Navigating AI Bio-risk Frontier: A Four-Domain Framework

Tanya Sarawagi, Suryesh Kumar Namdeo

The authors are researchers at the Indian Institute of Science, Bengaluru, and the views expressed are personal.

Summary

As AI accelerates biological research, governance frameworks to ensure safety and security struggle to keep pace. This article introduces a four-domain framework spanning AI-augmented biodesign, dual-use LLMs, AI-driven misinformation, and AI-enhanced cyber-bio risks. It evaluates India's existing regulatory architecture against these threats and proposes safeguards beyond aspirational principles.

Introduction

Artificial intelligence (AI) is transforming biology at a remarkable speed. From accelerating protein design and structure prediction to democratising access to complex bioscience knowledge, AI advances promise breakthroughs in medicine, agriculture, research, and sustainability. Nevertheless, this same acceleration brings in new vulnerabilities and threats in technical, social, and security aspects that demand urgent practical governance, rather than aspirational principles alone[1]. However, there is limited prioritisation of the AI–Bio nexus in most countries, as it is perceived more as a futuristic issue by policymakers, and they focus on more urgent near-term challenges due to being constrained by limited resources. This perception gap risks allowing governance to lag behind deployment, rather than co-evolve with capability[2, 3].

In view of the emerging risks, this article explores the four critical domains of the AI and biosecurity interface:

(1) Securing AI-augmented Biodesign tools

(2) Responsible use of LLMs for Dual-Use research

(3) AI-driven misinformation and disinformation to obfuscate biothreat response, and

(4) AI-enabled cyber bio-risks.

Securing AI-augmented Biodesign Tools

AI-bio platforms like AlphaFold and RoseTTAFold, and other generative design

models (the AI systems that create new biological designs), have revolutionised protein structure prediction and synthetic biology. Compressing the design and development cycles from years to days enables rapid drug discovery and metabolic engineering. These tools are increasingly embedded across academic, start-up, and industrial pipelines, amplifying both their scientific value and systemic risk. For example, Coalition for Epidemic Preparedness Innovations (CEPI) is creating an agentic AI- based Pandemic Preparedness Engine which harnesses the advances of new generative AI edition, in addition to their predictive model for vaccine development based on known data on virus structures and sequences, their epidemiological spread, and immune system responses.

This tool would significantly reduce the vaccine design and development time from several years to a few months, while enabling security guardrails in the system principles[4].

As the need for specialised expertise declines, biodesign tools are increasingly used not only by expert malign actors but also by less skilled or unsupervised users, expanding pathways for unsafe or unregulated biodesign. So, the risks involve the misuse by experts as well as the scaling of poorly governed participation in high-impact biological design. With the exponential growth of available biodesign tools, it has become hard to track adequate safety measures to prevent misuse.

Voluntary norms and fragmented safeguards are insufficient when design capabilities scale faster than institutional oversight. Responsible deployment, traceable usage records, and screening of AI-generated DNA and protein sequences are essential to mitigate biosecurity risks [5].

Responsible use of Large Language Models (LLMs) for Dual-Use research

LLMs are becoming an everyday gateway to biological knowledge, from summarizing literature to translating jargon and suggesting experimental protocols. However, their training data can blend peer-reviewed research with unverified or unvalidated information and opinion pieces, which could significantly reduce the reliability of the information produced [6]. This epistemic flattening blurs the boundary between validated knowledge and speculative or unsafe practices. The inaccurate and confident but misleading answers are more dangerous than honest uncertainty and evidence-based reasoning. In biosciences, such 'hallucinated certainty' can translate directly into material risk [7].

As public dependency on AI-sourced information increases, misguided enthusiasts may attempt 'harmless' experiments in their kitchen, a prime example of unsupervised research. Such Do-It-Yourself (DIY) biology or biohacking experimentation could create biosafety and biosecurity risks. This represents a shift from institutional laboratories to informal settings, where neither biosafety norms nor regulatory oversight are enforced. AI may act as a primary guidance tool by suggesting protocols, reagents, and instrument usage, which could potentially contribute to unintended pathogenicity, laboratory exposure, or environmental release [8].

AI-driven Misinformation and Disinformation to Obfuscate Biothreat Response

Beyond technical errors, AI can amplify misinformation and disinformation, shaping public perception during biological crises. Misinformation refers to false or inaccurate information shared without malicious intent,

while disinformation involves the deliberate spread of false narratives to mislead or manipulate public opinion. Coordinated campaigns exploiting AI- generated content like synthetic text, images, or deepfakes can undermine trust in health authorities, delay response efforts, and fuel panic, especially when emotionally charged or fear-inducing narratives are algorithmically amplified by the social media platform [9].

