Position: Restricted Release of Advanced Biological Models Safeguards Biosecurity

Jonathan Feldman
Georgia Institute of Technology
Atlanta, GA
jonathanfeldman@gatech.edu

Tal Feldman
Yale Law School
New Haven, CT
tal.feldman@yale.edu

Abstract

Recent advances in generative protein design, protein language models, and genomic language models have unlocked the ability to generate novel biomolecular sequences with unprecedented efficiency. These systems promise major breakthroughs in drug discovery, synthetic biology, and fundamental research but also create a high-stakes national security challenge. Unlike general-purpose language models designed for broad public use, these are highly specialized systems with capabilities that only a small community of researchers legitimately needs. Unrestricted open-source release of such models lowers the expertise and resource thresholds required to engineer pathogenic proteins or other hazardous biomolecules, making them attractive tools for malicious actors. We argue that safeguarding national security requires ensuring that high-risk models are available only to trusted researchers and institutions with appropriate biosecurity capacity, while maintaining broad support for open scientific progress in lower-risk domains. We evaluate three approaches for constraining distribution: governmental regulation, coordinated self-governance within the research community, and architectural or dataset-level interventions such as the targeted exclusion of pathogenic sequences. By weighing the feasibility and limitations of each, we argue for proactive safeguards that both protect national security and sustain a vibrant research and innovation ecosystem.

1 Introduction

Generative artificial intelligence and molecular biology have converged to create highly specialized tools capable of designing proteins and genomes with unprecedented precision [1–3]. Protein language models, genomic design systems such as Evo 2, structure predictors, and generative design platforms, like RFdiffusion, are transforming the biotechnology landscape [1, 4–6]. These systems can generate novel enzymes, targeted vaccines, engineered antibodies, and synthetic genomes far faster than traditional methods [2, 7, 8]. As a result, therapeutic development is accelerating, research bottlenecks are shrinking, and solutions are emerging for biological challenges once considered beyond reach [9, 10].

Artificial intelligence is now a central driver of biological discovery. Protein language models trained on vast datasets can predict properties such as fitness, structural stability, and function, and can generate entirely new proteins [6, 11]. In the fight against SARS-CoV-2, such models were used to predict viral fitness, anticipate immune escape, accelerate vaccine and therapeutic development, and model possible evolutionary pathways [12–14]. Their integration into research pipelines is enabling breakthroughs across medicine, agriculture, and industrial biotechnology [15, 16].

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: The Second Workshop on GenAI for Health: Potential, Trust, and Policy Compliance.

Yet the same capabilities that enable progress also create significant dual-use risks [10, 13, 17]. Unlike general-purpose language models designed for wide public use, these biological design systems are highly specialized, with applications relevant only to a narrow set of researchers. In the wrong hands, they could lower the expertise, time, and cost required to engineer synthetic threats [10, 17]. Protein language models can propose mutations that enhance transmissibility or resistance to treatment; genomic design systems can generate optimized viral genomes; and structure-aware systems such as RFdiffusion can refine and validate these sequences in silico [6, 18, 19]. This design–validation cycle mirrors legitimate biomedical research, but for a malicious actor it could accelerate the creation of dangerous agents, some capable of evading current detection systems [20].

Safeguards such as real-time sequence filtering, red-teaming to uncover misuse pathways, and screening for pathogenic molecules are among the most practical tools for reducing the likelihood of AI-enabled biological threats [10, 21]. Many defenses work by detecting homology to known viruses or proteins, which can block most low-resource misuse by individuals or groups without significant expertise [22, 23]. But this approach cannot yet reliably stop radically novel designs or deter well-funded actors with access to custom models and synthesis infrastructure [10, 20, 23]. To preserve the effectiveness of safeguards, it is our position that two conditions are essential. First, high-risk protein and genomic design models must shift from open to closed-source access. Second, their distribution must be limited to vetted institutions with demonstrated biosecurity capacity. Open release of model weights or unrestricted code enables easy circumvention of safety controls, while closed access allows developers and oversight bodies to adapt safeguards—such as inference-time screening with BLAST or structural homology search—as models advance [23].

