Trends in Biotechnology

CellPress

Science & Society

Governance strategies for biological AI: beyond the dual-use dilemma

Alex B. Lu[1,2] and Anna C.F. Lewis[2,3,*]

A common framing for governing artificial intelligence (AI) in the biological sciences is to focus on risk mitigation owing to the technology's potential 'dual-use' for both beneficial and harmful applications. This is a reactive policy approach, and a broader framing that urges policymakers to actively champion the benefits alongside mitigating risks is needed, including through targeted investments aimed at securing public priorities.

The dual-use dilemma

At the forefront of excitement about AI is the potential to massively accelerate the process of biological discovery. A recent flurry of announcements shows the contours of this promise: progress in biological models and progress in accelerating the scientific process itself. These technologies are often framed as 'dual-use': they can be used to benefit humanity, but are also associated with biosecurity risks, including the development of bioweapons. The dual-use framing encourages a focus on governance tasks that mitigate these risks, often obscuring the benefits [1,2]. While current policy priorities intend to prevent 'burdensome requirements' on the AI industry[i], there may be particular motivation to minimize biosecurity risks. In this article we highlight the importance of governance strategies that – alongside mitigating risk – actively pursue the promise of biological AI, and do so in a way that steers towards public priorities.

The promise

Whereas familiar models such as OpenAI's GPT series are trained on human language, there are also large language models (LLMs) trained instead on the language of life: biological sequence data. One example is ESM3, which can predict the structure and function of protein sequences and design novel proteins [3]. Another example, Evo2, can design novel genomes up to the size of bacterial genomes and make predictions about the function of DNA sequences [4]. Both are open sourced.

These models give scientists new capabilities. To capitalize on this, however, they must be used for specific purposes, requiring sufficiently granular research questions connected to well-motivated research programs. Because LLMs such as OpenAI's GPT-4 are trained on the near totality of published scientific knowledge, they can assist this process. In the context of well-defined problems, existing models can handle complex problem-solving and knowledge integration tasks [5].

Any output the biological models generate requires extensive follow-up, usually in laboratories. This yields new data that can be fed into a model, resulting in a recursive design–make–test–analyze (DMTA) cycle where AI is involved at every stage of the scientific process, including directing real-world experiments with minimal human oversight. For example, the Senate's National Security Commission on Emerging Biotechnology recommends that the National Science Foundation (NSF) create a series of autonomous cloud laboratories, enabling researchers to generate high-throughput and high-quality data to train AI models[ii]. While early implementations will likely encounter technical and methodological challenges, the field is poised for rapid expansion.

Mitigating risks

As many commentators have flagged, these technologies are dual-use (see Box 1 for a case study) [1]. Concern about the potential for novice actors with malicious intent to produce bioweapons is shared by the frontier AI companies[iii]. In May 2025, Anthropic announced that they could not conclusively rule out the possibility that their latest model could supply bad actors with dangerous bioweapons knowledge[ii].

How can we mitigate biosecurity risks? In the spirit of proactive risk management, there have been developer-led efforts to guide safe model design[iv] [4]. Risk can be reduced through choices made during model development, and, in the case of closed-source models, through including refusal mechanisms. The success of these strategies can then be assessed via robust red teaming (Box 2). Red teaming is a common strategy for LLMs, but it can come with inherent risks with biological AI: unlike validating text, where outputs remain confined to digital environments, generative outputs of these models cannot be fully assessed through *in silico* methods alone, requiring instead real-world experimentation. An additional challenge is that many beneficial tasks are hard to distinguish from malicious intent, making this approach complex for balancing functionality and risk.

