models are trained once and rely on that historical data to answer queries. Dynamic models can be continually retrained or augmented training with an ability to search the web and provide information that is more current. The behavior of models that rely only on their original training data is easier to anticipate, whereas the behavior of models that change over time or that retrieve information on an ongoing basis can be more difficult to anticipate and may limit the effectiveness of evaluation.

Experts expressed uncertainty and a range of opinions about how these evaluations should be done and what would constitute a "safe" AI model. It is also unclear to what extent red-teaming should test the outputs of these models to determine whether they are genuine risks or whether the ideas or designs provided will ultimately fail. An in-depth evaluation of the risks posed by these tools could run into legal and ethical barriers. For example, assessing whether an LLM can acquire a controlled chemical would be illegal, and using a benign chemical as a proxy would not fully answer the most relevant question. Similarly, evaluators can test whether LLMs and AI biodesign tools will volunteer suggestions for novel hazardous biological agents, but testing whether the designs work might be irresponsible.

Testing may involve building an adversarial agent or set of requests, which would require knowledge of potential hazards. Experts raised the concern that a resource of this type could include information hazards (see page 39) both because it may contain specific, high-consequence risks and because it could act as a roadmap for those seeking to bypass model safeguards. One technical solution could be to map sensitive information to less sensitive proxies, and then test the model and implement safeguards on the basis

of these proxies. However, some experts cautioned that model safeguards could perform well on proxy data without fully capturing the risks that would exist with the actual data.

Monitoring Models

Experts generally believe that it will be important to evaluate AI models for their misuse potential prior to their release; however, they have low confidence that they can capture all forms of misuse, particularly because model capabilities are not fully explored prior to their release. Therefore, a few experts suggested mechanisms to monitor the behavior of AI models after their release. If a model is run on a developers' computing infrastructure and made available through an API, then developers can directly monitor the model's outputs in response to prompts to determine whether the model is providing potentially harmful information. To make monitoring of AI models easier, AI oversight models could monitor model outputs and flag concerning results for human review.

Absent direct control of AI models, AI model developers or others could implement reporting mechanisms for users to flag concerning behaviors or outputs. For example, systems could be established to support public reporting of cases in which an AI model has resulted in harm⁶¹ or reporting of risks directly to the developers or a third party. As an incentive to report potential risks, "bug bounties" could provide financial rewards to the reporter.⁶²

Controlling Access to Models

Many experts believe that controlling access to AI models is a fundamental strategy to prevent their misuse. They point out that any technical safeguards that are incorporated into a model to reduce its potential for misuse can be stripped out if the full model is released as an open-source tool. Some LLMs and many AI biodesign tools are fully open source.

Many larger LLMs, including GPT-4, Claude, and PaLM, use APIs that allow the model developer to keep the model itself closed while enabling users to enter queries and receive outputs. These APIs enable the model developer to monitor user prompts and restrict outputs of potentially harmful information. They also enable the developer to maintain control over the model, including through any of the technical safeguards described earlier, and to update the model when needed. Some LLM developers restrict access to their models to a small group of beta testers early in their development as part of a staged release.⁶³ This type of staging provides an opportunity for evaluation and the mitigation of potential risks in advance of broad distribution.

Many experts believe that more developers of AI biodesign tools should consider access controls to reduce the risk of misuse; however, many of these tools are developed in the academic community, where strong cultural norms support open-source resources. Often, tools are developed collaboratively across loosely affiliated groups of people, and requirements for publication include the need for peer review, which often includes access to published AI models. The expectation for many AI biodesign tools is that they will be used as a foundation for future work, including alterations of the tool itself to improve it and to apply it in novel ways. Furthermore, one knowledgeable expert pointed out that maintaining control of a model and establishing APIs requires institutional infrastructure that may not be available to many academics.

Several experts in academia believe that cultural norms supporting open-source AI biodesign tools could change, but this change would require awareness-raising and engagement across the academic community, including with publishers and funders. These experts emphasized that the benefits and drawbacks of restricting access to these tools would have to be carefully weighed to ensure that this approach does not limit further study, collaboration, or scientific progress. Because there is a wide range of biodesign tools, it will be particularly important to evaluate them for misuse potential and to ensure that any access restrictions are commensurate with the risks that they pose. Additionally, restrictions could disproportionately affect researchers in low-income countries, raising important equity considerations.

Some experts suggested that access controls for AI biodesign tools could incorporate customer screening or “Know Your Customer” requirements. For example, developers of these tools could restrict model access to individuals who have institutional affiliations and reasonable use cases. It is unclear how such a system would be implemented by the wide range of model developers, many of whom are in academia and are not currently equipped to screen users. A centralized credentialing system that verifies users (as described on page 41) could help in this regard.

AI models could also monitor the use of AI-bio capabilities and identify concerning behavior

Many experts believe that more developers of AI biodesign tools should consider access controls to reduce the risk of misuse; however, many of these tools are developed in the academic community, where strong cultural norms support open-source resources.

by users. Several experts were optimistic about the ability of AI to analyze patterns of behavior, such as gathering information from an LLM on specific topics combined with purchasing life sciences products, to identify customers with potentially malicious intent. A similar project has demonstrated the value of this type of monitoring of publicly available data for detecting high-risk or illicit nuclear trade.⁶⁴

A few experts raised the possibility that governments could control access to AI models by implementing export controls on models that meet specific requirements. Owing to the challenge of restricting access to software tools, experts see this approach primarily as a way to slow the spread of these tools rather than a means to prevent their use. Also, it may be difficult to implement export controls on tools developed as open-source resources, including many biodesign tools.

Controlling Access to Computing Infrastructure

A small number of experts raised the possibility of controlling (or monitoring) access to high-performance computing infrastructure to ensure that powerful AI models are developed only by responsible users. This infrastructure includes resources provided by large cloud computing vendors such as Amazon and Microsoft, as well as government-funded infrastructure provided for national research and development efforts. Because training state-of-the-art AI models, particularly LLMs, requires large amounts of computational power, access to computational infrastructure may provide an opportunity for overseeing or restricting the training of large AI models. For example, to ensure that model developers are legitimate, access to these resources could require a license. Cloud computing providers could require staged safety checks of AI models or other safeguards as part of their usage agreements. Chip manufacturers could also impose limits on the hardware itself, similar to how some graphics cards limit the speed at which

they can be used to mine cryptocurrency; the feasibility of this method is unclear and warrants further review.⁶⁵

However, many experts were broadly pessimistic about controlling access to computing infrastructure as a means to reduce biosecurity risks. They did not believe that high-performance computing infrastructure was needed in order to build an LLM capable of being misused to cause harm. Many of the largest LLMs are trained on supercomputers, but techniques have been developed to fine-tune large models on modest computing resources, such as a personal laptop.⁶⁶ Many AI biodesign tools can also be trained with modest computing resources. Furthermore, model developers are actively pursuing methods to decrease the amount of computational infrastructure that is needed. Therefore, these experts believe that computational infrastructure may not provide a meaningful opportunity for oversight in the future.

Controlling Access to Data

A few experts believe that restricting access to specialized or particularly harmful data could help reduce potentially harmful outputs from AI models and could prevent bad actors from training their own models. Experts listed a wide range of data, including, for example, pharmaceutical company databases on protein and chemical toxicity, publicly available pathogen genomes, gain-of-function research, and information related to historical bioweapons programs. They disagree about what types of data should be restricted, and many are skeptical about the effectiveness of controlling access to data for biosecurity purposes. Much of the data described are already publicly and redundantly available on the Internet, and it would be very difficult to prevent some types of models, including LLMs, from accessing such data. One solution could be for AI developers to agree not to use publicly available data to train models. A suite of resources is now available to verify whether sensitive data were used to

Accurately identifying meaningful risks will require collaboration with a range of experts in synthetic biology, infectious disease, biosecurity, national security, and other fields.

train a model,⁶⁷ allowing verification of these commitments.