Securing model access must be complemented by safeguards at the synthesis stage. DNA synthesis providers should screen orders not only for exact sequence matches but also for structural similarity to dangerous proteins or genomes. This two-tiered approach—governing both the design and production phases—would substantially raise barriers to malicious use while preserving the open scientific ecosystem that drives legitimate progress.

It is critical to emphasize that these safeguards, while essential, are not sufficient on their own to fully prevent future biosecurity risks. A resilience-based strategy, which focuses on the rapid development, validation, and deployment of therapeutics, vaccines, and other countermeasures, must be pursued alongside preventative measures [24, 25]. Effective biosecurity requires a multifaceted approach that both limits the potential for misuse through controlled access and oversight and strengthens our capacity to respond quickly to emerging threats.

We argue that high-risk biological design models, including those for proteins and genomes, should not be openly released. Instead, access should be limited to trusted researchers and institutions under strict oversight. Striking this balance is essential: fostering a vibrant research and innovation ecosystem while ensuring that these powerful technologies do not become tools for creating the next generation of biological threats.

2 The Case for Closed-Source Models

Advanced protein and genomic design models present an unusually high biosecurity risk. These systems can engineer viral proteins with enhanced immune escape, increased binding affinity, or other pathogenic traits [10, 17]. Models and frameworks such as EVEscape, EVE-Vax, and VIRAL can map mutational pathways to generate novel, high-risk variants of existing viruses [8, 13, 26]. Generative design platforms like RFdiffusion and ESM3 extend this further, producing entirely new proteins that share little or no sequence similarity with known pathogens yet may act as structural homologs [6, 10, 27]. Such designs could evade sequence-based diagnostics and detection systems, creating substantial national security concerns if misused [20].

These models enable misuse in several technical ways. They can optimize protein sequences for stability, binding affinity, or immune evasion, effectively providing a "blueprint" for pathogenic proteins [13, 28, 29]. They can explore mutational spaces far beyond natural evolution, revealing sequences with enhanced virulence or resistance to existing therapeutics [12, 26, 30, 31]. Theoretically, model latent-space manipulation can be exploited to create entirely novel proteins that retain functional activity but remain invisible to homology-based screens. Even actors without full experimental pipelines could identify high-risk candidates and outsource synthesis to third parties—underscoring the need for preemptive safeguards [9, 10, 32].

Open access to these models would substantially undermine existing protections. Controlled-access deployment allows operators to implement inference-time safeguards such as keyword filtering, homology screening, and embedding-based similarity detection. For example, the ESM3 API automatically screens outputs against databases of viral and pathogenic proteins, blocking high-risk sequences [33]. While not foolproof—particularly against radically novel designs or well-funded state-level actors—such measures remain the most practical way to reduce misuse by lower-resource or opportunistic groups [32, 34].

Closed-sourcing also enables monitoring and accountability. Outputs can be logged for review, and suspicious users flagged for oversight [35, 36]. Software frameworks can maintain these logs even if limited code or weights are shared with vetted researchers.

While critics contend that openness accelerates scientific progress, full release of model weights is not necessary for reproducibility or legitimate research. Replication and validation can be supported through controlled access mechanisms, such as vetted data sharing and audited compute environments. The incremental scientific value of unrestricted weight release is minimal compared to the substantial security risks, particularly in biological design, where generated sequences must undergo experimental validation to assess clinical relevance [10, 29]. Moreover, the threat is not confined to state actors; many malicious activities are undertaken by individuals or small groups with limited resources. By increasing the cost and complexity of misuse, controlled access provides critical time for detection, mitigation, and response.

Importantly, closed-sourcing does not preclude scientific progress. Advanced biological design still requires substantial wet-lab and computational infrastructure [10, 37]. Secure APIs and restricted-weight sharing—modeled on services like the AlphaFold Server or controlled ESM3 API—can provide researchers with access while preserving screening protocols, audit logs, and inference-time safeguards [33, 38]. Closed-source distribution ensures that as capabilities advance, safeguards remain enforceable. Once model weights are openly released, they cannot be recalled or retroactively regulated—an irreversible exposure of dual-use technology [39].

