Opportunities, challenges and responsible innovation using generative artificial intelligence for biological sequence design

Maximilian Sprang

masprang@uni-mainz.de

University Medical Center of the Johannes Gutenberg University https://orcid.org/0000-0002-8438-4747

Hassan Hassan

Johannes Gutenberg University Mainz

Yasha Ektefaie

Broad Institute of MIT and Harvard

Anne Schmieder

TU Dresden

Alix Wasilenko

University of Zurich https://orcid.org/0000-0002-6492-6932

Moritz Hanke

Johns Hopkins University https://orcid.org/0009-0003-9893-8098

Rosa Martín-Peña

Leibniz University Hannover / CAIMed

Rachael Harkness

University of Edinburgh

Jasper Götting

SecureBio

Alan Murphy

Cold Spring Harbor Laboratory https://orcid.org/0000-0002-2487-8753

Johannes Mayer

University Medical Center of the Johannes Gutenberg University

Marco Schäfer

Robert Koch Institute https://orcid.org/0000-0003-3854-6415

André Friedrich

Central Committee for Biosecurity

Kerstin Lenhof

University Medical Center Göttingen / CAIMed https://orcid.org/0000-0003-1311-0367

Sophia Rudorf
Institute of Cell Biology and Biophysics https://orcid.org/0000-0001-7653-3733
Gigi Gronvall
Johns Hopkins Center for Health Security https://orcid.org/0000-0003-2514-146X
Jan Kellmann
Philipps-Universität Marburg
Simon Anders
Heidelberg University https://orcid.org/0000-0003-4868-1805
Berkan Aysan
Independent Researcher
Nicole Saverschek
Stiftung der Deutschen Wirtschaft (sdw)
Kathrin Schuldt
Bernhard Nocht Institute for Tropical Medicine
Marc Güell
ICREA Research Professor at Pompeu Fabra University https://orcid.org/0000-0003-4000-7912
Katharina Ladewig
Robert Koch Institute
Georges Hattab
Robert Koch Institute https://orcid.org/0009-0006-7349-4699

Perspective

Keywords:

Posted Date: June 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9617635/v1

Additional Declarations: There is **NO** Competing Interest.

Opportunities, challenges and responsible innovation using generative artificial intelligence for biological sequence design

Maximilian Sprang[1,2], Hassan Ahmed Hassan[1,3], Yasha Ektefaie[4,5], Anne Schmieder[6,7,8], Alix Wasilenko[9], Moritz Hanke[10], Rosa E. Martín-Peña[11,12], Rachael Harkness[13], Jasper Götting[14], Alan Murphy[15], Johannes U Mayer[1,2], Marco Schäfer[16], André Friedrich[17], Kerstin Lenhof[18,19], Sophia Rudorf[20], Gigi Gronvall[21], Jan Kellmann[17,22], Simon Anders[23], Berkan Aysan[24], Nicole Saverschek[25], Kathrin Schuldt[26], Marc Güell[16], Katharina Ladewig[16,28,+], and Georges Hattab[16,29,+]

[1]Department of Dermatology, University Medical Centre of the Johannes Gutenberg University Mainz, Mainz, Germany
[2]Institute of Quantitative and Computational Biosciences (IQCB), Johannes Gutenberg University Mainz, Mainz, Germany
[3]DeOxy Tech Ltd, Manchester, UK
[4]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
[5]Broad Institute of MIT and Harvard, Cambridge, MA, USA
[6]ScaDS.AI Dresden/Leipzig – Center for Scalable Data Analytics and Artificial Intelligence, Dresden/Leipzig, Germany
[7]Institute for Drug Discovery, Faculty of Medicine, Leipzig University, Leipzig, Germany
[8]Institute of Protestant Theology, Technical University Dresden, Dresden, Germany
[9]Institute of Biomedical Ethics and History of Medicine (IBME), University of Zurich, Zurich, Switzerland
[10]Johns Hopkins Center for Health Security, Baltimore, MD, USA
[11]CAIMed – Lower Saxony Center for Artificial Intelligence and Causal Methods in Medicine, Hannover, Germany
[12]CELLS – Centre for Ethics and Law in the Life Sciences, Leibniz University Hannover, Hannover, Germany
[13]Baillie-Gifford Pandemic Science Hub, University of Edinburgh, Edinburgh, UK
[14]SecureBio, Cambridge, MA, USA
[15]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA
[16]Centre for Artificial Intelligence in Public Health Research (ZKI-PH), Robert Koch Institute, Berlin, Germany
[17]Zentrale Kommission für Biologische Sicherheit, Bundesministerium für Verbraucherschutz und Lebensmittel, Germany
[18]University Medical Center Göttingen, Göttingen, Germany
[19]CAIMed – Lower Saxony Center for Artificial Intelligence and Causal Methods in Medicine, Göttingen, Germany
[20]Institute for Cell Biology and Biophysics, Leibniz University Hannover, Hannover, Germany
[21]Department of Environmental Health and Engineering, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
[22]Center for Synthetic Microbiology (SYNMIKRO), Philipps-Universität Marburg, Marburg, Germany
[23]BioQuant Center, University of Heidelberg, Heidelberg, Germany
[24]Independent Researcher
[25]Stiftung der Deutschen Wirtschaft (sdw) gGmbH
[26]Infectious Disease Surveillance and Control Lab, Bernhard Nocht Institute for Tropical Medicine (BNITM), Hamburg, Germany
[27]Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain
[28]Medical University Lausitz – Carl Thiem (MUL-CT), Cottbus, Germany
[29]Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin, Germany
[+] *These authors jointly supervised this work.*

