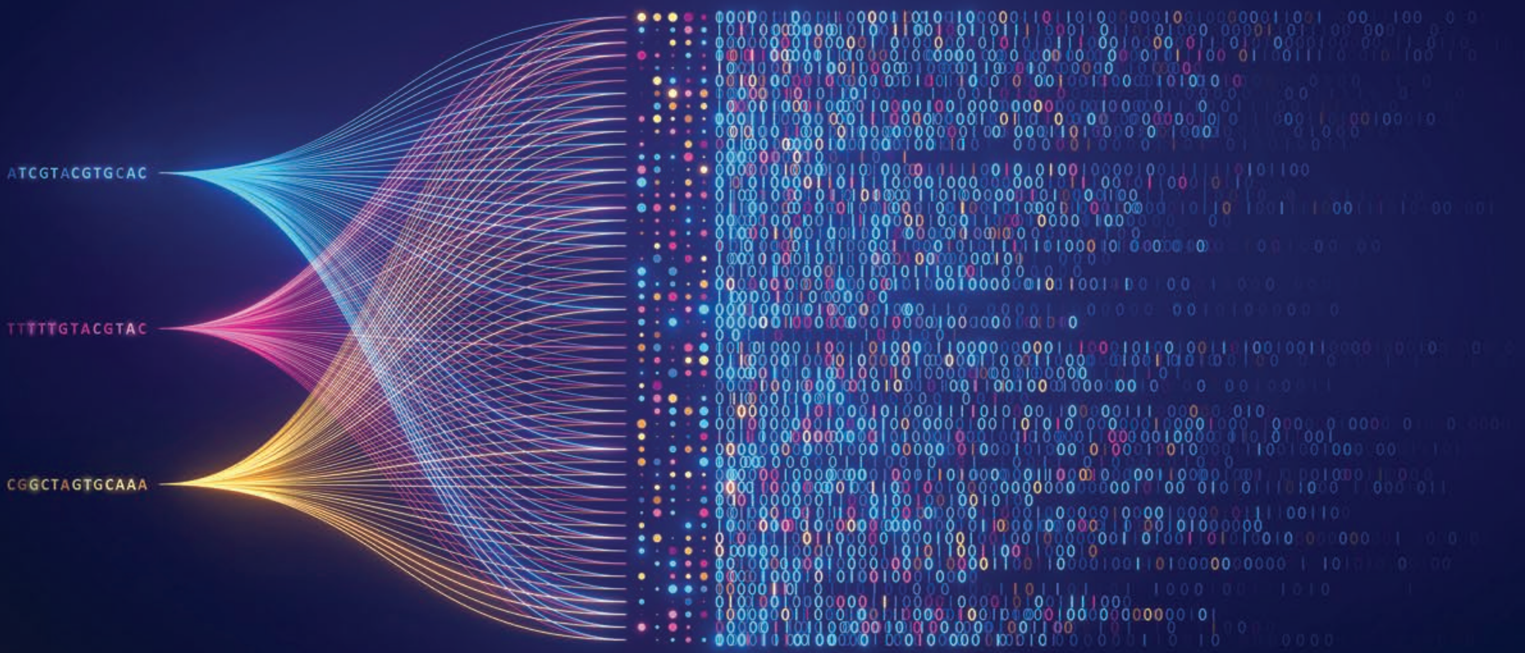

The Convergence of Artificial Intelligence and the Life Sciences:

Safeguarding Technology, Rethinking Governance, and Preventing Catastrophe

OCTOBER 2023

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. 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.

> **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.**

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 and carrying out experiments using laboratory robotics, analyzing results, and forming updated hypotheses. These capabilities have the potential to speed up the scientific process in a number of ways, including by scaling up and outsourcing work, reducing the number of experiments that need to be performed, and removing time constraints and errors inherent in human labor. Although some chemistry research has been completely automated in this way, automation of work involving living systems has proven more challenging, and only parts of the process can currently be automated. It is unclear if or when more AI advances will make it possible to fully automate systems to support life science research.

Biosecurity Implications

The level of risk that AI-bio capabilities may pose for biosecurity, and which tools pose the greatest risk, are the subject of some disagreement within the life sciences and AI expert communities. However, experts broadly agree that while offering benefits, these capabilities will enable a wide range of users to engineer biology and are therefore also likely to pose some biosecurity risks. Without appropriate safeguards, a malicious actor with little expertise in biology could use LLMs to become familiar with pathogens that could be used to cause catastrophic harm. LLMs also could provide access to publicly available information on how to obtain such agents and locate relevant equipment, facilities, and opportunities for outsourcing. However, significant barriers would remain, including funding, infrastructure, access to materials, and tacit knowledge of how to successfully work in a laboratory. Furthermore, current LLMs are unlikely to generate toxin or pathogen designs that are not already described in the public literature, and it is likely they will only be able to do this in the future by incorporating more specialized AI biodesign tools.

AI biodesign tools, in contrast, may be able to generate toxin or pathogen designs that are not found in nature. Some of these could be more harmful than versions that may evolve naturally. Using these types of tools currently requires some expertise, but they will likely become easier to use in the near future. Significant uncertainty remains about if or when AI biodesign tools might be able to generate reliable designs for biological agents that are as complex as pathogens, and there are major barriers to converting a digital design into biological reality, including generating, testing, and deploying these agents.

Notwithstanding the risks, AI-bio capabilities will also benefit society and bolster biosecurity and pandemic preparedness. In addition to broadly enabling scientific progress, AI models are already aiding pathogen biosurveillance systems, the development of medical countermeasures, and other aspects of pandemic preparedness and response. In the future, AI could improve biosecurity and pandemic preparedness in a wider variety of ways, including by predicting supply-chain shortages during public health emergencies and detecting unusual or potentially dangerous behaviors among AI model users or life science practitioners.

> **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.**

Opportunities for Risk Reduction

Reducing biosecurity risks associated with AI-bio capabilities without unduly hindering their beneficial uses is paramount, and a number of approaches are possible. A successful, layered defense will include several key components: establishing guardrails for AI-bio capabilities, strengthening controls at the interface where digital designs become physical biological systems, and bolstering pandemic preparedness. This report focuses primarily on building stronger guardrails for AI-bio capabilities because this area 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.

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

Authors

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

NTI | bio

OCTOBER 2023

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.

Read the full report

NTI
BUILDING A SAFER WORLD

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