3 Strategies for Managing Closed-Source Protein Models

3.1 Federal Roles in Managing High-Risk Protein Models

Managing access to advanced protein and genomic design models requires safeguards that reduce the risk of misuse without stifling scientific progress. Most biological and computational research should remain entirely unaffected. But a narrow class of systems with unusually high capabilities poses distinctive national security concerns, and the federal government has a central role to play in ensuring they are managed responsibly. Oversight in this context should remain focused and unobtrusive—designed to preserve innovation while creating meaningful barriers to misuse.

The first step is to define clearly which models fall into this category. Risk should be tied to demonstrated capability rather than model size or training data. Systems that can reliably propose immune-escape mutations, design functional toxins, or generate novel proteins that act as functional or structural homologs of dangerous pathogens warrant special treatment. Developers of such models should be expected to provide a concise description of the system's capabilities, safeguards, and intended use contexts. Such metadata, compiled in a simple federal registry, creates transparency without forcing every new model through an onerous approval pipeline.

Access to these models should be limited to vetted researchers and institutions with appropriate biosafety and governance infrastructure. Federal agencies are well placed to coordinate this process, which should resemble familiar applications for grants or high-performance computing resources: structured, documented, but not unduly burdensome. Once a laboratory demonstrates that it meets baseline standards, it should be able to access multiple models without repeated applications. Temporary supervised access can be provided to peer reviewers, allowing reproducibility to be maintained even when model weights are not openly distributed.

For commercial organizations, compliance would be minimal. Companies already monitor usage, protect intellectual property, and restrict access to proprietary algorithms. Reporting on safeguards such as red-teaming, filtering, and query logging is aligned with existing practice and does not require new bureaucracy. Academic institutions face a different challenge, given their emphasis on open science. Here the balance should be to continue publishing research findings freely, while ensuring

that access to high-risk models occurs through secure APIs or supervised weight-sharing, with clear records of institutional approval coordinated through federal channels.

Technical measures further reduce misuse risk while keeping research efficient. Hosted access allows inference-time safeguards such as sequence and structure screening, anomaly detection, and query-pattern monitoring. Logs of model outputs, preserved in tamper-evident form but with appropriate privacy protections, create accountability and allow for after-the-fact review in the event of suspected misuse. None of these measures are perfect, but together they raise the cost of malicious activity while imposing minimal friction on legitimate research.

Oversight of model access should also be paired with safeguards at the synthesis stage. DNA synthesis providers form the final barrier between a digital design and a physical organism. This approach aligns with recent federal initiatives such as the Biden administration's directives on DNA synthesis screening and the Trump administration's AI Action Plan, both of which emphasize responsible innovation through integrated oversight mechanisms [40, 41]. Universal screening of orders for both sequence and structural similarity to pathogens should become standard practice, integrated into existing provider workflows. Compliance can be tied to funding eligibility and procurement preferences rather than solely strict punitive measures, maintaining a considerate approach while establishing clear expectations.

Because biotechnology and AI research are global, federal oversight must also be coordinated internationally. Shared minimum standards for model access, safeguards, and synthesis screening can prevent adversaries from exploiting regulatory gaps, while mutual recognition of vetted institutions reduces duplicative bureaucracy across jurisdictions. To avoid unnecessary restrictions, well-qualified laboratories in partner countries should be supported so that compliance costs do not exclude them from legitimate research.

Oversight should remain adaptive, with requirements subject to periodic review and sunset clauses that can be adjusted as evidence evolves. This form of collaborative governance ensures that rules remain flexible and do not become rigid barriers that hinder scientific progress.

Essentially, a narrow set of biological design models should be treated as sensitive and subject to controlled access under coordinated federal oversight. Researchers and institutions seeking to use them would undergo a straightforward vetting process and then be permitted access through secure channels with built-in safeguards. DNA synthesis providers would apply standardized screening, and allies would coordinate internationally to avoid gaps and duplicative burdens. This approach preserves an open and vibrant research ecosystem while ensuring that the most powerful biological design tools are managed with the seriousness their national security implications demand.

3.2 Community-Led Standards and Consortia for Managing High-Risk Protein Models

An alternative to federal oversight is a community-led model in which universities, companies, and research organizations voluntarily coordinate standards for responsible development and access. Many commercial entities already have incentives to keep high-capability biological models closed to protect intellectual property, but they should also be encouraged to adopt and publicize strong safeguards. Because corporate incentives vary, complementary self-governance by the academic and nonprofit research community is essential.