May 26, 2026

Abstract

The convergence of artificial intelligence (AI) and biotechnology is creating unprecedented capabilities for designing biological sequences, ranging from proteins and nucleic acids to entire genomes. However, significant barriers hinder the translation of generative AI models into real-world biological applications, including biased and fragmented training data, opaque model architectures that resist interpretation, and the increasing disparity between the speed of computational sequence generation and the throughput of experimental validation. As these barriers are overcome, the dual-use potential of these technologies is simultaneously expanded: the same advances that enable therapeutic design also lower the threshold for misuse, ranging from the

generation of new pathogens to circumventing existing biosecurity screening. In order to address this issue, the Wübben Sandpit on Generative AI for Biological Sequences was convened, bringing together researchers from the fields of molecular biology, AI, biosecurity, ethics and policy. This white paper presents the resulting consensus recommendations, structured as a defence-in-depth framework spanning three domains. Firstly, the governance of sensitive training data should be achieved through risk-stratified access and funding-stage review. Secondly, the governance of models is addressed through lifecycle risk assessment, safety benchmarking and tiered deployment. Thirdly, the governance of nucleic acid synthesis is strengthened through screening, supply chain accountability and dedicated funding for countermeasure research. Rather than treating security as an impediment to innovation, the framework positions responsible governance as a prerequisite for realising the transformative potential of AI-driven biological design.

1 Introduction

The life sciences are undergoing a fundamental paradigm shift, transitioning from a descriptive discipline to a predictive, engineering-led endeavour. This transformation is driven by the convergence of advanced physical technologies, such as quantum-enhanced sensing, with sophisticated engineering platforms, such as nanopore long-read sequencing and the industrialisation of wet-lab automation. A key outcome of this convergence is the sharp, exponential decline in genomic sequencing costs, which has triggered a data explosion that far exceeds the analytical capacity of traditional heuristic and statistical methods.

This deluge of information is at the heart of the tensions and opportunities of modern biotechnology, and is the primary impetus for the Generative Artificial Intelligence (GenAI) for Biological Sequences' Wübben Sandpit: the dual-use nature of Artificial Intelligence (AI). While AI offers unprecedented capabilities for biological discovery, the same pattern-recognition architectures that decode life's complexity also pose significant risks to global biosecurity and the integrity of scientific practice. Navigating this landscape requires striking a balance between fostering rapid innovation and implementing robust safeguards against the accidental or intentional misuse of biological data.

1.1 The Promise of Biological AI

Before addressing the regulatory and technical hurdles identified by the sandpit participants, it is important to consider the potential transformative impact of these technologies. AI-driven models are set to transform our approach to global health and ecological stewardship. By enabling the automated design of highly specific bacteriophages and bespoke biomolecules, AI could provide a definitive solution to the escalating crisis of antimicrobial resistance (AMR).

Furthermore, the integration of AI into the life sciences extends to:

- Precision ecology: utilising genome-based modelling to optimise conservation strategies and predict ecosystem responses to climate change.
- Predictive diagnostics: moving from reactive medicine to proactive health management through the analysis of individual multi-omic profiles.
- Rational drug discovery: accelerating the identification of novel therapeutics by simulating complex biomolecular interaction networks in silico.

However, the realisation of this 'bio-intelligent' future is currently obstructed by systemic bottlenecks. To bridge the gap between theoretical modelling and real-world application, three interacting categories of translational barriers must be addressed at the AI-life science interface:

- Data scarcity and quality: there is an urgent need for large-scale, high-fidelity and annotated datasets that accurately represent the full stochasticity of biological sequence space. Unlike in previous eras of bioinformatics, the 'AI-first' approach requires a level of data granularity and interoperability that currently does not exist. This has sparked a global race for access to curated biological 'ground truth' data, often at the expense of open science principles.
- Model interpretability: While contemporary deep learning architectures exhibit remarkable predictive power, they remain largely 'black boxes'. In high-stakes domains, such as human clinical

trials or the management of strategic ecologies, the inability to understand the logic behind a model's output poses a significant epistemic risk. Therefore, developing eXplainable AI (XAI) for biology is not just a technical preference, but a safety requirement.

- The synthesis-verification gap: the velocity of *in silico* sequence generation currently outpaces the throughput of physical laboratory synthesis and functional validation. This creates an imbalance where generative models can design sequences that bypass current biosecurity screening protocols, creating a 'security vacuum'. This mismatch slows the iterative cycle of testing and verification, hindering the ability of researchers and regulators alike to keep pace with algorithmic advancements.