A few experts believe that restricting access to pathogen genome data in particular would unduly hinder legitimate scientific research, public health, and biosecurity efforts. In addition to affecting research on pathogens, removing pathogen data from more general biological data sets would substantially reduce those data sets because an outsized proportion of DNA and protein sequence records in public databases originate from pathogens. As a result, the removal of access to these data could hamper broader efforts, such as development of AI protein design tools or protein structure prediction. Other challenges to controlling data for this purpose are more systematic. For example, many experts believe that it would be difficult to decide which data should be restricted and who should be responsible for controlling access. Depending on the type of data, legitimate model developers would also need exceptions or ways to access the restricted data. All of these questions raise important issues of equity and access.

Coordinating Efforts in AI Guardrail Development

Many experts, particularly those familiar with LLMs, pointed to a need for collaboration and diverse expertise to effectively identify and reduce biosecurity risks that might arise from AI models. Accurately identifying meaningful risks will require collaboration with a range of experts in synthetic biology, infectious disease, biosecurity, national security, and other fields—in addition to model

developers and others familiar with advances in AI. Risk assessments of this type are very new, and multidisciplinary efforts will likely be needed to understand and track how the risk landscape is changing over time.

Some LLM developers have already begun to collaborate with biosecurity experts and others to evaluate and reduce biosecurity risks related to the misuse of their models. However, these efforts are ad hoc, and few opportunities exist for model developers to learn from others' experiences. Furthermore, awareness and understanding of risks and potential solutions vary widely across developers, and the development and dissemination of best practices could benefit the entire community. A few experts pointed out that very little information exists about how developers of smaller models and those in other parts of the world are approaching these issues.

Experts disagree to some extent about how open collaboration among model developers and others should be. A more open process would best ensure participation and engagement of a wide range of model developers, including those from non-Western countries, and from a broader set of experts in biosecurity and related fields. However, the need to successfully develop and adopt biosecurity safeguards would have to be balanced with the need to limit information hazards. Furthermore, it is not clear how much AI model developers, particularly those developing LLMs, will be willing to share about their technical safeguards because the details of how these methods are implemented might reveal proprietary information or raise intellectual property concerns.

Managing Information Hazards

Experts repeatedly expressed concern that efforts to reduce risks related to the use of AI in the life sciences could generate information that could enable a malicious actor. These information hazards could include a list of pathogens or ideas that biosecurity experts believe are especially dangerous, a database of protein functions of concern, or a resource that highlights types of data or information as especially enabling for development of bioweapons. This challenge is compounded by the need for broad collaboration among model developers, biosecurity experts, and others, as described earlier.

Despite the risk, many experts stressed the need to develop these resources so that model developers can effectively implement safeguards. A few experts, including some familiar with LLMs, have been frustrated at the reluctance of experts to describe risks with specificity, which they say has hindered efforts to mitigate potential harms and exacerbated uncertainty about the level of risk.

Experts concerned about information hazards were divided about how they should be managed. A few believe that risk reduction efforts likely to raise significant information hazards should take place in closed environments within government intelligence or security agencies. Some believe that government-supported development of these risk reduction tools and resources should be done collaboratively with a restricted set of trusted model developers, companies, institutions,

or individuals who need to have access to this type of information to develop or implement safeguards. One knowledgeable expert suggested that a formalized mechanism independent of governments would be best positioned to balance the need to share information with the need to prevent dissemination of information hazards.

Existing legal frameworks also restrict what types of information can be shared with AI model developers, model testers, and others. Export control laws currently capture technical information that could be used in a biological weapons program,⁶⁸ but it is not clear what types of resources might fall into this category. For example, experts familiar with current software tools and databases for DNA sequence screening report significant uncertainty about whether these resources fall under export control regimes. It is likely that the same uncertainty will apply to the range of tools developed in the future for understanding and reducing risks related to AI-bio capabilities. Several experts believe that export control laws should be clarified to facilitate responsible sharing of information.

Another important point raised by security experts is that any data that are publicly released or documents that governments declassify are likely to be quickly incorporated into LLMs and will be more readily available to the public. This factor should be considered as governments and others determine what types of information and databases should be released and how.

Experts repeatedly expressed concern that efforts to reduce risks related to the use of AI in the life sciences could generate information that could enable a malicious actor. Despite the risk, many experts stressed the need to develop these resources so that model developers can effectively implement safeguards.

Bolstering Biosecurity at the Digital-Physical Interface

The digital-physical interface in biology refers to the point when an actor begins to develop a digital design produced by an AI model into biological reality. As described in the previous section, access to biological components, laboratory infrastructure, and scientific skills are significant hurdles that a malicious actor would face in attempting to misuse AI-bio capabilities to generate a biological agent. For this reason, many experts believe that the digital-physical interface is an important point for oversight. Biosecurity could be improved at this interface in several ways, including strengthening biosecurity frameworks for DNA synthesis screening and improving and expanding customer screening practices to a broader range of providers of life sciences products, services, and infrastructure.

Strengthening DNA Synthesis Screening

Many experts believe that expanding DNA synthesis screening practices is an important way to reduce the risk that a malicious actor will gain access to pathogen or toxin DNA. As noted earlier, many DNA providers already conduct biosecurity screening, which includes both screening of customers to ensure their legitimacy and screening of DNA orders to determine if the sequences match known pathogen or toxin DNA.⁶⁹ In 2010, the U.S. government issued guidance that recommends these practices among DNA providers,⁷⁰ and an updated version of that guidance is expected soon. This type of screening is not yet required anywhere in the world, and incentives are lacking. Many experts believe that strengthening this framework by establishing regulations or other types of incentives will be important as AI-bio capabilities could enable a broader range of people to attempt to misuse biology. To further expand DNA

synthesis screening internationally, a few experts suggested additional support for the International Biosecurity and Biosafety Initiative for Science,⁷¹ an organization incubated by NTI and expected to launch in late 2023 to strengthen global norms and provide resources needed for DNA synthesis screening.

Several experts also pointed out that new technical approaches for DNA sequence screening will be needed to keep pace with new DNA and protein designs from AI biodesign tools. Current DNA sequence screening tools evaluate sequences based on their similarity with known pathogen and toxin DNA and cannot detect sequences designed to cause harm, if they diverge significantly from natural sequences. AI protein design tools will enable the development of novel toxins with the same function as known toxins but different biological sequences.

A few experts called for dedicated funding to support a research program to bring DNA sequence screening up to date with current threats. A program of this type would entail developing AI models that can identify novel sequences designed to function in the same way as known hazards. Experts pointed out that implementing and sharing these models with DNA providers internationally will require a reevaluation of legal frameworks, including export control rules, that can hinder the sharing and updating of DNA sequence screening tools. These models would also require the development of detailed resources and databases containing functions of concern, which could pose an information hazard and so should be developed responsibly. An additional solution to this problem could be for DNA synthesis companies to require certifiable origins—which detail the inputs used to design an output—for orders generated by AI biodesign tools (see box 8).

Several experts highlighted the need to expand customer screening practices to parts of the life science supply chain beyond DNA providers.

Expanding Customer Screening

An important part of protecting the digital-physical interface for biology is screening customers to ensure that they have a legitimate use for life science products and services. As mentioned earlier, many DNA providers conduct customer screening as part of their biosecurity oversight,⁷² but few, if any, guidelines or requirements exist for customer screening by other vendors of life sciences products, services, or infrastructure. Furthermore, LLMs have been shown to flag opportunities for outsourcing of laboratory skills and infrastructure, for example, to contract research organizations,⁷³ which could unknowingly facilitate the development of a harmful biological agent. It will be important for these types of organizations to take biosecurity precautions.

Several experts highlighted the need to expand customer screening practices to parts of the life science supply chain beyond DNA providers.⁷⁴ Other types of providers—such as academic core facilities, cloud labs, and contract research organizations—could also adopt customer screening practices. This would help ensure that they do not provide equipment, tools, or services to illegitimate users who may wish to cause harm. Some experts recommended identifying which vendors provide the materials and equipment most needed for the development of dangerous biological agents and focusing particular attention on those vendors. A few also mentioned strengthening frameworks for sharing of materials and products among researchers to ensure that products purchased for legitimate purposes were not obtained and misused by third parties.