The psychological dimension of biothreats, that is, impact on public trust and behaviors, which includes heightened fear, erosion of institutional trust, behavioral non-compliance, and polarisation, is often overlooked [10]. However, it can significantly undermine containment strategies during accidental or deliberate outbreaks. Historical outbreak responses show that communication failure can be as damaging as the responsible pathogen. Proactive monitoring for early detection of such AI-framed canards and clear risk communication with the public and interdepartmental agencies are key defenses [11]. AI itself can be repurposed defensively to detect coordinated manipulation, synthetic media, and narrative amplification patterns in real time.

Cyber Bio-risks: Data and Infrastructure Vulnerability

Modern biology is deeply digital, from genomic databases to digitally controlled automated laboratories and cloud-based bioinformatics pipelines. AI plays a pivotal role in optimizing these systems, automating data analysis, and improving workflow efficiency. This creates expanded opportunities for cyberattacks, including unauthorised access, theft of sensitive biological data, sabotage of high-containment laboratories, exploitation of supply chains of biomaterials, manipulation in workflows or protocols, and tampering with digitally connected lab instruments. AI-driven algorithms, when exploited by malicious actors, can facilitate targeted attacks, manipulate datasets, or even override safety protocols in laboratory systems, further amplifying the risk of these vulnerabilities.[12]

This affects the bigger picture because compromised research processes will lead to scientifically invalid outputs and flawed publications, which raise questions about trust in scientific oversight and validation.

AI itself is a double edged sword here: adversaries can use it to probe vulnerabilities, but defenders can deploy AI for anomaly detection, intrusion prevention, and supply chain integrity (e.g., monitoring robotics logs, sequencing pipelines, and cloud workflows for deviations). Contemporary cybersecurity literature shows that supervised and deep learning approaches reduce false positives and catch zero day behaviors (a secret loophole about which the authorities are unaware; however, an attacker finds it and utilizes it to perpetrate the system, giving zero days to the authorities to play defense) [13]. These techniques should be adapted for use in bioinformatics and laboratory automation environments. Further, harmonised standards, robust encryption, and immutable audit trails combined with AI- enabled smart techno-legal measures could be prioritised for addressing these challenges. These measures are particularly critical for high-containment labs and sensitive research infrastructure.

Pragmatic Path Forward

Taken together, these four domains identify and categorise the diverse biosecurity challenges due to the advent of AI. This further necessitates a clear and workable policy strategy to address them. Fragmented ethical guidance must now be complemented by enforceable technical and institutional

controls. Concretely, this means adopting auditable AI biodesign workflows, involving checks on genetic sequences and end users of synthetic DNA, aligned with the emerging international best practices. Parallelly, provenance-aware LLM outputs [14]should be mandated, including clear disclosure of data sources and uncertainty. The government should deploy infodemic-aware communications strategies to counter misinformation during health emergencies. Finally, cybersecurity in digital laboratories should be strengthened by utilising automated systems to flag unusual or high-risk activity. With such practical safeguards, AI–Bio can deliver scientific and economic benefits while minimizing foreseeable harm.

Leveraging India's Existing Frameworks for AI-Biosecurity:

Mentioned below are the key existing ethical and regulatory frameworks that can be utilised to partially address AI–Bio governance challenges:

1) **Digital Personal Data Protection Act, 2023**

   The DPDP Act establishes individual rights over digital personal data and imposes obligations on data fiduciaries, the entities that collect and process data, including notice and consent, purpose limitation, data minimisation, accuracy, and reasonable security safeguards. For AI- Bio, this creates a legal framework for the protection of sensitive personal data, including health and genomic data, by mandating lawful processing, defined purposes, and accountability mechanisms. However, the DPDP Act is fundamentally data-centric rather than technology- or risk-specific and therefore does not directly address governance gaps arising from AI-driven biodesign, automated laboratory systems, or dual-use bioengineering applications.

2) **India AI Governance Guidelines, 2025**

   The guidelines emphasise a 'whole-of-government' approach, in which all line ministries, sectoral regulators, technical standard bodies and other public institutions work together to develop and implement AI policy. Interestingly, AI is framed as a dual-use, national security-relevant technology, creating a policy space to deal with AI-biosecurity risks. Establishing bodies such as the AI Safety Institute (AISI) and the AI Governance Group (AIGG) could be the ground for AI education and safe AI awareness programs. These initiatives could be supported by the Technology & Policy Expert Committee (TPEC) by the inclusion of experts from biosecurity in the committee to underscore the AI-linked biosecurity issues. Further, the techno-legal 'compliance by design' approach is strongly endorsed, which requires embedding guardrails and safety measures in high-risk AI models.