In this approach, a decentralized consortium of participating institutions would set baseline requirements that any high-risk biological design model must meet before dissemination or demonstration. Crucially, the consortium would establish a strong expectation that such models remain closed-source, with access mediated through secure APIs or supervised sharing with approved institutions.

To make compliance feasible, the consortium could coordinate funding to offset the costs of responsible hosting—maintaining secure portals, monitoring systems, and compute infrastructure. By lowering the expense of secure deployment, well-resourced standards help make closed, responsible access the easier choice.

Community accountability would also play a role. Participating institutions would be expected to log and report suspicious or malicious activity, with responses handled through established academic or institutional processes. The consortium could also coordinate with major funders and publishers so that compliance becomes a condition of grant eligibility or publication. In practice, this would align

incentives: responsible actors gain recognition and access, while persistent non-compliance carries reputational and funding consequences.

Finally, the consortium should promote ongoing research into model evaluation and biosecure benchmarking. Shared datasets, standardized protocols, and collaborative red-teaming would raise the technical bar for responsible release and ensure that safety assessments are comparable and reproducible.

This decentralized approach cannot guarantee perfect coverage. But by embedding the expectation of closed access into professional norms, pooling resources for secure hosting, and aligning publishers and funders around shared standards, the research community can significantly reduce risks while preserving the openness and collaboration that drive innovation in the life sciences.

3.3 Pathogen Exclusion: A Dataset-Level Intervention

Achieving either comprehensive federal oversight or broad community concordance will take sustained effort and time. As an interim measure, one practicable dataset-level intervention is to restrict the open-sourcing of protein design models, protein language models, and genomic language models trained on pathogenic or viral sequences.

By excluding such data—as in the open-source versions of ESM3 or Evo 2—the generative capacity of models to produce hazardous proteins can be meaningfully reduced [4, 27, 31]. This safeguard is not foolproof, since homologous proteins and shared domains may still present risks, but it represents a narrower and more feasible step than wholesale regulation of all models, and one more likely to attract community and policy support.

Under this approach, researchers could still develop internal models trained on comprehensive datasets, including pathogenic proteins, but these would remain closed and accessible only through secure portals or private hosting. Only models trained on non-pathogenic data would be eligible for open-source release. This compromise preserves the benefits of transparency and collaboration for the vast majority of biological research while lowering the likelihood that open-source systems could be directly misused to design dangerous sequences.

4 Conclusion

The rapid advancement of generative protein design models, protein language models, and genomic language models has brought the life sciences to a pivotal juncture. These systems offer transformative opportunities for scientific discovery and biomedical innovation, yet their capacity to generate novel, potentially pathogenic biomolecules introduces unprecedented biosecurity risks. Unrestricted open-source dissemination of such models removes critical barriers to misuse, placing powerful capabilities within reach of actors lacking the necessary oversight or accountability.

This paper has argued that some, and in certain cases all, high-capability biological design models should remain closed-source to mitigate these risks. Through the examination of governmental regulation, coordinated academic self-governance, and targeted exclusion of pathogenic data from training sets, we have outlined multiple pathways for restricting access. While each carries limitations, timely consensus and decisive action can ensure that the impact on legitimate research remains minimal.

By implementing safeguards before dangerous capabilities become entrenched in the open domain, the research community can protect the integrity of scientific progress while strengthening global biosecurity. Swift, coordinated measures will not only reduce the risk of catastrophic misuse, but also set a precedent for responsible governance in the age of powerful generative biological technologies.

References

[1] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael

Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John M. Jumper. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630 (8016):493–500, June 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07487-w. URL https://www.nature.com/articles/s41586-024-07487-w. Publisher: Nature Publishing Group.

[2] Jeffrey A. Ruffolo and Ali Madani. Designing proteins with language models. *Nature Biotechnology*, 42(2):200–202, February 2024. ISSN 1546-1696. doi: 10.1038/s41587-024-02123-4. URL https://doi.org/10.1038/s41587-024-02123-4.