Expanding the number and type of institutions and vendors expected to conduct customer screening will be challenging. Customer screening by DNA providers, for example, is not universal, and current methods are burdensome, inconsistent, inefficient, and ad hoc. To solve this problem, a few experts pointed to methods used by other sectors that have implemented “Know Your Customer” approaches that could be adapted for the life sciences.⁷⁵ Others suggested a centralized customer verification framework that would give consumers credentials that they could take to a range of life sciences and AI model providers.⁷⁶ Centralizing screening would allow a single organization to perform effective screening rather than depending on many providers to do so independently. Furthermore, a centralized system also could make it possible to track a constellation of behaviors and purchases made using provided credentials, providing an opportunity to identify concerning patterns indicative of malicious intent. For example, repeated attempts to evade DNA synthesis screening could be logged, allowing for flagging of penetration testing done in attempts to access restricted materials. In the future, AI models could be developed to automate many aspects of customer screening.

It is worth noting that implementing a credentialing process or centralized screening system for life sciences practitioners would require significant outreach to those communities and could face resistance in a culture that has been dedicated to expanding access to the tools of engineering biology. Although many scientists who work with pathogens or in public health may be attuned to the risks and willing to participate, those who work on broader engineering biology pursuits may not.

Advancing Pandemic Preparedness

Many experts believe that bolstering pandemic preparedness and broader public health infrastructure will be critical for reducing biosecurity risks in the future, including those that arise from the misuse of AI tools. As discussed in the previous section, many of these capabilities can be improved using AI-bio capabilities (see box 7). A few experts described this approach as an “arms race” in which those working to prevent and respond to pandemics and engineered threats will need better funding and more powerful tools than those seeking to cause harm. A few experts cautioned that the application of AI models to pandemic preparedness should be done carefully and responsibly. A mistake by an AI model used for public health could have serious health consequences, shake public confidence, and cause de-investment in these approaches. It will be important to understand the limitations of these models and include human oversight in their implementation.

Some experts also believe that the level of effort and resources required to prevent the misuse of AI-bio capabilities and to adopt a proactive stance on biosecurity, including for infrastructure, labor, and investment in public health, will be far more than the level of effort and resources that may be required to produce a harmful biological agent and release it. The level of investment by governments and others to safeguard public health is already high, but the widespread availability of AI-bio capabilities may further reinforce the need for pandemic preparedness capabilities.

Roles and Responsibilities

Reducing risks at the intersection of AI and the life sciences will necessarily involve a wide range of actors from international organizations and governments to individual AI model developers, users, and scientists.⁷⁷ Key actions identified by experts include establishing new types of coordination bodies and oversight mechanisms, bolstering governance frameworks, funding risk reduction initiatives, and pursuing many ways to develop, incentivize, and implement safeguards for AI models. Importantly, the most powerful and impactful AI models are being developed by industry and academia, which have their own strengths, weaknesses, and constraints (box 9). These groups will play important roles in the oversight of these technologies, alongside governments, funders, and other actors.

Many experts emphasized the need for an all-hands-on-deck approach that incorporates a geographically diverse set of stakeholders, including those in North America, Europe, Africa, and Asia. A few experts believe that a formal international body should be established, along the same lines as the International Atomic Energy Agency, to focus on safety issues related to AI. However, others believe that AI is not well suited to formal, central oversight of this type because it is changing too rapidly and because it will intersect with too many different sectors, types of risks, and governance bodies and tools.

Governments

Experts saw several different roles for governments in developing guardrails for AI models, providing incentives for AI model developers to adopt good practices, bolstering biosecurity frameworks at the digital-physical interface for biology, and supporting programs for improved pandemic preparedness and response. Many experts believe that governments could develop oversight for AI model developers that might include regulations, a licensing system for model developers, or guidance. Some experts proposed establishing oversight for AI models that meet specific thresholds and criteria, for example, for LLMs, being trained on a certain number of graphics processing units (GPUs) or containing a certain number of parameters. Development of such models could be restricted, or their dissemination could be regulated through export controls. Policies could also be established that require or recommend that model developers implement good biosecurity practices. Requirements for good biosecurity practices might include, among others, that each model include technical safeguards to prevent misuse, undergo rigorous evaluation for potential misuse, or incorporate staged safety checks during development (as outlined earlier, under Guardrails for AI Models). One possibility is that governments could implement oversight by requiring cloud computing vendors to include good biosecurity practices in contractual agreements with their customers.

Many experts highlighted the need for governments to support and improve coordination of broader efforts to reduce biosecurity risks from AI-bio capabilities. This could include establishing intergovernmental cooperation or secure fora that include non-governmental stakeholders to discuss risks and options for risk reduction. Evaluation and red-teaming efforts for AI models currently are ad hoc and inconsistent, and governments could help establish or fund third-party evaluations to improve standardization and to help developers who lack the expertise or capabilities to evaluate

their models. Several experts also believe that governments should provide funding for coordination efforts and for research programs to develop resources needed to support risk reduction activities.

As mentioned earlier, many experts believe that it will be critical to improve security at the interface where digital designs become biological reality. Governments already play an important role in DNA synthesis screening and could work to strengthen these frameworks by expanding requirements that providers of synthetic DNA conduct sequence and customer screening. Experts also highlighted the need to develop improved, AI-enabled DNA synthesis screening tools that will keep pace with AI protein design tools. These tools will require significant funding; experts believe that government funding would be needed because commercial incentives for such tools are lacking. One expert suggested that a U.S. national laboratory should support these types of efforts.

AI Model Developers

Nearly all experts see a key role for developers in creating models responsibly and overseeing how these tools are used. Leading LLM developers have already made commitments to invest in the safety and security of their models,⁷⁸ including to reduce biosecurity risks, and they could incorporate several ideas for guardrails discussed earlier. A few experts also believe that AI model developers could strengthen cybersecurity to better protect their models from tampering or theft, establish internal whistleblower protections for individuals who raise safety or security concerns, and implement other practices to reduce the potential for misuse of AI models. Many experts also highlighted a need for model developers to work collaboratively with other developers and outside experts to develop best practices and establish norms for responsible behavior. Some LLM developers have already begun these types of efforts, and these should be coordinated and expanded.⁷⁹

Many experts believe that developers of AI biodesign tools hold some responsibility for how their tools are used or misused, but potential safeguards, controls, and opportunities for oversight are underdeveloped. Model developers may need to work with funders and broader academic communities to better define their role.

Non-governmental Funders

Non-governmental funders might play several roles in reducing risks related to the intersection of AI and the life sciences. Funders, particularly those who fund academic life sciences research, have the opportunity to shape how AI biodesign tools are developed and the types of guardrails that are incorporated, as many developers of AI models in academia lack awareness, incentives, and resources for implementing safeguards. Funders could also require evaluations of the research that they fund to determine whether it will result in AI models with the potential for misuse or will generate data that could contribute to information hazards. A few experts also saw a key role for non-governmental, philanthropic funding to support international, collaborative efforts to develop resources, best practices, and durable norms for responsible AI model development and dissemination.

Other Key Stakeholders

In addition to governments, AI model developers, and non-governmental funders, experts mentioned potential roles for a variety of other actors in reducing risks from AI-bio capabilities:

- **Cloud computing vendors** can implement requirements for their use in the development of AI models and could monitor usage of their resources by AI model developers.
 - **Insurers** can evaluate risks and determine whether and how the potential for misuse of AI models might affect liability insurance, which may contribute to the creation of effective incentives to implement safeguards.
 - **Legal experts** can evaluate liability and legal frameworks to determine how they intersect with AI-bio capabilities and implementation of guardrails.
 - **Institutional review bodies** can require evaluations of AI models for their potential for misuse and ask about appropriate guardrails for those that pose risks.
 - **Publishers and conferences** can evaluate whether new AI models should be published in full or whether a less open approach should be taken.
 - **Civil society** can convene multidisciplinary groups to develop resources, best practices, and approaches for reducing risks.
- **DNA providers** can bolster DNA synthesis screening activities and work collaboratively on efforts to develop improved screening mechanisms.
 - **Cloud labs, contract research organizations,** and other life sciences vendors can implement customer screening practices and more closely monitor their orders and requested services.