In the context of the USA, there has been a policy shift away from mitigation frameworks concentrated on the 'responsible development and use' of AI to prioritize promoting the country's dominance in the AI industry. While this suggests a light regulatory touch, there are reasons to believe that US-based policymakers may be particularly motivated to stem off biosecurity risks because of developments outside of biological AI. Most prominently, the White House in 2025 remarked that the virus that caused the COVID pandemic was caused by a biosecurity lapse, a laboratory leak from the Wuhan Institute of Virology[v]. In June 2025, the NIH terminated funding for research that seeks knowledge through engineering new capabilities for pathogens

Box 1. Dual-use case study

Urbina *et al.* [6] demonstrated the ease with which AI-powered biomedical tools can be appropriated for creating weapons of mass destruction. Their model, MegaSyn, is typically used as a molecular generator to discover therapeutic inhibitors of disease by penalizing compound toxicity and rewarding target activity. By tweaking their model to reward rather than penalize toxicity, their previously harmless MegaSyn discovered over 40 000 molecules more toxic than the most known toxic chemical agents. The authors stated: 'we had transformed our innocuous generative model from a helpful tool of medicine to a generator of likely deadly molecules'.

(i.e., gain-of-function research) because of its biosecurity risk[vi]. Given this emerging concern to mitigate risks, policymakers must carefully consider how those strategies affect innovation capacity and how they detract from securing the promise of biological AI.

Policy strategies for supporting innovation

A policy limitation of the dual-use framing is that it invites too heavily a focus on the trade-offs between aggregate risks versus aggregate benefits, framing policy as needed to mitigate the worst risks in a way that detracts from the benefits as little as possible. While hard choices to protect biosecurity are necessary, a proactive agenda focused on capturing the benefits of AI as applied to the biological sciences is also needed [7].

Part of this benefit-focused strategy would be direct investment. Biological research is widely seen as pre-competitive, which motivates financial support from government funders. In particular, there are areas where investment incentives can be insufficient for the marketplace, such as the identification of novel antibiotics or treatments for ultra-rare diseases. Funding for AI-enabled advances in these areas could advance basic science as well as its translational potential. Collaboration in this space could set the stage for needed broader collaboration between governments, users, researchers, and industry [8].

Beyond this, governments should champion audacious AI-meets-biology projects, of similar ambition to the Human Genome Project. Examples include the proposal for the US Department of the Interior to fund a 'Sequencing Public Lands' initiative, designed to collect, sequence, and archive the non-human life spread across diverse environments of the USA to train models[ii], as well as the US Department of Energy's proposal to create a 'Web of Biological Data', a platform to share biological AI data generated across industry, academia, and governments.

Another way to promote innovation is to lower the barriers to effective model capability testing. This could be achieved by defining robust benchmarks specifically for biological AI risk testing, since pushing risk mitigation responsibility to developers can burden innovation[vii]. Providing centralized capability assessments would accelerate development, and these strategies could build off of efforts from the NIST US AI Safety Institute [9] and the UK's AI Security Institute[viii]. Governments could then require testing against these benchmarks. Governments could also require actors in this space to follow community-established norms for model design and release.

Box 2. Red teaming

Red teaming, a security strategy to deliberately elicit dangerous outputs from AI, requires a nuanced adoption in biological applications due to the potential for creating informational hazards and the need for unpredictable laboratory validation. For biological AI applications, the goal is finding the closest proxy validation, where outputs can be assessed as close to the endpoint as possible without dangerous physical manifestation.

Some strategies include: (i) Automated benchmarking → asking standardized questions that assess model knowledge capabilities, from low-level to severe risks. (ii) Assistant task evaluations → assessing the model's ability to provide assistive guidance in accomplishing specific tasks by posing plausible questions, open-ended inquiries, and real-world interaction scenarios. (iii) Uplift studies → comparing how the model can enhance human capabilities when using a given model versus not. (iv) Expert assessments → leveraging specialists to identify domain-specific concerns in model outputs.

Governments can also actively monitor the implementation of AI for biology to assess whether there are application areas where benefits could be sped up. Ideas for innovative use cases that could benefit humanity could come from anywhere, and governments can play a role to help uncover and encourage these.

Another governance task is investment in human capital, which creates the foundation for rapid innovation. Securing the benefits of bio-AI will require individuals to speak the languages of computer science and the relevant biology topics. Government-funded training programs that bridge these disciplines, as well as open forum discussion opportunities for students, will allow a larger population to engage with the broader implications of their work.