[3] Miruna Cretu, Charles Harris, Ilia Igashov, Arne Schneuing, Marwin Segler, Bruno Correia, Julien Roy, Emmanuel Bengio, and Pietro Liò. SynFlowNet: Design of Diverse and Novel Molecules with Synthesis Constraints, April 2025. URL http://arxiv.org/abs/2405.01155. arXiv:2405.01155 [cs].

[4] Garyk Brixi, Matthew G. Durrant, Jerome Ku, Michael Poli, Greg Brockman, Daniel Chang, Gabriel A. Gonzalez, Samuel H. King, David B. Li, Aditi T. Merchant, Mohsen Naghipourfar, Eric Nguyen, Chiara Ricci-Tam, David W. Romero, Gwanggyu Sun, Ali Taghibakshi, Anton Vorontsov, Brandon Yang, Myra Deng, Liv Gorton, Nam Nguyen, Nicholas K. Wang, Etowah Adams, Stephen A. Baccus, Steven Dillmann, Stefano Ermon, Daniel Guo, Rajesh Ilango, Ken Janik, Amy X. Lu, Reshma Mehta, Mohammad R. K. Mofrad, Madelena Y. Ng, Jaspreet Pannu, Christopher Ré, Jonathan C. Schmok, John St John, Jeremy Sullivan, Kevin Zhu, Greg Zynda, Daniel Balsam, Patrick Collison, Anthony B. Costa, Tina Hernandez-Boussard, Eric Ho, Ming-Yu Liu, Thomas McGrath, Kimberly Powell, Dave P. Burke, Hani Goodarzi, Patrick D. Hsu, and Brian L. Hie. Genome modeling and design across all domains of life with Evo 2, February 2025. URL https://www.biorxiv.org/content/10.1101/2025.02.18.638918v1. Pages: 2025.02.18.638918 Section: New Results.

[5] Jonathan Feldman and Jeffrey Skolnick. Af3complex yields improved structural predictions of protein complexes. *Bioinformatics*, 41(8):btaf432, 07 2025. ISSN 1367-4811. doi: 10.1093/bioinformatics/btaf432. URL https://doi.org/10.1093/bioinformatics/btaf432.

[6] Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas F. Milles, Basile I. M. Wicky, Nikita Hanikel, Samuel J. Pellock, Alexis Courbet, William Sheffler, Jue Wang, Preetham Venkatesh, Isaac Sappington, Susana Vázquez Torres, Anna Lauko, Valentin De Bortoli, Emile Mathieu, Sergey Ovchinnikov, Regina Barzilay, Tommi S. Jaakkola, Frank DiMaio, Minkyung Baek, and David Baker. De novo design of protein structure and function with RFdiffusion. *Nature*, 620(7976):1089–1100, August 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06415-8. URL https://www.nature.com/articles/s41586-023-06415-8. Publisher: Nature Publishing Group.

[7] Changge Guan, Fangping Wan, Marcelo D. T. Torres, and Cesar de la Fuente-Nunez. Improving functional protein generation via foundation model-derived latent space likelihood optimization, January 2025. URL https://www.biorxiv.org/content/10.1101/2025.01.07.631724v1. Pages: 2025.01.07.631724 Section: New Results.

[8] Noor Youssef, Sarah Gurev, Fadi Ghantous, Kelly P. Brock, Javier A. Jaimes, Nicole N. Thadani, Ann Dauphin, Amy C. Sherman, Leonid Yurkovetskiy, Daria Soto, Ralph Estanboulieh, Ben Kotzen, Pascal Notin, Aaron W. Kollasch, Alexander A. Cohen, Sandra E. Dross, Jesse Erasmus, Deborah H. Fuller, Pamela J. Bjorkman, Jacob E. Lemieux, Jeremy Luban, Michael S. Seaman, and Debora S. Marks. Computationally designed proteins mimic antibody immune evasion in viral evolution. *Immunity*, 58(6):1411–1421.e6, June 2025. ISSN 1074-7613. doi: 10.1016/j.immuni.2025.04.015. URL https://www.cell.com/immunity/abstract/S1074-7613(25)00178-5. Publisher: Elsevier.