Recommendations: A Proposed Path Forward for Governance of AI-Bio Capabilities

The application of AI to engineering living systems will have far-reaching implications that include major potential benefits across many types of applications—such as the development of vaccines and therapeutics, broader advances in pandemic preparedness capabilities, and more fundamental advances in human health and beyond. However, these same technologies can also be misused to cause a wide range of harms, potentially including a global biological catastrophe. The rapid pace of AI advances coupled with accelerating developments in modern bioscience and biotechnology requires a radically new approach and a layered defense to reduce associated emerging biological risks. Effective governance approaches will require focused engagement by governments, AI model developers, the scientific community, non-governmental biosecurity organizations, funders, and international fora.

The findings of this report, as noted earlier, are based on interviews and engagement with a wide range of experts in AI, the life sciences, biosecurity and pandemic preparedness, and other key areas. The recommendations provided here build on these findings but were developed by the authors alone and do not necessarily reflect the views of the experts who participated in this project.



Establish an international “AI-Bio Forum” to develop AI model guardrails that reduce biological risks

- The Forum should serve as a venue for developing and sharing best practices for implementing effective AI-bio guardrails, identifying emerging biological risks associated with ongoing AI advances, and developing shared resources to manage these risks. It should inform efforts by AI model developers in industry and academia, governments, and the broader biosecurity community, and it should establish global norms for biosecurity best practices in these communities.
- Regular meetings of the Forum should provide opportunities to raise concerns, evaluate new ideas, and develop solutions on an ongoing basis.
- The Forum should be composed of key stakeholders and experts, including AI model developers in industry and academia and biosecurity experts within government and civil society, and it should act in concert with other initiatives focused on governance of AI more broadly.
- The Forum should develop a strategy for managing potential information hazards and confidential information associated with this work.

Develop a radically new, more agile approach to national governance of AI-bio capabilities

- To address emerging risks associated with rapidly advancing AI-bio capabilities, which can be difficult to anticipate, national governments should establish agile and adaptive governance approaches that can monitor AI technology developments and associated biological risks, incorporate private sector input, and rapidly adjust policy. Traditional regulatory oversight mechanisms are not equipped to match the exponential rate of change in this field, and many opportunities for risk reduction will depend on implementation by AI model developers in industry and academia. Government policymakers should explore

innovative approaches, such as dramatically streamlining rule-making procedures; rapidly exchanging information or co-developing policy with non-governmental AI experts; or explicitly empowering agile, non-governmental bodies that are working to develop and implement AI guardrails and other biological risk reduction measures.

- Governments should plan to try multiple types of approaches because some innovative governance ideas could fail. In addition, governments should incorporate sunset provisions for experimental governance bodies or processes, proactively evaluate successes and limitations, and update approaches based on lessons learned.

Implement promising AI model guardrails at scale

AI model developers should implement the most promising already developed guardrails that reduce biological risks without unduly limiting beneficial uses. They should collaborate with other entities, including the AI-Bio Forum described above, to establish best practices and develop resources to support broader implementation. Governments, biosecurity organizations, and others should explore opportunities to scale up these solutions nationally and internationally, through funding, regulations, and other incentives for adoption. Existing guardrails that should be broadly implemented include the following:

- **AI model evaluations**, including red-teaming, to preemptively identify and characterize biosecurity risks. Evaluations should begin before a model is widely available and should be conducted by multidisciplinary teams with expertise in AI, biosecurity, and microbiology.
- **Methods for AI model users to proactively report hazards**. Model developers should establish ways for users to report when a model has provided potentially harmful biological information. These reports should contribute to ongoing efforts to evaluate and update models with improved safeguards even after the model is widely available.
- **Technical safeguards to limit harmful outputs from AI models**. The state of the art for these safeguards is likely to change over time. Current promising approaches include training models to refuse to engage on particular topics or requiring models to provide outputs based on a “constitution” or set of rules determined by the developer. These should be evaluated and updated on an ongoing basis as models advance.
- **Access controls for AI models**. A promising approach for many types of models is the use of APIs that allow users to provide inputs and receive outputs without access to the underlying model. Maintaining control of a model ensures that built-in technical safeguards are not removed and provides opportunities for ensuring user legitimacy and detecting any potentially malicious or accidental misuse by users.

Pursue an ambitious research agenda to explore additional AI guardrail options for which open questions remain

AI model developers should work with biosecurity experts in government and civil society to explore additional options for AI model guardrails on an ongoing basis, experimenting with new approaches, and working to address key open questions and potential barriers to implementation. Priority areas for exploration include the following:
















- **Controlling access to AI biodesign tools.** Many strategies for safeguarding AI models depend on managing, overseeing, and limiting access, and such strategies should be widely adopted. Many LLM developers already employ APIs to maintain control of their models; however, some LLMs and many current AI biodesign tools are open-source resources. Biodesign tools are often developed by academic scientists in collaborative groups, who consider open sharing of resources an important norm for scientific advancement. Funders, publishers, and other key players should work with biosecurity experts and the academic community to reevaluate open-source norms and publication requirements for some types of AI biodesign tools. Key open questions include:
 - » Should access to some types of biodesign tools be limited to legitimate users? What types of models?
 - » How would legitimate users be verified?
 - » To what extent would limiting access prevent beneficial uses?
 - » Are there additional barriers to implementing access controls for biodesign tools (e.g., funding, infrastructure, or know-how among model developers)? How should these be overcome?
- **Managing access to computational resources needed to train models.** Because significant computational infrastructure is currently required to develop the largest, most advanced AI models, controlling access—for example, to cloud computing resources held by private companies—could help ensure that such models are developed with appropriate safeguards. However, available computational resources continue to expand rapidly, and there are strong incentives to reduce the amount of computational power needed for these models. Key open questions include:
 - » Will managed access to computational resources provide a meaningful chokepoint for model development in the future, given the rapid decline of computational power required to develop new AI models?
 - » What types of incentives would effectively ensure that vendors of cloud computing and other services enforce requirements for use of their resources?
 - » What types of biosecurity safeguards or AI model features should be required?

- **Managing access to data needed to train models.** It is possible that limiting the availability of some types of data from being used to train AI models could reduce biological risks. However, there are many potential benefits and drawbacks to this approach that depend on the types of data in question. For example, removing publicly available pathogen genome data from the Internet may be infeasible and, if pursued, could cause more harm than good by undermining important, beneficial work, such as bioscience research and biosurveillance. It may be more feasible and effective to manage access to databases that are currently privately held because of intellectual property or privacy protection needs, such as private databases linking protein structure to function or databases that include patient medical data. Key open questions include:
 - » Are there specific types of data that should not be used or should be used in limited ways for incorporation into AI models? What types of data? For what types of models?
 - » How will legitimate model development that uses restricted data be verified?

Strengthen biosecurity controls at the interface between digital design tools and physical biological systems

- Tool developers in industry, academia and non-governmental organizations should develop new AI tools to strengthen DNA sequence screening approaches to capture novel threats and improve the robustness of current approaches.
- Governments, international bodies, and others should work to strengthen DNA synthesis screening frameworks. This work should include improving incentives for DNA providers and others to conduct sequence screening and customer screening through establishment of regulations, funding requirements, financial support for DNA providers that comply, provision of resources to make screening easier, and support for international bodies that support DNA synthesis screening practices such as the International Biosecurity and Biosafety Initiative for Science.
- Governments and other key players should expand requirements and incentives for customer screening to a wide range of providers of life sciences products, infrastructure, and services, including cloud labs, contract research organizations, and academic core facilities. This could include support for a third-party verification system for life sciences customers.