These governance strategies rely on an accurate understanding of evolving capabilities of biological AI, and hence require tracking of the state of the art. Moreover, because the impact of AI-enabled biology will not be restrained by national borders, international cooperation across governance tasks will be necessary; this could build off international efforts such as the 2025 Paris AI action summit and the International AI Safety Summit.

Making hard choices

Finally, as the field progresses, the extent of the risks and benefits will become clearer, and the balance of risk mitigation and investment strategies may need to shift. Governments should develop robust public participation processes to clarify the trade-offs at stake and inform the hard choices that may be needed. Even at a time when investment in science is not assured, robust public participation – involving deliberation between scientists, technologists, and the public – can help ensure that innovations do indeed steer towards public priorities. Human capital investment should extend to ensuring a scientifically informed public and a shared, democratic base of stakeholders invested in innovation.

Concluding remarks

There are strategies for mitigating risks that strike a promising balance between government regulation and scientific discovery, many of which do not impose regulatory burden, and which could be workable in a changing policy landscape. The dual-use framing for technology rightly draws attention to biosecurity risks, but it can sideline policy approaches for capitalizing on the benefits. In addition to mitigating biosecurity concerns, policymakers should actively champion innovation through targeted investments in under-incentivized areas, create ambitious national initiatives, devise capability testing standards, establish benefits monitoring systems, and support public steering capacities.

Declaration of interests

The authors declare no competing interests.

Resources

[i] www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/

[ii] www.biotech.senate.gov/final-report/chapters/

[iii] www-cdn.anthropic.com/6be99a52cb68eb70eb9572b4cafad13df32ed995.pdf

[iv] https://responsiblebiodesign.ai/

[v] www.whitehouse.gov/lab-leak-true-origins-of-covid-19/

[vi] https://grants.nih.gov/grants/guide/notice-files/NOT-OD-25-127.html

[vii] www.brookings.edu/articles/balancing-market-innovation-incentives-and-regulation-in-ai-challenges-and-opportunities/

[viii] cdn.prod.website-files.com/663bd486c5e4c81588db7a1d/68778c08bd1d69a31d4775e5_Elicitation%20Best%20Practices%202.pdf

[1] Metabolism Unit, Massachusetts General Hospital, Boston, MA, USA
[2] Harvard Medical School, Boston, MA, USA
[3] Division of Genetics, Brigham and Women's Hospital, Boston, MA, USA

*Correspondence:
aclewis@bwh.harvard.edu (A.C.F. Lewis).

https://doi.org/10.1016/j.tibtech.2025.09.012

References

1. Molla, K.A. *et al.* (2025) The Spirit of Asilomar: lessons for the next era of biotechnology governance. *Trends Biotechnol.* 43, 1809–18122
2. Bloomfield, D. *et al.* (2024) AI and biosecurity: the need for governance. *Science* 385, 831–833
3. Hayes, T. *et al.* (2024) Simulating 500 million years of evolution with a language model. *Science* 387, 850–858
4. Brixi, G. *et al.* (2025) Genome modeling and design across all domains of life with Evo 2. *bioRxiv* Published online February 21, 2025. https://doi.org/10.1101/2025.02.18.638918
5. Microsoft Research AI4ScienceMicrosoft Azure Quantum (2023) The impact of large language models on scientific discovery: a preliminary study using GPT-4. *arXiv* Published online December 8, 2023. https://doi.org/10.48550/arXiv.2311.07361
6. Urbina, F. *et al.* (2022) Dual use of artificial-intelligence-powered drug discovery. *Nat. Mach. Intell.* 4, 189–191
7. Allen, D. *et al.* (2025) A roadmap for governing AI: technology governance and power-sharing liberalism. *AI Ethics* 5, 3355–3377
8. National Academies of Sciences, Engineering, and Medicine (2025) *The Age of AI in the Life Sciences: Benefits and Biosecurity Considerations*, The National Academies Press, Washington, DC https://doi.org/10.17226/28868
9. NIST (2025) *Managing Misuse Risk for Dual-Use Foundation Models*, National Institute of Standards and Technology Published online January 2025. https://doi.org/10.6028/NIST.AI.800-1.2pd