[9] Sriram Kosuri and George M. Church. Large-scale de novo DNA synthesis: technologies and applications. *Nature Methods*, 11(5):499–507, May 2014. ISSN 1548-7105. doi: 10.1038/

nmeth.2918. URL https://www.nature.com/articles/nmeth.2918. Publisher: Nature Publishing Group.

[10] Dianzhuo Wang, Marian Huot, Zechen Zhang, Kaiyi Jiang, Eugene I Shakhnovich, and Kevin M Esvelt. Without Safeguards, AI-Biology Integration Risks Accelerating Future Pandemics. 2025. doi: 10.13140/RG.2.2.29765.15849. URL https://rgdoi.net/10.13140/RG.2.2.29765.15849. Publisher: Unpublished.

[11] Dina Listov, Casper A. Goverde, Bruno E. Correia, and Sarel Jacob Fleishman. Opportunities and challenges in design and optimization of protein function. *Nature Reviews Molecular Cell Biology*, 25(8):639–653, August 2024. ISSN 1471-0080. doi: 10.1038/s41580-024-00718-y. URL https://www.nature.com/articles/s41580-024-00718-y. Publisher: Nature Publishing Group.

[12] Brian L. Hie, Kevin K. Yang, and Peter S. Kim. Evolutionary velocity with protein language models predicts evolutionary dynamics of diverse proteins. *Cell Systems*, 13(4):274–285.e6, April 2022. ISSN 2405-4720. doi: 10.1016/j.cels.2022.01.003.

[13] Marian Huot, Dianzhuo Wang, Jiacheng Liu, and Eugene Shakhnovich. Few-Shot Viral Variant Detection via Bayesian Active Learning and Biophysics, March 2025. URL https://www.biorxiv.org/content/10.1101/2025.03.12.642881v1. Pages: 2025.03.12.642881 Section: New Results.

[14] Md Shahadat Hossain, A. Q. M. Sala Uddin Pathan, Md Nur Islam, Mahafujul Islam Quadery Tonmoy, Mahmudul Islam Rakib, Md Adnan Munim, Otun Saha, Atqiya Fariha, Hasan Al Reza, Maitreyee Roy, Newaz Mohammed Bahadur, and Md Mizanur Rahaman. Genome-wide identification and prediction of SARS-CoV-2 mutations show an abundance of variants: Integrated study of bioinformatics and deep neural learning. *Informatics in Medicine Unlocked*, 27:100798, January 2021. ISSN 2352-9148. doi: 10.1016/j.imu.2021.100798. URL https://www.sciencedirect.com/science/article/pii/S2352914821002677.

[15] Anshu Ankolekar, Lisanne Eppings, Fabio Bottari, Inês Freitas Pinho, Kit Howard, Rebecca Baker, Yang Nan, Xiaodan Xing, Simon LF Walsh, Wim Vos, Guang Yang, and Philippe Lambin. Using artificial intelligence and predictive modelling to enable learning healthcare systems (lhs) for pandemic preparedness. *Computational and Structural Biotechnology Journal*, 24:412–419, 2024. ISSN 2001-0370. doi: https://doi.org/10.1016/j.csbj.2024.05.014. URL https://www.sciencedirect.com/science/article/pii/S2001037024001600.

[16] Andreas Holzinger, Katharina Keiblinger, Petr Holub, Kurt Zatloukal, and Heimo Müller. Ai for life: Trends in artificial intelligence for biotechnology. *New Biotechnology*, 74:16–24, 2023. ISSN 1871-6784. doi: https://doi.org/10.1016/j.nbt.2023.02.001. URL https://www.sciencedirect.com/science/article/pii/S1871678423000031.

[17] Mengdi Wang, Zaixi Zhang, Amrit Singh Bedi, Alvaro Velasquez, Stephanie Guerra, Sheng Lin-Gibson, Le Cong, Yuanhao Qu, Souradip Chakraborty, Megan Blewett, Jian Ma, Eric Xing, and George Church. A call for built-in biosecurity safeguards for generative AI tools. *Nature Biotechnology*, 43(6):845–847, June 2025. ISSN 1546-1696. doi: 10.1038/s41587-025-02650-8.