RECOMMENDATIONS	RESPONSIBILITY					
	National Governments	AI Model Developers	Life Science Research Community	Biosecurity Organizations	Funders	International AI-Bio Forum
✔ indicates primary responsibility						
Develop and implement AI model guardrails. (continued)						
Support the scale-up of guardrail solutions nationally and internationally.	✔	●	●	●	●	●
Explore additional options for guardrail development.	●	●	●			✔
<i>Develop additional or more effective guardrails, particularly for biodesign tools.</i>		✔	●			●
<i>Create mechanisms and principles for responsibly and equitably controlling access to AI models.</i>	●	✔	●	●		●
<i>Explore if and how to manage access to large computational resources.</i>	✔		●			●
<i>Explore if and how to manage access to data with potential for misuse.</i>	●		●	✔		●
Develop radically new national governance approaches.						
Establish agile and adaptive governance approaches that can monitor developments in AI tools and biological risks.	✔					●
Try multiple governance approaches and evaluate their effectiveness.	✔					●

RECOMMENDATIONS	RESPONSIBILITY					
 indicates primary responsibility						
	National Governments	AI Model Developers	Life Science Research Community	Biosecurity Organizations	Funders	International AI-Bio Forum
Strengthen biosecurity controls at the interface between digital design tools and physical biological systems.						
Develop new AI tools to strengthen DNA sequence screening.						
Work together to strengthen DNA synthesis screening frameworks.						
<i>Improve incentives for DNA providers to conduct biosecurity screening of customers and DNA sequences.</i>						
Expand requirements and incentives for customer screening to a wide range of providers of life sciences products and services.						
Use AI tools to build next-generation pandemic preparedness capabilities.						
Increase investment in pandemic preparedness and response, including the development of AI-bio capabilities.						

Conclusion

The convergence of AI and the life sciences marks a new era for biosecurity and offers tremendous potential benefits, including for pandemic preparedness and response. Yet, these rapidly developing capabilities also shift the biological risk landscape in ways that are difficult to predict and have the potential to cause a global biological catastrophe. The recommendations in this report provide a proposed path forward for taking action to address biological risks associated with rapid advances in AI-bio capabilities. Effectively implementing them will require creativity, agility, and sustained cycles of experimentation, learning, and refinement.

The world faces significant uncertainty about the future of AI and the life sciences, but it is clear that addressing these risks requires urgent action, unprecedented collaboration, a layered defense, and international engagement. Taking a proactive approach will help policymakers and others anticipate future technological advances on the horizon, address risks before they fully materialize, and ultimately foster a safer and more secure future.

Appendix A: Participants

Ms. Tessa Alexanian

Ending Bioweapons Fellow
The Council on Strategic Risks

Dr. Sion Bayliss

Research Fellow
University of Bristol

Dr. Rocco Casagrande

Managing Director
Gryphon Scientific

Dr. Lauren Cowley

Senior Lecturer, Milner Centre for Evolution
University of Bath

Dr. James Diggans

Distinguished Scientist, Bioinformatics and Biosecurity
Twist Bioscience

Dr. Kevin Esvelt

Director, Sculpting Evolution Group
MIT Media Lab

Dr. Rob Fergus

Research Director
Google DeepMind

Dr. Michal Galdzicki

Data Czar
Arzeda

Dr. John Glass

Professor and Leader, Synthetic Biology Group
J. Craig Venter Institute

Dr. Logan Graham

Member of Technical Staff
Anthropic

Dr. Nathan Hillson

Department Head of BioDesign, Biological Systems and Engineering Division

Lawrence Berkeley National Laboratory

Dr. Stefan A. Hoffmann

Research Associate, Manchester Institute of Biotechnology
University of Manchester

Dr. John Lees

Group Leader
European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)

Dr. Alan Lowe

Associate Professor and Turing Fellow
University College London / Alan Turing Institute

Dr. Becky Mackelprang

Associate Director for Security Programs
Engineering Biology Research Consortium

Dr. Jason Matheny

Chief Executive Officer
RAND Corporation

Dr. Greg McKelvey

Assistant Director for Biosecurity
U.S. Office of Science and Technology Policy

Dr. Chuck Merryman

Vice President of Biology
ThinkingNode Life Science

Dr. Michael Montague

Senior Scholar and Research Scientist, Center for Health Security

Johns Hopkins University

Dr. Sella Nevo

Senior Information Scientist
RAND Corporation

Ms. Antonia Paterson

Science Manager, Responsible Development and Innovation
Google DeepMind

Dr. Ryan Ritterson

Executive Vice President of Research
Gryphon Scientific

Mr. Jonas Sandbrink

Researcher in Biosecurity
Oxford University

Dr. Clara Schoeder

Research Group Leader, Institute of Drug Discovery
Leipzig University

Dr. Reed Shabman

Deputy Director, Office of Data Science and Emerging Technologies
U.S. National Institute of Allergy and Infectious Diseases

Dr. Sarah Shoker

Research Scientist
OpenAI

Dr. Lynda Stuart

Executive Director, Institute for Protein Design
University of Washington

Appendix B: Examples of AI Models

These tables are not exhaustive, but provide examples of existing AI models, their characteristics, and whether or not they are openly available. Information listed in these tables was obtained from publicly available sources.

TABLE B.1: LARGE LANGUAGE MODELS

MODEL	DEVELOPER	PARAMETERS	OPEN SOURCE?
Natural language LLMs and their applications			
GPT-4	OpenAI	Unspecified, likely >1 trillion	No
PaLM	Google	340 billion	No
MT-NLG	NVIDIA	540 billion	No
GPT-3	OpenAI	175 billion	No
Claude	Anthropic	Unspecified	No
Pi	Inflection	Unspecified	No
LLaMA2	Meta	65 billion	Yes
MPT-30B	MosaicML	30 billion	Yes
Science-specific LLMs and their applications			
BioGPT	Microsoft	347 million	Yes
Elicit	Ought	Unspecified	No
scite	scite	Unspecified	No

TABLE B.2: BIODESIGN TOOLS

MODEL	DESCRIPTION	CITATIONS IN GOOGLE SCHOLAR	OPEN SOURCE?
Protein design			
ProteinBERT	Protein language model	199	Yes
RoseTTAFold	Protein structure prediction	2,585	Yes
AlphaFold-2	Protein structure prediction	15,717	No - requires API

MODEL	DESCRIPTION	CITATIONS IN GOOGLE SCHOLAR	OPEN SOURCE?
Protein design (continued)			
ProteinMPNN	Protein structure prediction	274	Yes
ESM-2, ESMFold	Protein structure prediction	407	Yes
RFDiffusion	Protein design	50	Yes
xTrimopGLM	Performs 15 design and prediction tasks on protein sequences	3	No
Design of DNA, biological circuits, and cells			
DNABERT-2	Foundation model for DNA sequences	267	Yes
DeepCRISPR	Design of CRISPR guide RNA	322	Yes
Enformer	Predicts the impact of genetic variants on gene expression	356	Yes
Sei / DeepSEA	Predicts the impact of genetic variants on gene expression	1,927	Yes
ExpressionGAN	Generates DNA sequences to control the expression of proteins	1,927	Yes
DeepMEL	Generates DNA sequences to control the expression of proteins	23	Yes
DeepBind	Predicts the DNA binding specificities of proteins	2,671	Yes
Benchling	DNA sequence design and editing platform	(no publication)	No
Cello 2.0	Genetic circuit design	44	Yes
novoStoic	Biological pathway design	82	Yes
RetroPath2	Biological pathway design	190	Yes
Automatic Recommendation Tool (ART)	Recommends changes in design-build-test-learn cycles to optimize metabolic engineering	140	Non-commercial and commercial licenses

TABLE B.3: AUTOMATED SCIENCE

TOOL	AUTOMATION STEP	DESCRIPTION	OPEN SOURCE?
ResearchRabbit	Search literature	Recommends relevant research papers for a literature search	No
MineTheGap	Generate hypotheses	Identifies promising gaps in the scientific literature	No
xT SAAM	Design experiments	Designs experiments for additive manufacturing	No
Emerald Cloud Lab	Perform experiments	Automates laboratory work	Symbolic Lab Language is open source for research use
Opentrons	Perform experiments	Uses laboratory robotics	Yes
Microsoft Copilot	Write software	Writes software collaboratively with a user	No
INDRA	Generate hypotheses, interpret results	Resource for representing and learning from scientific knowledge	Yes
Adam	Full cycle	Automated scientist for gene function discovery	Some publicly available elements
Eve	Full cycle	Automated scientist for drug discovery	Yes

About the Authors

Sarah R. Carter, Ph.D.