3) **ICMR Ethical Guidelines for AI in Healthcare (2023)**

   The Indian Council for Medical Research (ICMR) guidelines emphasize patient-centric principles, including accountability, consent, privacy, safety, non-discrimination, and clinical validation. These guidelines require human decision-making for all important clinical and medical matters, thereby reducing some AI threats. These guidelines are well-suited for clinical AI applications but have limited reach beyond healthcare delivery and do not involve non-medical usage or core research biology, particularly in AI-assisted biodesign and cyber-biosecurity.

What's Next: Closing the Gaps

To meaningfully govern AI-Bio, India should move beyond fragmented oversight and adopt targeted expansion of existing regulations.

Key priorities could include:

- Encouraging adoption of global sequence and customer screening standards across DNA synthesis providers as well as similar checks on small molecule and protein producers;
- Aligning lab automation and cloud bioinformatics pipelines with sector-specific cybersecurity baselines;
- Operationalizing provenance labeling and model accountability for bioscience-focused LLMs;
- Establish cyber rapid response teams to quickly respond to misinformation and disinformation campaigns related to CBRN threats; and
- Develop national capabilities to evaluate AI models and tools for potential biosecurity threats.

Together with DPDP protections and ICMR ethics guidelines, these steps would strengthen India's capacity to safely scale AI-Bio innovation while maintaining public trust, scientific integrity, and national security.

Endnotes:

[1] Bhardwaj, Abhaya, Shristi Kishore, and Dhananjay K. Pandey. 2022. "Artificial Intelligence in Biological Sciences" *Life* 12, no. 9: 1430. https://doi.org/10.3390/life12091430

[2] Wünn, Tina. "Exploring AI-Biosecurity Governance in the Global South." *Nuclear Threat Initiative*, December 5, 2024. https://www.nti.org/risky-business/exploring-ai-biosecurity-governance-in-the-global-south/.

[3] "Generative AI for Biosciences: Emerging Threats and Roadmap to Biosecurity." arXiv preprint arXiv:2510.15975 (2025). https://arxiv.org/abs/2510.15975.

[4] Stuart, Lynda M., Rick A. Bright, and Eric Horvitz. "AI-Enabled Protein Design: A Strategic Asset for Global Health and Biosecurity." *NAM Perspectives*. National Academy of Medicine, October 28, 2024. https://doi.org/10.31478/202410d.

[5] Dessimoz, C., Thomas, P.D. AI and the democratization of knowledge. *Sci Data* 11, 268 (2024). https://doi.org/10.1038/s41597-024-03099-1

[6] Wang, et al. "Large Language Model Agents for Biological Intelligence across Genomics, Proteomics, Spatial Biology, and Biomedicine." *Briefings in Bioinformatics* (2026). https://academic.oup.com/bib/article/doi/10.1093/bib/bbag110/8540361.

[7] Farquhar, Sebastian, Jannik Kossen, Lorenz Kuhn, and colleagues. "Detecting Hallucinations in Large Language Models Using Semantic Entropy." *Nature* 630, no. 8017 (2024): 625–630. https://doi.org/10.1038/s41586-024-07421-0.

[8] Undheim, Trond Arne. "The Whack-a-Mole Governance Challenge for AI-Enabled Synthetic Biology: Literature Review and Emerging Frameworks." *Frontiers in Bioengineering and Biotechnology* 12 (2024): 1359768. https://doi.org/10.3389/fbioe.2024.1359768.

[9] Naffi, Nadia. "Deepfakes and the Crisis of Knowing." *UNESCO*, October 1, 2025. https://www.unesco.org/en/articles/deepfakes-and-crisis-knowing.

[10] Tegomoh, Bryan. *The Public Health AI Handbook: Evaluating AI Tools for Public Health Practice*. 2025. https://publichealthaihandbook.com/

[11] Jay K. Varma, "The AI Threat to Public Health No One Is Thinking About: A Fake Bioterrorist Attack," *STAT*, May 27, 2025, https://www.statnews.com/2025/05/27/artificial-intelligence-bioterrorism-deepfake-public-health-threat/.

[12] Elgabry, Mohamed, and Simon Johnson. "Cyber-Biological Convergence: A Systematic Review and Future Outlook." *Frontiers in Bioengineering and Biotechnology* 12 (2024): 1456354. https://doi.org/10.3389/fbioe.2024.1456354.

[13] Al Siam, Abdullah, Nuruzzaman Faruqui, Akm Azad, and Mohammad Ali Moni. "Securing the Unseen: A Comprehensive Exploration Review of AI-Powered Models for Zero-Day Attack Detection." *Expert Systems* 43, no. 3 (2026): e70217. https://doi.org/10.1111/exsy.70217.