[18] Philip Hunter. Security challenges by ai-assisted protein design. *EMBO reports*, 25(5):2168–2171, 2024. doi: https://doi.org/10.1038/s44319-024-00124-7. URL https://www.embopress.org/doi/abs/10.1038/s44319-024-00124-7.

[19] Cheng Peng, Jiayu Shang, Jiaojiao Guan, Donglin Wang, and Yanni Sun. Viralm: empowering virus discovery through the genome foundation model. *Bioinformatics*, 40(12):btae704, 11 2024. ISSN 1367-4811. doi: 10.1093/bioinformatics/btae704. URL https://doi.org/10.1093/bioinformatics/btae704.

[20] Bruce J. Wittmann, Tessa Alexanian, Craig Bartling, Jacob Beal, Adam Clore, James Diggans, Kevin Flyangolts, Bryan T. Gemler, Tom Mitchell, Steven T. Murphy, Nicole E. Wheeler, and Eric Horvitz. Strengthening nucleic acid biosecurity screening against generative protein design tools. *Science*, 390(6768):82–87, 2025. doi: 10.1126/science.adu8578. URL https://www.science.org/doi/abs/10.1126/science.adu8578.

[21] Alida Zárate, Lorena Díaz-González, and Blanca Taboada. Virdetect-ai: a residual and convolutional neural network–based metagenomic tool for eukaryotic viral protein identification. *Briefings in Bioinformatics*, 26(1):bbaf001, 01 2025. ISSN 1477-4054. doi: 10.1093/bib/bbaf001. URL https://doi.org/10.1093/bib/bbaf001.

[22] Weizhong Li and Adam Godzik. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics (Oxford, England)*, 22(13):1658–1659, July 2006. ISSN 1367-4803. doi: 10.1093/bioinformatics/btl158.

[23] G. S. Miller and R. Fuchs. Post-processing of BLAST results using databases of clustered sequences. *Bioinformatics*, 13(1):81–87, February 1997. ISSN 1367-4803. doi: 10.1093/bioinformatics/13.1.81. URL https://doi.org/10.1093/bioinformatics/13.1.81.

[24] Jonathan Feldman and Tal Feldman. Resilient Biosecurity in the Era of AI-Enabled Bioweapons, August 2025. URL http://arxiv.org/abs/2509.02610. arXiv:2509.02610 [q-bio].

[25] Tal Feldman and Jonathan Feldman. Opinion | AI just created a working virus. The U.S. isn't prepared for that. *The Washington Post*, September 2025. ISSN 0190-8286. URL https://www.washingtonpost.com/opinions/2025/09/25/artificial-intelligence-advance-virus-created/.

[26] Nicole N. Thadani, Sarah Gurev, Pascal Notin, Noor Youssef, Nathan J. Rollins, Daniel Ritter, Chris Sander, Yarin Gal, and Debora S. Marks. Learning from prepandemic data to forecast viral escape. *Nature*, 622(7984):818–825, October 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06617-0. URL https://www.nature.com/articles/s41586-023-06617-0. Publisher: Nature Publishing Group.

[27] Thomas Hayes, Roshan Rao, Halil Akin, Nicholas J. Sofroniew, Deniz Oktay, Zeming Lin, Robert Verkuil, Vincent Q. Tran, Jonathan Deaton, Marius Wiggert, Rohil Badkundri, Irhum Shafkat, Jun Gong, Alexander Derry, Raul S. Molina, Neil Thomas, Yousuf Khan, Chetan Mishra, Carolyn Kim, Liam J. Bartie, Matthew Nemeth, Patrick D. Hsu, Tom Sercu, Salvatore Candido, and Alexander Rives. Simulating 500 million years of evolution with a language model, July 2024. URL https://www.biorxiv.org/content/10.1101/2024.07.01.600583v1. Pages: 2024.07.01.600583 Section: New Results.

[28] Jaspreet Pannu, Doni Bloomfield, Robert MacKnight, Moritz S. Hanke, Alex Zhu, Gabe Gomes, Anita Cicero, and Thomas V. Inglesby. Dual-use capabilities of concern of biological AI models. *PLOS Computational Biology*, 21(5):e1012975, May 2025. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1012975. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1012975. Publisher: Public Library of Science.