Principal, Science Policy Consulting LLC

Dr. Sarah R. Carter is the principal at Science Policy Consulting LLC. For more than 12 years, she has focused on advanced biotechnology tools and capabilities, the bioeconomy, biosecurity screening frameworks, and international norms for biosecurity. In recent years, she has been a senior consultant in support of NTI on projects related to the development of an international common mechanism for DNA synthesis screening and the implications of AI for biosecurity. She is also a senior fellow at the Federation of American Scientists and has worked with other non-governmental organizations, companies, academic institutions, and U.S. government agencies. Previously, she worked in the Policy Center of the J. Craig Venter Institute and at the White House Office of Science and Technology Policy. She is a former AAAS S&T Policy Fellow and a former Mirzayan S&T Fellow of the National Academies. She earned her Ph.D. from the University of California, San Francisco, and her bachelor's degree from Duke University.

Nicole E. Wheeler, Ph.D.

Turing Fellow, The University of Birmingham

Dr. Wheeler is a Turing Fellow who runs a research group at the University of Birmingham. She has a background in biochemistry and microbial genomics, and experience in developing machine learning methods for predicting the effects of genetic variation on the virulence of pathogens. She has provided expertise on bioinformatics and machine learning for genomic pathogen surveillance for several international programs, and her group develops novel computational methods for flagging emerging infectious disease threats, managing the spread of infectious diseases, and safeguarding emerging capabilities at the interface of AI, biosecurity and synthetic biology. She is also actively involved in public outreach and the development of governance frameworks to ensure the safe and responsible development of biotechnologies.

Sabrina Chwalek

Technical Consultant, Global Biological Policy and Programs

Sabrina Chwalek is a technical consultant for Global Biological Policy and Programs at NTI. Chwalek is a rising senior at Brown University, studying computer science with a focus on artificial intelligence and machine learning. Previously, Chwalek worked as a research assistant for the Horizon Institute for Public Service, where she contributed to their efforts to map the biosecurity landscape in U.S. policy and support the next generation of biosecurity professionals. She also worked for a non-profit focused on promoting the development of safe AI.

Christopher R. Isaac, M.Sc.

Program Officer, Global Biological Policy and Programs

Mr. Christopher Isaac is a program officer for Global Biological Policy and Programs at NTI. Isaac has been involved with synthetic biology through the Internationally Genetically Engineered Machines (iGEM) Competition since the start of his scientific career and brings with him a mixture of skills in policy, biochemistry, and programming. Isaac holds a B.Sc. in biological sciences with a minor in philosophy and a M.Sc. in biochemistry (bioinformatics) from the University of Lethbridge. He is an alumnus of the Emerging Leaders in Biosecurity Fellowship at the Johns Hopkins Center for Health Security, a member of the iGEM Safety and Security Committee, and a Schmidt Futures International Strategy Forum Fellow.

Jaime M. Yassif, Ph.D.

Vice President, Global Biological Policy and Programs

Dr. Jaime Yassif has 20 years of experience working at the interface of science, technology, public health, and international security within government and civil society. As NTI’s vice president for Global Biological Policy and Programs, she oversees the organization’s work to reduce catastrophic biological risks, strengthen biosecurity and pandemic preparedness, and drive progress in advancing global health security. Yassif previously served as a program officer at Open Philanthropy, where she led the Biosecurity and Pandemic Preparedness initiative, recommending and managing approximately \$40 million in biosecurity grants, which rebuilt the field and supported work in several key areas. Prior to this, she served as a science and technology policy advisor at the U.S. Department of Defense and worked on the Global Health Security Agenda at the U.S. Department of Health and Human Services. Dr. Yassif holds a Ph.D. in biophysics from the University of California Berkeley, an M.A. in science and security from the King’s College London War Studies Department, and a B.A. in biology from Swarthmore College.

Endnotes

- ¹ Bo Chen et al., "xTrimoPGLM: Unified 100B-Scale Pre-trained Transformer for Deciphering the Language of Protein" (preprint, July 14, 2023), bioRxiv, <https://doi.org/10.1101/2023.07.05.547496>.
- ² Hanchen Wang et al., "Scientific Discovery in the Age of Artificial Intelligence," *Nature* 620, no. 7972 (2023): 47–60, <https://doi.org/10.1038/s41586-023-06221-2>.
- ³ For example, see <https://biolm.ai/ui/home/>.
- ⁴ NTI, "Biosecurity and Risk Reduction Initiative," <https://www.nti.org/about/programs-projects/project/fostering-biosecurity-innovation-and-risk-reduction/>.
- ⁵ NTI, "Common Mechanism to Prevent Illicit Gene Synthesis," March 22, 2019, <https://www.nti.org/analysis/articles/common-mechanism-prevent-illicit-gene-synthesis/>.
Sarah R. Carter, Jaime M. Yassif, and Christopher R. Isaac, *Benchtop DNA Synthesis Devices: Capabilities, Biosecurity Implications, and Governance* (Washington, DC: NTI, 2023).
NTI, "NTI and World Economic Forum Release New Report on DNA Synthesis Technologies," January 9, 2020, <https://www.nti.org/news/nti-and-world-economic-forum-release-new-report-dna-synthesis-technologies/>.
- ⁶ NTI, "International Biosecurity and Biosafety Initiative for Science (IBBIS)," October 5, 2022, <https://www.nti.org/about/programs-projects/project/international-biosafety-and-biosecurity-initiative-for-science-ibbis/>.
- ⁷ Arul Siromoney and Rani Siromoney, "A Machine Learning System for Identifying Transmembrane Domains from Amino Acid Sequences," *Sadhana* 21 (1996): 317–25, <https://link.springer.com/article/10.1007/BF02745526>.
Richard Fox, "Directed Molecular Evolution by Machine Learning and the Influence of Nonlinear Interactions," *Journal of Theoretical Biology* 234, no. 2 (2005): 187–99, <https://www.sciencedirect.com/science/article/pii/S0022519304005697>.
D. B. Kell, "Metabolomics, Machine Learning and Modelling: Towards an Understanding of the Language of Cells," 33, no. 3 (June 2005): 520–24, <https://portlandpress.com/biochemsoctrans/article/33/3/520/82871/Metabolomics-machine-learning-and-modelling>.
- ⁸ OpenAI, "Introducing ChatGPT," November 30, 2022, <https://openai.com/blog/chatgpt>.
- ⁹ Eric Schmidt, "This Is How AI Will Transform the Way Science Gets Done," *MIT Technology Review*, July 5, 2023, <https://www.technologyreview.com/2023/07/05/1075865/eric-schmidt-ai-will-transform-science/>.
- ¹⁰ Michael Chui et al., "The Economic Potential of Generative AI: The Next Productivity Frontier" (McKinsey & Company, New York, NY, June 14, 2023), <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier>.
- ¹¹ NVIDIA, "NVIDIA Unveils Next-Generation GH200 Grace Hopper Superchip Platform for Era of Accelerated Computing and Generative AI," press release, August 8, 2023, <https://nvidianews.nvidia.com/news/gh200-grace-hopper-superchip-with-hbm3e-memory>.
- ¹² David McCandless, Tom Evans, and Paul Barton, "The Rise and Rise of A.I. Large Language Models (LLMs)," *Information Is Beautiful*, July 27, 2023, <https://informationisbeautiful.net/visualizations/the-rise-of-generative-ai-large-language-models-llms-like-chatgpt/>.
- ¹³ Sida Peng et al., "The Impact of AI on Developer Productivity: Evidence from GitHub Copilot," ArXiv, (2023). <https://arxiv.org/pdf/2302.06590.pdf>.
Noy, Shakked, and Whitney Zhang, "Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence," *Science*, (2023). <https://doi.org/adh2586>.
Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond, "Generative AI at Work," ArXiv, (2023). <https://arxiv.org/abs/2304.11771>.
- ¹⁴ For more information, see the scite website at, <https://scite.ai/>.
For more information, see the Elicit website at, <https://elicit.org/>.
- ¹⁵ Yevgen Chebotar and Tianhe Yu, "RT-2: New Model Translates Vision and Language into Action," Google Deepmind, July 28, 2023, https://www.deepmind.com/blog/rt-2-new-model-translates-vision-and-language-into-action?utm_source=keywordblog&utm_medium=referral&utm_campaign=rt2.
Julián N. Acosta, Guido J. Falcone, Pranav Rajpurkar, and Eric J. Topol, "Multimodal Biomedical AI," *Nature Medicine* 28, no. 9 (2022): 1773–84, <https://www.nature.com/articles/s41591-022-01981-2>.
- ¹⁶ Jason Wei et al., "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models" (updated manuscript, January 10, 2023), arXiv, <https://doi.org/10.48550/arXiv.2201.11903>.
- ¹⁷ Jason Wei et al., "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models."
- ¹⁸ John Jumper et al., "Highly Accurate Protein Structure Prediction with AlphaFold," *Nature* 596 (2021): 583–89, <https://doi.org/10.1038/s41586-021-03819-2>.
Andriy Kryshchak et al., "Critical Assessment of Methods of Protein Structure Prediction (CASP)—Round XIV," *Proteins: Structure, Function, and Bioinformatics* 89, no. 12 (2021): 1607–17, <https://onlinelibrary.wiley.com/doi/10.1002/prot.26237>.
- ¹⁹ Joseph L. Watson et al., "De Novo Design of Protein Structure and Function with RFdiffusion," *Nature* 620 (2023): 1089–1100, <https://doi.org/10.1038/s41586-023-06415-8>.
- ²⁰ J. Dauparas et al., "Robust Deep Learning–Based Protein Sequence Design Using ProteinMPNN," *Science* 378, no. 6615 (2022): 49–56, <https://www.science.org/doi/10.1126/science.add2187>.