The following sections provide a detailed examination of these translational challenges, followed by a comprehensive analysis of the associated dual-use risks. We conclude with a set of strategic recommendations formulated by attendees of the Sandpit to guide the responsible integration of AI into the biological sciences. While recent work has addressed individual aspects of this challenge, including tiered data governance [BBC+26] and a review of the risk landscape [HAİ+26], the present work provides the first integrated, expert consensus spanning the full biological risk chain from data through models to synthesis. The Sandpit brought together an international team of 20 participants including the organisers and two moderators from more than 20 institutions, spanning molecular biology, artificial intelligence, biosecurity, ethics, policy, and public health, as reflected in the interdisciplinary authorship of this work.

Table 1: The consensus of the Wübben Sandpit attendees on recommendations for mitigating the bio-risk of AI in genomic sequence design. The green background refers to recommendations for mitigating risks in the data space, the blue background to recommendations for mitigating risks in the model space, and the orange background to recommendations for mitigating risks in the synthesis and translation space.

Overview of recommendations
Initiate an expert consensus process to classify dual-use pathogens.
Create initiatives to raise scientists' awareness of dual-use research and datasets.
Establish a tiered data access system with proportionate access controls.
Introduce funding-stage reviews and requirements for projects that generate dual-use data.
Establish a structured risk-benefit assessment process throughout the AI model lifecycle.
Develop safety frameworks that combine *in silico* evaluation with controlled experimental validation.
Create tiered access platforms to control access to models and allow independent risk assessments.
Strengthen screening, detection and registries across nucleic acid synthesis providers.
Implement Know Your Customer (KYC) rules and Radio Frequency Identification (RFID) tracking for critical components in the supply chain.
Establish or strengthen clear governance for nucleic acid synthesis screening.
Create a sector-level insurance fund to support research into countermeasures and risks.

2 Translational barriers

To truly benefit from GenAI for biological sequence design, we must address the translational barriers that currently prevent models from delivering on their promise. These barriers can be grouped into three interacting categories: **Data**, relating to the quantity, quality and accessibility of training data; **Models**, relating to verification, interpretability and the institutional environment; and **Synthesis**, relating to the physical realisation of AI-designed sequences. The first two categories are discussed in detail below, while synthesis barriers are primarily considered to be engineering challenges.

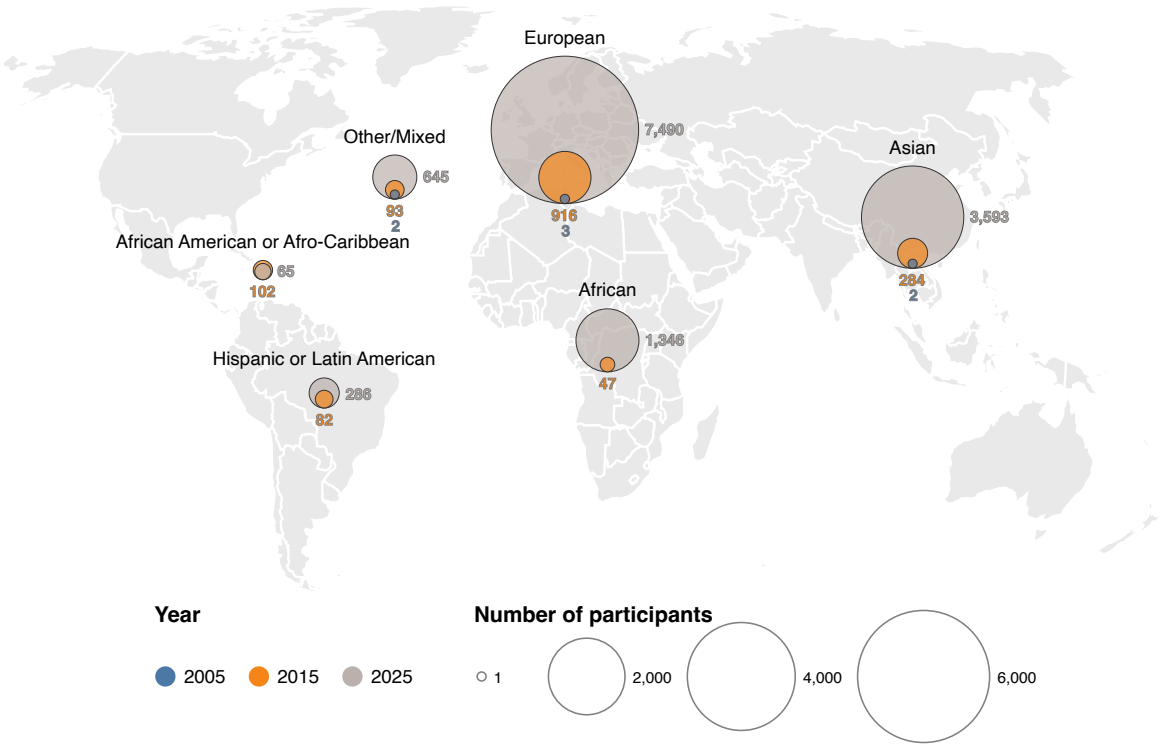

Figure 1: Number of individuals included in genomic studies in 2005 and 2025, stratified by reported ancestry. The dataset used here is from the NHGRI-EBI GWAS Catalog: a curated collection of all human genome-wide association studies (GWAS). Although we are seeing an increase in the representation of non-European genomes, the majority of these are still European. Furthermore, most of the Asian class consists of Chinese and Japanese genomes.