[29] Marian Huot, Pierre Rosenbaum, Cyril Planchais, Hugo Mouquet, Rémi Monasson, and Simona Cocco. Generative model of sars-cov-2 variants under functional and immune pressure unveils viral escape potential and antibody resilience. *bioRxiv*, 2025. doi: 10.1101/2025.05.12.653592. URL https://www.biorxiv.org/content/early/2025/05/13/2025.05.12.653592.

[30] Brian L. Hie, Varun R. Shanker, Duo Xu, Theodora U. J. Bruun, Payton A. Weidenbacher, Shaogeng Tang, Wesley Wu, John E. Pak, and Peter S. Kim. Efficient evolution of human antibodies from general protein language models. *Nature Biotechnology*, 42(2):275–283, February 2024. ISSN 1546-1696. doi: 10.1038/s41587-023-01763-2. URL https://www.nature.com/articles/s41587-023-01763-2. Publisher: Nature Publishing Group.

[31] Samuel H. King, Claudia L. Driscoll, David B. Li, Daniel Guo, Aditi T. Merchant, Garyk Brixi, Max E. Wilkinson, and Brian L. Hie. Generative design of novel bacteriophages with genome language models. *bioRxiv*, 2025. doi: 10.1101/2025.09.12.675911. URL https://www.biorxiv.org/content/early/2025/09/17/2025.09.12.675911.

[32] Leyma P. De Haro. Biosecurity Risk Assessment for the Use of Artificial Intelligence in Synthetic Biology. *Applied Biosafety: Journal of the American Biological Safety Association*, 29(2):96–107, June 2024. ISSN 1535-6760. doi: 10.1089/apb.2023.0031. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11313549/.

[33] Evolutionary Scale faqs. URL https://forge.evolutionaryscale.ai/faq.

[34] Renan Chaves de Lima, Lucas Sinclair, Ricardo Megger, Magno Alessandro Guedes Maciel, Pedro Fernando da Costa Vasconcelos, and Juarez Antônio Simões Quaresma. Artificial intelligence challenges in the face of biological threats: emerging catastrophic risks for public health. *Frontiers in Artificial Intelligence*, 7, May 2024. ISSN 2624-8212. doi: 10.3389/frai.2024.1382356. URL https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2024.1382356/full. Publisher: Frontiers.

[35] Peter P. Swire. What should be hidden and open in computer security: Lessons from deception, the art of war, law, and economic theory, 2001. URL https://arxiv.org/abs/cs/0109089.

[36] Yi Dong, Ronghui Mu, Gaojie Jin, Yi Qi, Jinwei Hu, Xingyu Zhao, Jie Meng, Wenjie Ruan, and Xiaowei Huang. Building guardrails for large language models, 2024. URL https://arxiv.org/abs/2402.01822.

[37] Douglas M. Fowler and Stanley Fields. Deep mutational scanning: a new style of protein science. *Nature Methods*, 11(8):801–807, August 2014. ISSN 1548-7105. doi: 10.1038/nmeth.3027. URL https://www.nature.com/articles/nmeth.3027. Publisher: Nature Publishing Group.

[38] AlphaFold Server. URL https://alphafoldserver.com/.

[39] Francisco Eiras, Aleksandar Petrov, Bertie Vidgen, Christian Schroeder, Fabio Pizzati, Katherine Elkins, Supratik Mukhopadhyay, Adel Bibi, Aaron Purewal, Csaba Botos, Fabro Steibel, Fazel Keshtkar, Fazl Barez, Genevieve Smith, Gianluca Guadagni, Jon Chun, Jordi Cabot, Joseph Imperial, Juan Arturo Nolazco, Lori Landay, Matthew Jackson, Phillip H. S. Torr, Trevor Darrell, Yong Lee, and Jakob Foerster. Risks and opportunities of open-source generative ai, 2024. URL https://arxiv.org/abs/2405.08597.

[40] America's AI Action Plan. https://www.ai.gov/action-plan.

[41] Framework for Nucleic Acid Synthesis Screening | OSTP, April 2024. URL https://bidenwhitehouse.archives.gov/ostp/news-updates/2024/04/29/framework-for-nucleic-acid-synthesis-screening/.