21 Ewen Callaway, "Scientists Are Using AI to Dream Up Revolutionary New Proteins," news release, Nature, September 15, 2022, <https://www.nature.com/articles/d41586-022-02947-7>.

22 Christian Jäckel, Peter Kast, and Donald Hilvert, "Protein Design by Directed Evolution," *Annual Review of Biophysics* 37 (2008): 153–73, <https://www.annualreviews.org/doi/abs/10.1146/annurev.biophys.37.032807.125832>.

23 Bo Chen et al., "xTrimoPGLM: Unified 100B-Scale Pre-trained Transformer for Deciphering the Language of Protein" (preprint, July 14, 2023), bioRxiv, <https://doi.org/10.1101/2023.07.05.547496>.

24 Bo Chen et al., "xTrimoPGLM."

25 Engineering Biology Research Consortium (EBRC), "Review Progress in the Field: An Assessment of Short-Term Milestones in EBRC's Roadmap, Engineering Biology," <https://roadmap.ebrc.org/>.

26 For more information, see the Synthetic Biology Open Language portal at <https://sbolstandard.org/>.

27 Jan Zrimec et al., "Controlling Gene Expression with Deep Generative Design of Regulatory DNA," *Nature Communications* 13, no. 1 (2022): 1–17, <https://doi.org/10.1038/s41467-022-32818-8>.

28 Zihao Chen et al., "Artificial Intelligence in Aptamer–Target Binding Prediction," *International Journal of Molecular Sciences* 22, no. 7 (2021), <https://doi.org/10.3390/ijms22073605>.

29 Christopher E. Lawson et al., "Machine Learning for Metabolic Engineering: A Review," *Metabolic Engineering* 63 (January 2021): 34–60, <https://doi.org/10.1016/j.ymben.2020.10.005>.

30 Maren Wehrs et al., "Engineering Robust Production Microbes for Large-Scale Cultivation," *Trends in Microbiology* 27, no. 6 (2019): 524–37, <https://doi.org/10.1016/j.tim.2019.01.006>.

31 Christopher J. Hartline et al., "Dynamic Control in Metabolic Engineering: Theories, Tools, and Applications," *Metabolic Engineering* 63 (January 2021): 126, <https://doi.org/10.1016/j.ymben.2020.08.015>.

32 Linfeng Zhang et al., "Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics," *Physical Review Letters* 120, no. 14 (2018): 143001, <https://link.aps.org/doi/10.1103/PhysRevLett.120.143001>.

33 Ross D. King et al., "The Automation of Science," *Science* 324, no. 5923 (2009): 85–89. [10.1126/science.1165620](https://doi.org/10.1126/science.1165620).

34 Kevin Williams et al., "Cheaper Faster Drug Development Validated by the Repositioning of Drugs against Neglected Tropical Diseases," *Journal of the Royal Society Interface* 12, no. 104 (2015), <http://doi.org/10.1098/rsif.2014.1289>.

35 Benjamin Burger et al., "A Mobile Robotic Chemist," *Nature* 583, no. 7815 (2020): 237–41, <https://www.nature.com/articles/s41586-020-2442-2>.

36 Daniil Boiko, Robert MacKnight, and Gabe Gomes, "Emergent Autonomous Scientific Research Capabilities of Large Language Models," arXiv, April 11, 2023, <https://arxiv.org/pdf/2304.05332.pdf>.

37 Gemma Conroy, "Scientists Used ChatGPT to Generate an Entire Paper from Scratch—But Is It Any Good?," news release, Nature, July 11, 2023, <https://www.nature.com/articles/d41586-023-02218-z>.

38 Vahe Tshitoyan et al., "Unsupervised Word Embeddings Capture Latent Knowledge from Materials Science Literature," *Nature* 571, no. 7763 (2019): 95–98, <https://doi.org/10.1038/s41586-019-1335-8>.

39 John P. A. Ioannidis, "Why Most Published Research Findings Are False," *PLOS Medicine* 2, no. 8 (2005): e124, <https://doi.org/10.1371/journal.pmed.0020124>.

40 Gemma Conroy, "Scientists Used ChatGPT to Generate an Entire Paper from Scratch—But Is It Any Good?," news release, Nature, July 11, 2023, <https://www.nature.com/articles/d41586-023-02218-z>.

41 Yuting Xu et al., "Deep Dive into Machine Learning Models for Protein Engineering," *Journal of Chemical Information and Modeling* 60, no. 6 (April 6, 2020): 2773–90, <https://doi.org/10.1021/acs.jcim.0c00073>.

42 Eirini Kalliamvakou, "Research: Quantifying GitHub Copilot's Impact on Developer Productivity and Happiness," GitHub, September 7, 2022, <https://github.blog/2022-09-07-research-quantifying-github-copilots-impact-on-developer-productivity-and-happiness/>.

43 Daniil Boiko, Robert MacKnight, and Gabe Gomes, "Emergent Autonomous Scientific Research Capabilities of Large Language Models," arXiv, April 11, 2023, <https://arxiv.org/pdf/2304.05332.pdf>.

44 For more information, see the INDRA website at <http://www.indra.bio/>.

45 Toby Shevlane et al., "Model Evaluation for Extreme Risks" (submitted manuscript, May 24, 2023), arXiv, <https://arxiv.org/abs/2305.15324>.

46 Emily H. Soice et al., "Can Large Language Models Democratize Access to Dual-Use Biotechnology?" (submitted manuscript, June 6, 2023), arXiv, <https://arxiv.org/abs/2306.03809>.

47 Sarah R. Carter, Jaime M. Yassif, and Christopher R. Isaac, *Benchmark DNA Synthesis Devices: Capabilities, Biosecurity Implications, and Governance* (Washington, DC: NTI, 2023), <https://www.nti.org/analysis/articles/benchmark-dna-synthesis-devices-capabilities-biosecurity-implications-and-governance/>.

48 Gavin R. Meehan et al., "Phenotyping the Virulence of SARS-CoV-2 Variants in Hamsters by Digital Pathology and Machine Learning" (preprint, August 1, 2023), bioRxiv, <https://www.biorxiv.org/content/10.1101/2023.08.01.551417v1>.