2.1 Data

The challenge

Data is the driving force behind every generative model. For systems that design biological sequences, such as proteins, nucleic acids or entire genomes, the training corpus must capture enough of nature's combinatorial diversity to enable meaningful generalisation. However, current biological datasets not only describe nature, but also encode historical, economic and taxonomic priorities that determine which organisms are sequenced and which are not. Even where relevant sequence–function data exists, it is often inaccessible due to legal constraints, institutional incentives and technical fragmentation. Where data is accessible, it is often noisy, poorly labelled and produced much more slowly than models can process it. Unless these structural imbalances in *bias*, *access*, *quality*, and the *physical–digital feedback loop* are actively addressed, generative models will inherit and amplify them at scale.

Bias enters at multiple levels. For example, sequence databases are dominated by organisms of commercial, medical or model-organism interest, leaving vast swathes of microbial, environmental and invertebrate diversity poorly represented [TGB+17, Cho24, AL23]. This taxonomic and species bias directly constrains the design space available to generative models. For example, a protein language model trained predominantly on mammalian and bacterial sequences would struggle to propose functional designs inspired by extremophilic enzymes or deep-sea viral proteins. Survivorship and filtering bias compound the problem, as quality control pipelines routinely discard fragmented, divergent or poorly annotated sequences—systematically removing the very biology that could expand a model's generative capacity [DS24]. Temporal bias is particularly relevant for sequence design targeting rapidly evolving organisms. The spread of antimicrobial resistance genes among bacteria since the introduction of antibiotics and viral quasi-species dynamics during an outbreak are often not cap-

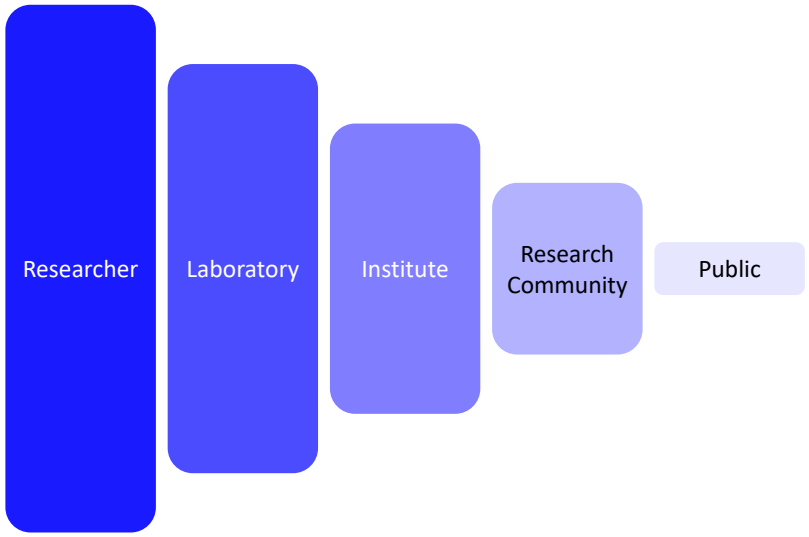

Figure 2: The process of data sharing dilutes ownership of the data from the people who generated it to the public. This leads to the segregation of datasets. This refers to data withheld not due to biosecurity concerns, but due to a reluctance to share data under systemic pressure. The problem is not just withheld raw data: often, data is shared but the annotation is not fully accessible, which can introduce bias when using the dataset.

tured in static database snapshots. This means that models may learn outdated sequence–function relationships [BTWH18, oGSoA20]. Functional annotation bias further distorts what models can learn: annotations cluster around well-studied pathways and commercially relevant functions, while the vast majority of predicted open reading frames are characterised only by homology—or not at all [BDCuR+13, MRPCC+25].

Even when relevant data exists, siloing prevents it from being accessible 2. Citation-based promotion systems, grant-cycle pressures and domain-specific software stacks can create a culture of isolation, often discouraging collaboration. Labels and curated datasets frequently hide behind institutions or individuals who are unwilling to share due to uncertainty over attribution. This even extends within organisations, where data and responsibilities are frequently not exchanged across teams. The variety of software used to store and process biological datasets creates implicit technical barriers that make pooling data across groups difficult. Intellectual property considerations create their own distortions. As GenAI can produce functional homologues, different sequences with equivalent functions, traditional sequence-based patents are becoming less meaningful. The resulting uncertainty over IP defensibility affects what data is generated and shared in the first place.

