NSCEB
National Security Commission on Emerging Biotechnology

White Paper 3: Risks of AIxBio

NSCEB White Paper Series on AIxBio

January 2024

Policymakers should carefully consider the nuances of the risks associated with artificial intelligence (AI) models that interact with biotechnology (AIxBio) when considering oversight mechanisms, like regulation, in order to promote safe and secure innovation without stifling advancement of the technology. This white paper provides a current snapshot of specific information related to AIxBio and risk and describes the concerns associated with different categories of AI tools.

Overview of assessing AIxBio risks

Each type of AI model and platform poses a different set of bio-related considerations and risks. This paper differentiates the risks associated with two primary AI tools: Large Language Models (LLMs) and Biological Design Tools (BDTs).[1] Popular, publicly-accessible tools such as ChatGPT (which is an LLM) have very different capabilities and risks compared with a model that is specifically built for understanding and predicting biology (such as a BDT). Policymakers can assess the risk using the following three factors:

1) The type of tool used (LLM vs. BDT)
2) The entity or person using the tool and the user's skills with biological experimentation (from amateur to expert); and
3) The intent of the user (good intentions vs. bad intentions)

The table[2] to the right summarizes some key considerations for different AI models and skills of the actor.

Large Language Models (LLMs)

Popularized by natural language programs such as ChatGPT, LLMs have varying levels of scientific capability where the models can sometimes provide very helpful scientific information and sometimes provide wildly inaccurate information.

Based on the current state of LLMs, these tools may be able to quickly assist both experts and amateurs with the collection and synthesis of biological information. This is a benefit to well-intentioned researchers who are trying to quickly learn about an area of study, better understand complex technical language, or troubleshoot an experiment. However, this capability could also provide information to a bad actor looking to create something harmful.

	LLMs	BDTs
Amateur Biotechnologist	**Advantages:** • May help compile information • Can summarize existing knowledge on pathogens and suggest known methods for development and dispersal **Risks** • Could make novice researchers believe they have the full knowledge needed for bioweapon development	**Advantages:** • Likely inaccessible to amateur lacking deep biological and computer science expertise **Risks** • None of concern currently
Expert Biotechnologist	**Advantages:** • May help compile information or provide tips to troubleshoot biological experiments • Could help brainstorm ideas for development or dispersal of new pathogens or toxins **Risks** • None of concern currently	**Advantages:** • Could decrease research time by designing more targeted experiments that quickly lead to positive results **Risks** • Could design more harmful pathogens • Could design a pathogen with enhanced properties to evade screening or detection • Small potential for unintentional design of harmful biological properties due to biases or inaccuracies in training data

At this time, LLMs do not significantly increase the risk of the creation of a bioweapon as LLMs do not provide new information or information on how to conduct biological laboratory experiments, beyond what is already available on the internet.[3,4] For example, while a bad actor may use an LLM to quickly determine what genetic mutation can increase the speed at which a virus spreads, that genetic mutation is not new information because it already is available in published papers.[5] Additionally, the process of introducing a genetic mutation into a virus is complicated and requires extensive hands-on lab training that cannot be provided by an LLM. Before LLMs become more sophisticated, it is important that programmers and developers think about early guardrails that might prevent the generation of new insights (rather than the presentation of existing, available information) that could pose new bio-related threats.

Biological Design Tools (BDTs)

While LLMs digest natural language data, BDTs digest biological data and produce biologically relevant predictions and simulations. At the moment, using BDTs requires a substantial level of understanding of both biology and computer science, whereas using LLMs does not.

Amateur users are unlikely to use a BDT, but experts who are bad actors could complement extensive scientific training with specific AI models to more effectively generate new pathogen designs, develop synthetic DNA strands that subvert screening guardrails, or improve the efficiency of bioweapon production.[6] As with any AI system, BDTs rely on the quality of their training data, and sometimes the data can have significant limitations such as a lack of completeness or unintentional biases.

Capturing intent

While the intent of an individual using an LLM or BDT is difficult to determine, it is an important consideration when discussing guardrails or safety regulations. For example, in the case of both LLMs and BDTs, it is possible that an AI model could lead a well-intentioned user to accidentally create a more harmful biological entity by producing incorrect or biased information. If the models are not properly trained on complete datasets, which may result in specific biases in their outputs, there is the possibility for accidental misuse. Future safeguards should balance the need to prevent misuse of AI models for biotechnology research while accounting for the possibility of incorrect information generated by these models.

Current gaps in risk assessments

Much of the information on AI tools in biotechnology is derived from proofs-of-concepts conducted by researchers. There are few empirical studies that quantify the actual risk of specific LLMs and BDTs leading to the production of harmful biological agents.[7] For example, BDTs have been used to digitally generate potentially risky genetic sequences, but research has yet to show if the synthesized sequences could be used to create a harmful biological agent. Future policies or industry practices could establish safe and systematic ways to assess the risk posed by AI systems like LLMs and BDTs to lead to the actual creation of harmful biological agents.[8]

Policymakers could also consider safeguards for future technologies that combine LLMs and BDTs. A future LLM chatbot could help both experts and amateur scientists program a more sophisticated BDT that generates risky protocols. While this scenario is still far from a reality, one company announced the first LLM chatbot interface to control a suite of BDTs, opening the possibility of those without computer science expertise to more easily use BDTs.[9] However, the technology is proprietary and can only be used by people who work at the company.

As indicated by the rapid advancement of the field and the nuance in current technologies, addressing the various risks that the convergence of AI and biotechnology poses requires a dynamic approach rather than a one-size-fits-all solution.[10]

Sources

1. For a more in-depth explanation of LLMs and BDTs, see the Commission's White Paper #1
2. Adopted from: https://cset.georgetown.edu/wp-content/uploads/20230047-AI-BioRisk-Extended-Explainer-FINAL-v2.pdf
3. https://foreignpolicy.com/2023/11/05/ai-artificial-intelligence-chatbot-bioweapon-virus-bacteria-genetic-engineering/
4. https://www.rand.org/pubs/research_reports/RRA2977-2.html
5. https://www.nti.org/analysis/articles/the-convergence-of-artificial-intelligence-and-the-life-sciences/
6. https://cset.georgetown.edu/publication/ai-and-biorisk-an-explainer/
7. https://www.helenabiosecurity.org/
8. https://www.schumer.senate.gov/imo/media/doc/Alexander%20Titus%20-%20Statement.pdf
9. https://ir.recursion.com/news-releases/news-release-details/recursion-unveils-lowe-drug-discovery-software-jp-morgan
10. https://www.schumer.senate.gov/imo/media/doc/Sean%20McClain.pdf

For any questions about this series of white papers, or the AIxBio work at the National Security Commission on Emerging Biotechnology, please contact us at ideas@biotech.senate.gov

Staff at the National Security Commission on Emerging Biotechnology authored this paper with input from the expert Commissioners. The content and recommendations of this white paper do not necessarily represent positions officially adopted by the Commission.