49 Jose Antonio Lanz, "Meet Chaos-GPT: An AI Tool That Seeks to Destroy Humanity," Decrypt, April 13, 2023, <https://decrypt.co/126122/meet-chaos-gpt-ai-tool-destroy-humanity>.

50 European Centre for Disease Prevention and Control, "New Tools for Public Health Experts: Outbreak Detection and Epidemiological Reports," news release, December 17, 2018, <https://www.ecdc.europa.eu/en/news-events/new-tools-public-health-experts-outbreak-detection-and-epidemiological-reports>.

- 51 Amber Barton and Caroline Colijn, "Genomic, Clinical and Immunity Data Join Forces for Public Health," *Nature Reviews Microbiology* 21, no. 639 (2023), <https://doi.org/10.1038/s41579-023-00965-4>.
- 52 Xiaodong Guo et al., "Aptamer-Based Biosensor for Detection of Mycotoxins," *Frontiers in Chemistry* 8 (2020), <https://doi.org/10.3389/fchem.2020.00195>.
- 53 Alfredo Quijano-Rubio et al., "De novo Design of Modular and Tunable Protein Biosensors," *Nature* 591 (2021), <https://www.nature.com/articles/s41586-021-03258-z>
- 54 Ethan C. Alley et al., "A Machine Learning Toolkit for Genetic Engineering Attribution to Facilitate Biosecurity," *Nature Communications* 11, no. 1 (2020): 1–12, <https://doi.org/10.1038/s41467-020-19612-0>.
- 55 Francine J. Boonekamp et al., *ACS Synthetic Biology* 9, no. 6 (May 15, 2020): 1361–75, <https://doi.org/10.1021/acssynbio.0c00045>.
- 56 Dylan Matthews, "The \$1 Billion Gamble to Ensure AI Doesn't Destroy Humanity," *Vox*, July 17, 2023, <https://www.vox.com/future-perfect/23794855/anthropic-ai-openai-claude-2>.
- 57 Yuntao Bai et al., "Constitutional AI: Harmlessness from AI Feedback," (submitted manuscript, December 15, 2022), arXiv, (2023), <https://arxiv.org/abs/2212.08073>.
- 58 John Kirchenbauer et al., "A Watermark for Large Language Models," (preprint, revised June 6, 2023), arXiv, <https://arxiv.org/abs/2301.10226>; A Watermark for Large Language Models, demo, <https://huggingface.co/spaces/tomg-group-umd/lm-watermarking>.
- 59 Toby Shevlane, "Sharing Powerful AI Models," research post, Centre for the Governance of AI, January 20, 2022, <https://www.governance.ai/post/sharing-powerful-ai-models>.
- 60 Toby Shevlane et al., "Model Evaluation for Extreme Risks" (submitted manuscript, May 24, 2023), arXiv, <https://arxiv.org/abs/2305.15324>.
- 61 AI Incident Database, <https://incidentdatabase.ai/apps/incidents/>.
- 62 OpenAI, "Announcing OpenAI's Bug Bounty Program," April 11, 2023, <https://openai.com/blog/bug-bounty-program>.
- 63 For example, see Anthropic, "Claude 2," July 11, 2023, <https://www.anthropic.com/index/claude-2>.
- 64 Erin Dumbacher, Page Stoutland, and Jason Arterburn, *Signals in the Noise: Preventing Nuclear Proliferation with Machine Learning and Publicly Available Information* (Washington, DC: NTI, 2021), <https://www.nti.org/analysis/articles/signals-in-the-noise-preventing-nuclear-proliferation-with-machine-learning-publicly-available-information/>.
- 65 Matt Wuebbli, "GeForce Is Made for Gaming, CMP Is Made to Mine," NVIDIA, February 18, 2021, <https://blogs.nvidia.com/blog/2021/02/18/geforce-cmp/>.
Yonadav Shavit, "What Does It Take to Catch a Chinchilla? Verifying Rules on Large-Scale Neural Network Training via Compute Monitoring" (revised manuscript, May 30, 2023), arXiv, <https://arxiv.org/abs/2303.11341>.
- 66 Edward J. Hu et al., "LoRA: Low-Rank Adaptation of Large Language Models" (revised manuscript, October 16, 2021), arXiv, <https://arxiv.org/abs/2106.09685>.
- 67 Dami Choi, Yonadav Shavit, and David Duvenaud, "Tools for Verifying Neural Models' Training Data" (submitted manuscript, July 2, 2023), arXiv, <https://arxiv.org/abs/2307.00682>.
- 68 Bureau of Industry and Security, "Commerce Control List: Category 1—Materials, Chemicals, Microorganisms, and Toxins," Supplement No. 1 to Part 774, Export Administration Regulations, August 18, 2023, <https://www.bis.doc.gov/index.php/documents/federal-register-notices-1/3315-ccl1-11/file>.
Export Control Order 2008, U.K. S.I. 2008/3231, December 17, 2008, accessed September 7, 2023, <https://www.legislation.gov.uk/uksi/2008/3231/contents>.
- 69 For example, International Gene Synthesis Consortium, <https://genesynthesisconsortium.org/>; Sarah R. Carter, Jaime M. Yassif, and Christopher R. Isaac, *Benchtop DNA Synthesis Devices: Capabilities, Biosecurity Implications, and Governance* (Washington, DC: NTI, 2023), <https://www.nti.org/analysis/articles/benchtop-dna-synthesis-devices-capabilities-biosecurity-implications-and-governance/>.
- 70 Administration for Strategic Preparedness and Response, "Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA," U.S. Department of Health and Human Services, <https://aspr.hhs.gov/legal/syndna/Pages/default.aspx>.
- 71 The International Biosecurity and Biosafety Initiative for Science, <https://ibbis.bio/>.
- 72 For example, the Administration for Strategic Preparedness and Response, "Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA," U.S. Department of Health and Human Services, <https://aspr.hhs.gov/legal/syndna/Pages/default.aspx>; International Gene Synthesis Consortium, "Harmonized Screening Protocol® v2.0," November 19, 2017, <https://genesynthesisconsortium.org/wp-content/uploads/IGSCHARmonizedProtocol11-21-17.pdf>.
- 73 Emily H. Soice et al., "Can Large Language Models Democratize Access to Dual-Use Biotechnology?" (submitted manuscript, June 6, 2023), arXiv, <https://arxiv.org/abs/2306.03809>.
- 74 Sarah Carter and Diane DiEuliis, "Mapping the Synthetic Biology Industry: Implications for Biosecurity," *Health Security* 17, no. 5 (2019): 403–6, <https://doi.org/10.1089/hs.2019.0078>.
- 75 Dow Jones, "Understanding the Steps of a 'Know Your Customer' Process," Risk & Compliance Glossary, <https://www.dowjones.com/professional/risk/glossary/know-your-customer/>.
- 76 Global Alliance for Genomics & Health, "Passports," <https://www.ga4gh.org/product/ga4gh-passports/>.
- 77 Ian Bremmer and Mustafa Suleyman, "The AI Power Paradox: Can States Learn to Govern Artificial Intelligence—Before It's Too Late?" *Foreign Affairs*, August 16, 2023, <https://www.foreignaffairs.com/world/artificial-intelligence-power-paradox>.

⁷⁸ Google, “A New Partnership to Promote Responsible AI,” July 26, 2023, <https://blog.google/outreach-initiatives/public-policy/google-microsoft-openai-anthropic-frontier-model-forum>.

White House, “Fact Sheet: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI,” press release, 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/>.

⁷⁹ Google, “A New Partnership to Promote Responsible AI,” July 26, 2023, <https://blog.google/outreach-initiatives/public-policy/google-microsoft-openai-anthropic-frontier-model-forum>.

NTI:bio

Read more about NTI | bio's work to strengthen
biotechnology governance and biosecurity

www.nti.org/bio





1776 Eye Street, NW • Suite 600 • Washington, DC 20006 • @NTI_WMD • www.nti.org