Generative sequence design faces challenges beyond access issues, data quality and labelling. While unlabelled sequence data is abundant, functionally annotated data linking sequence to a measurable phenotype is scarce and expensive to produce. The labels that exist are often proxies: a binding-assay score stands in for therapeutic efficacy, and a fluorescence readout substitutes for enzymatic activity under process conditions. The cost of extracting accurate labels for individual projects can exceed the benefit they provide in isolation, even when such labels would be valuable collectively. Aleatoric and epistemic uncertainty, technical noise from instrumentation and sequencing facilities, library preparation artefacts and batch effects pervade biological measurements, directly translating into noisy fitness landscapes for generative models [SANF22, GYW22, SMANF24, DAİH26].

Finally and most critically, the physical–digital feedback loop has been broken. *In silico* sequence modelling and generation techniques vastly outpace laboratory synthesis and validation. For example, a generative model can propose hundreds of thousands of candidate sequences in hours, whereas producing even a modest validation panel in the laboratory takes weeks or months and substantial resources.

2.2 Models

The challenge

Although GenAI models operate at digital speed and produce vast numbers of candidate sequences, they remain bound to a slow, costly and error-prone biological reality. This creates a widening 'reality gap' in which model outputs cannot be validated, falsified or iteratively improved at the same pace at which they are produced. At the same time, many models operate as opaque statistical systems, making their internal reasoning difficult to explain to users or regulators. This undermines trust and accountability. The institutional environment surrounding biological research is not suited to GenAI. This is because it is designed for human-led, sequence-based innovation and does not accommodate the speed or logic of GenAI. This is true of the regulatory frameworks, intellectual property regimes and funding structures that surround it. Without progress in *verification*, *interpretability*, and *governance*, models risk becoming increasingly divorced from biological truth while seemingly successful.

As in the data domain, digital-physical decoupling is the defining challenge for generative sequence design, particularly with regard to validation. Although *in silico* generation is fast, parallel and continuous, wet-lab synthesis and characterisation are linear, expensive and time-consuming, creating a 'verification bottleneck'. Models generate enormous candidate pools, yet receive validation data for only a tiny fraction of these, which starves them of the negative examples needed for iterative improvement. This asymmetry is exacerbated by the implicit assumption of physicality, namely the widespread belief that *in silico* work is only valuable once validated in a laboratory setting. Additionally, the scarcity of reliable computational validation frameworks, such as digital twins or high-fidelity simulators, hinders the development of more autonomous models.

Current evaluation practices for biological generative models are fragmented and incomplete; multiple benchmark suites exist [MTH+23, PSW+24]. However, they rarely collaborate, rely on arbitrary downstream tasks and prioritise quality over power. Metrics are misaligned with goals, raising questions of optimisation. Understanding why a model proposes specific residues or motifs, whether it has learned genuine biophysical constraints or merely statistical regularities, is essential for responsible deployment. This highlights another translational barrier. Models exist on a spectrum from interpretable to opaque, and the lack of insight into their inner workings hinders adoption in high-stakes settings. Users need causal explanations centred on their background and role; a statistical prediction without a causal or mechanistic rationale is insufficient when the output is destined for synthesis and deployment in living systems [IDH25, Hat26]. Without interpretability, resources for subsequent experiments are allocated haphazardly. The risk of automation bias exacerbates this issue: as operational pressures push teams towards greater reliance on AI-generated designs, the expertise of humans is at risk of being marginalised. In systems capable of human language, the anthropomorphisation of AI can lead to a language model's confident output being treated as genuine biological understanding. This results in capabilities being overestimated and outputs being under-scrutinised [dARDT+25]. This can result in accountability becoming diffuse, with humans potentially blaming AI for decisions in order to avoid responsibility. Alternatively, there is a 'problem of many hands', where responsibility for an error is unclear.

Beyond the technical challenges of interpretability and model evaluation, the institutional environment introduces systemic frictions that hinder the translation of AI-driven discovery into tangible outcomes. Central to this is a growing misalignment in intellectual property (IP) frameworks. Historically, biological IP has been based on sequence uniqueness. However, the emergence of functional homologues, sequences that are structurally distinct yet biologically equivalent, challenges the long-term viability of sequence-based patent models. The current lack of consensus on how to formalise 'function-based' IP creates a regulatory vacuum that could undermine the R&D incentive structures that are fundamental to the biopharmaceutical industry. This challenge is particularly acute for non-coding genomic regions, where evolutionary constraints primarily affect regulatory output rather than specific nucleotide identities. In such regions, sequence-level similarity is an unreliable indicator of functional equivalence. This means that generative models could produce regulatory elements with the same function but with little sequence overlap. This further undermines sequence-based patent frameworks [NMRK26].

The institutional lag also affects the operational and regulatory sectors. Existing oversight regimes,

such as export controls for DNA synthesizers, were primarily designed for the movement of physical goods rather than the frictionless exchange of digital design workflows. This results in a kinetic asymmetry: while in silico development proceeds at digital speeds, physical validation is still held back by traditional compliance hurdles. Furthermore, these regulatory requirements often impose a disproportionate administrative burden on start-ups and smaller laboratories, potentially favouring established companies and reducing the diversity of actors within the innovation ecosystem.

