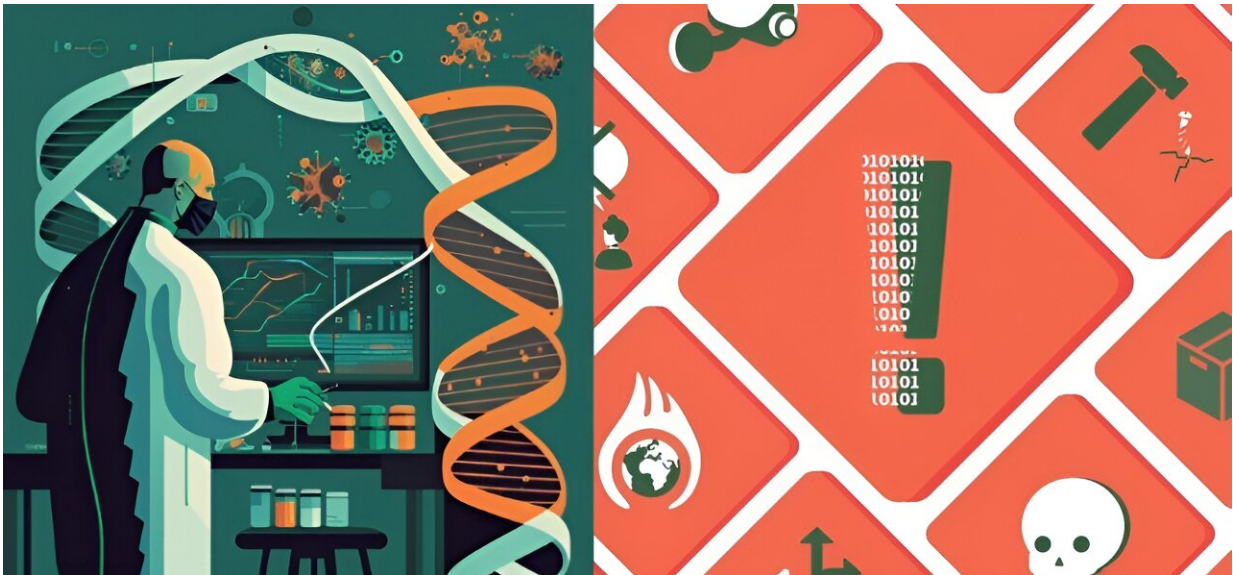


Improving safety of AI research for engineering biology

July 8 2024, by Laura Thomas



Proactively exploring data hazards in synthetic biology research. Credit: Thomas Gorochowski/University of Bristol

Hazards posed by using data-centric methods to engineer biology have been identified by experts at the University of Bristol with the aim of making future research safer.

The potential misuse of data-centric approaches in synthetic biology poses significant risk. The ease of access to [data science](#) tools may enable nefarious actors to develop harmful biological agents for

purposes such as bioterrorism or to disrupt [ecological systems](#) intentionally.

[The findings](#), published in *Synthetic Biology*, suggest additional Data Hazard labels that describe data related risks in area of synthetic biology.

- Uncertain accuracy of source data—The accuracy of the underlying data is not known and so its use may lead to erroneous results or introduce bias.
- Uncertain completeness of source data—Underlying data are of an uncertain completeness and have missing values that causes biased results.
- Integration of incompatible data—Data of different types and/or sources are being used together that may not be compatible with each other.
- Capable of ecological harm—This technology has the potential to cause broad ecological harm, even if used correctly.
- Potential experimental hazard—Translating technology into experimental practice can require safety precautions.

The work is the result of a collaboration between researchers from across the Bristol Center for Engineering Biology (BrisEngBio) and the Jean Golding Institute for Data Intensive Research.

Kieren Sharma, co-author and Ph.D. student working in AI for cellular modeling in the School of Engineering Mathematics and Technology said, "We're entering a transformative era where artificial intelligence and [synthetic biology](#) converge to revolutionize [biological engineering](#), accelerating the discovery of novel compounds, from life-saving pharmaceuticals to sustainable biofuels.

"Our study has uncovered potential risks associated with the specific types of data being used to train the latest systems biology models. For

instance, inconsistencies in measurements from complex and dynamic living organisms and privacy concerns that could compromise the safety of next-generation models trained on human genome data."

The project extends the work of the [Data Hazards project](#), which aims to create a clear vocabulary of the potential hazards of data science research.

Co-author and co-lead of the Data Hazards project, Dr. Nina Di Cara from the School of Psychological Science, explained, "Having a clear vocabulary of hazards makes it easier for researchers to think proactively about what the risks of their work are and to help put mitigating actions in place. It also makes communication easier for people working across fields who sometimes use different language to talk about the same issues."

To achieve these clear vocabularies, interdisciplinary collaboration is essential.

Dr. Daniel Lawson, Director of the Jean Golding Institute and Associate Professor in Data Science in the School of Mathematics noted that "As datasets grow in magnitude and ambition, increasingly sophisticated algorithms are developed to gain new insights. This complexity makes an un-siloed collaborative approach to identifying and preventing downstream harms essential."

Dr. Thomas Gorochowski, senior author and Associate Professor of Biological Engineering in the School of Biological Sciences, added, "Data science is set to revolutionize how we engineer biology to harness its unique capabilities to tackle global challenges covering the sustainable production of materials and fuels the development of innovative therapeutics.

"The extensions developed by our team will help bioengineers consider and discuss risks around data-centric approaches to their research and help ensure the huge benefits of bio-based solutions are realized in a safe way."

More information: Natalie R Zelenka et al, Data hazards in synthetic biology, *Synthetic Biology* (2024). [DOI: 10.1093/synbio/ysae010](https://doi.org/10.1093/synbio/ysae010)

Provided by University of Bristol

Citation: Improving safety of AI research for engineering biology (2024, July 8) retrieved 22 July 2024 from <https://phys.org/news/2024-07-safety-ai-biology.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.