Finally, current funding paradigms exacerbate the divide between digital proposals and physical reality. Granting bodies often prioritise developing novel algorithmic architectures over the essential yet capital-intensive work of data curation and experimental validation. Although computational modelling is comparatively cost-effective, closing the 'design-build-test' loop requires significant investment in 'wet' infrastructure; a resource for which dedicated funding remains disproportionately scarce. Until incentive structures reward translational infrastructure and cross-disciplinary validation consortia as much as methodological novelty, the gap between AI-generated hypotheses and biological verification is likely to continue expanding.

A third category of barriers lies in the synthesis of AI-designed sequences, the physical realisation of *in silico* designs in the laboratory. These barriers are primarily engineering-related, such as current limitations in the length and fidelity of oligonucleotide synthesis, the assembly of large constructs, and the *throughput* of functional screening pipelines. While we acknowledge these challenges, they are not unique to GenAI and do not constitute barriers to the deployment of artificial intelligence in this field. Rather, they constrain the speed at which the *design–build–test–learn* cycle can operate. As such, synthesis is more productively discussed in the context of dual-use risk mitigation (C.f., Section 6), where the governance of synthesis providers and supply chains plays a critical role in the defence-in-depth framework proposed by the sandpit.

3 Attempting to overcome translational barriers could raise concerns about the dual use of GenAI for genomic sequences

When we reduce these barriers, we increase the uptake and use of models, as well as the potential risks. Therefore, while we work to minimise these barriers, we must also consider how to mitigate and govern the emerging risks associated with the growing capabilities and dissemination of GenAI in biology.

In this section, we provide an overview of the dual-use risks associated with these models, from data collection and trained models to sequence synthesis and potential deployment in physical and biological systems. Risk is generated at the data level. Any information about sequence-to-function data, particularly that which includes pathogenic or toxic sequences, can be used by malicious actors to train models that can design sequences with known functions. When using pre-trained models, a large number of annotated sequences is not required. In their 2025 study, King et al. used 15,000 bacteriophage ($\Phi X174$) sequences to generate new sequences with as little as 40% nucleotide sequence identity [KDL+25]. This challenges one of the significant barriers to using public data, as good annotation is needed and encouraged, despite the potential for dual use. Funding bodies and researchers must play an active role in deciding which data should be withheld from the public if it includes dangerous sequences or sequence information. However, genomes of viruses that are considered high risk for misuse as biological weapons and those that have led to global pandemics are already in the public domain [PSW+22]. Another issue that will be mentioned repeatedly across different domains is the circumvention of security safeguards, as pathogenic data could be used to adversarially fine-tune pre-trained models that did not contain pathogens in their training corpus [BHM+25a].

The general availability of genomic data, despite its potential risks, lowers the barrier to training and designing one's own architectures for all kinds of actors. This enables small laboratories to conduct research, as well as potentially enabling malicious actors. The availability of huge models with potential frontier capabilities works similarly. As with the data stage, important actors in the modelling stage include researchers who open-source their models and state actors who provide the computing infrastructure to train them. While it is positive that these models are open-sourced, as this encourages collaboration and increases the speed at which scientific innovation occurs, this must be weighed against potential threats [RHC+, GFG+24]. Models with frontier capabilities can generate viable genomes and potentially assist in designing gain-of-function mutations or developing new-to-nature sequences. These could be weaponised or have negative environmental impacts by generating

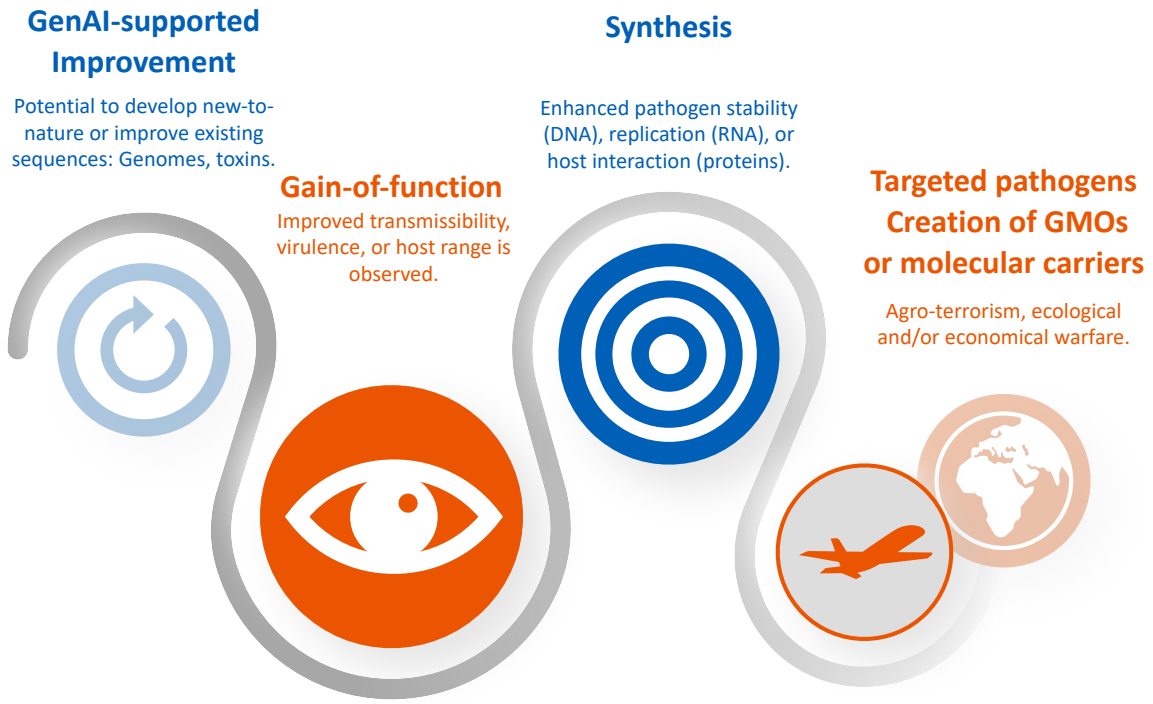

Figure 3: This infographic illustrates a series of events across multiple stages of the risk chain mentioned in Table 2 that culminate in a biohazardous outcome affecting the environment. The five main events relevant for biosequence modelling are shown. When used independently, they cause little to no harm. However, when they occur in sequence, they emerge as an insidious plan that threatens human health and has an ecological impact. 1. Generative AI (GenAI) is transforming synthetic biology by enabling the design of novel 'new-to-nature' sequences, including complete genomes and proteins such as toxins, which often exceed the complexity of naturally evolved organisms. These tools are trained on vast genetic datasets in order to learn the 'grammar' of life and create functional sequences that can bypass existing biosecurity screens. 2. Gain-of-function (GOF) research involves altering an organism's genome to enhance its existing abilities or confer new ones. These new abilities may include improved transmissibility, virulence, or host range. Examples of observed GOF at different molecular levels include modifying DNA/genomes to alter pathogen stability or virulence, engineering RNA to enhance replication or immune evasion, and mutating proteins to strengthen host interactions or expression patterns. 4. Synthesising these sequences is how they enter the real world. While most synthesis providers have implemented safeguards against misuse, these are not mandatory and can vary considerably between countries and providers. Furthermore, GenAI can be exploited to circumvent these safeguards. For instance, new-to-nature sequences are unlikely to be recognised by methods designed to detect known pathogens. Models can also be used to conceal a gain-of-function mutation by shuffling the sequence or employing alternative codons to encode a target protein. 5. Creation of GMOs or molecular carriers, which are then disseminated into the environment by a malicious actor. Depending on the malicious actor's scope and goals, these could be targeted pathogens or pandemic agents.

Table 2: Dual-use research of concern has many layers that connect with the biological risk chain at multiple points. We distinguish between actors that push research towards potential risks and the applications that harbour risks. Furthermore, we characterise the domains in which risks are created and can potentially be mitigated as follows: Data: focusing on the inherent risk in some biological data and its impact on the training and use of downstream models. Models: concerned with how models are trained and how a trained model can be used for malicious actions. Synthesis: bringing the designed biomolecule to life in a dish. Weaponisation: considering how the malicious use of models and synthetic biology could manifest.

	Data	**Model**	**Synthesis**	**Weaponisation**
Actors	Researchers publishing sequence-to-function information, state actors and funding bodies requiring open access data	Researchers building models and sharing weights/code publicly, state actors providing infrastructure	Process engineers developing cost-effective methods, synthesis providers lacking security, CROs and vendors in differing governance contexts	Lone wolves / terrorist groups, biohackers, organized crime, rogue states
Applications	Circumvention of existing security (adversarial fine-tuning), annotation can facilitate dual use	Lowering barriers to training models, potential to develop new-to-nature sequences (toxins, genomes), circumvention of existing security (design and synthesis of functional homologues)	DNA/RNA synthesis, cell-free expression systems, creation of GMOs, gene drives	Agroterrorism, ecological/economical warfare, targeted pathogens for specific populations, new pandemics

gene drives that outcompete natural species, for example in plants or insects.

Such systems must first be synthesised in order to be introduced to the environment. Those involved include process engineers, biochemists, commercial synthesis providers, and contract research organisations (CROs). The latter are particularly influenced by the governance framework in which they operate, which varies greatly between countries. While synthesis companies do check sequences sent to them for synthesis using various technical measures, this is not mandatory in many countries (see the IBBIS global synthesis map at globalsynthesismap.bio for more information). Returning to the modelling stage, models can also help malicious actors to circumvent these safeguards by designing functional homologues.

Finally, we should consider the actual weaponisation of these methods, particularly the gain-of-function and de novo generation capabilities of frontier AI models in biology. These models could be used to design new toxins. They could also be used to modify or enhance infectious bioagents. This could potentially lead to new epidemics or pandemics. Furthermore, the models could be used to create targeted pathogens. These would attack only a specific plant species. This could be exploited in the context of agroterrorism. Such attacks could be intended to cause environmental or ecological damage. Perpetrators could range from lone individuals to small groups to state actors. While the former currently lack the capabilities to do so, GenAI systems could lower this barrier at multiple points in the biological risk chain. As aforementioned, they can assist with the design and circumvention of safeguards, and they may also help with experimental design and laboratory work [QHY+25, dARDT+25, ZCQ+25, WHZ+26].

4 Strategic recommendations for mitigating dual-use research of concern risks in genomic GenAI

Throughout the Sandpit, a consensus emerged that mitigating the risks of Dual-Use Research of Concern (DURC) inherent in the computational design of genomic sequences requires a defence-in-depth architecture. Rather than relying on a single point of failure, the proposed framework takes a layered approach that covers all parts of the risk chain that have contact with civil academic research, from the initial curation of data and training of models to the final biological synthesis of sequence outputs.

A key principle of these recommendations is that governance must be dynamic and iterative. As Genomic GenAI capabilities advance at an unprecedented rate, classification systems and access controls must undergo regular technical audits. Participants identified 'upstream' levers, specifically funding mandates and data-sharing agreements, as the most effective instruments for embedding responsible practices before high-risk sequences reach the synthesis stage.

Although the objective is protective, these recommendations acknowledge certain operational limitations to ensure that security measures do not inadvertently impede legitimate genomic research. To prevent a 'bottleneck effect' on public health research or pandemic preparedness, we propose a tiered-access model. Moving away from a binary 'open vs. closed' data philosophy, we introduce a risk-stratified environment. For data there was a recent publication by Bloomfield et al in 2026 [BBC+26], that divides data in four tiers:

BDL-0 (Open Access) All biological data not named in higher tiers. No access controls imposed.

BDL-1 (Managed Access) Data anticipated to enable an AI system to learn general patterns of eukaryote-infecting viral genome and protein composition. Requires registered account, government-issued ID, agreement to terms of use; host must log usage and flag concerning activity.

BDL-2 (Identity Accreditation) Functional data anticipated to enable an AI system to learn properties of pandemic-capable viruses (environmental stability, host range, susceptibility, diagnostic evasion). Requires institutional affiliation, vouching institution, identity verification, and screening against government bad-actor lists.

BDL-3 (Use Approval) Functional data anticipated to enable an AI system to learn transmissibility, virulence, immune evasion, and countermeasure resistance of human-infecting viruses. Requires use-case screening; data accessed only via certified trusted research environment (TRE); prepublication model risk assessment if public health benefit cannot be demonstrated.

BDL-4 (Result Pre-screening) Data anticipated to enable design of pandemic-capable viral variants with enhanced transmissibility, virulence, or immune evasion. All requirements of lower tiers plus mandatory government review of results before publication; government retains right to restrict access to model code, weights, and outputs.

With such an approach governance can be a facilitator of secure innovation. The shift towards structured oversight in genomics is supported by successful models from related high-stakes biological fields, as well as by successes in data management, software security, and industrial manufacturing (see overarching justifications below; [BBC+26, WAB+25, CB26]).

Overarching justifications

- Data tiering (the provenance precedent): The UK Biobank and other similar repositories show that providing controlled access to sensitive genomic information improves data standardisation and documentation. This structured approach increases the resource's utility for the global research community while maintaining strict privacy and security safeguards.
- Model Red-Teaming (The AI Safety Precedent): In the broader AI domain, 'gated release' and adversarial red-teaming have become standard practice for high-capability models (e.g., Frontier Model Forum protocols). Applying these practices to GenAI in genomics enables the identification of potential biological risks, such as a model's capacity to design 'stealth' pathogens, before the weights are made publicly accessible.
- Synthesis Screening (The Physical-Gatekeeping Precedent): The International Gene Synthesis Consortium (IGSC) has already established a successful, industry-led screening framework for physical DNA orders. Expanding these protocols to include more sophisticated, AI-generated sequence variants ensures that the 'digital-to-biological' bridge is monitored by robust, standardised biosecurity checks.
- Institutional synthesis: Rather than proposing new regulatory bureaucracies, these recommendations are designed to integrate with existing institutional review processes and funding requirements. This approach minimises administrative friction while maximising compliance across the research lifecycle.

Overarching limitations

Risk classification is an inherently difficult problem because an AI's function is to identify patterns. Without knowing which datasets or models could enable harmful activity, there is a risk that restrictions will either be too broad or too narrow. This creates an 'all-or-nothing' dilemma. Furthermore, stringent regulations could restrict valuable research that could significantly benefit public health. They could also erode the openness of science and increase fragmentation across research.
Even without considering the misuse of guardrails by bad-faith actors, access restrictions could create a two-tiered research community or be subject to political pressure. Well-organised malicious actors might establish clandestine mirrors of classified datasets/models. At worst, this could give malicious actors (e.g., state actors developing biological weapons) an advantage over defensive research.

4.1 Data

We have the following recommendations for data: (1), (2), (3), and (4).

1. Standardised pathogen classification: Initiate a multidisciplinary consensus process to define 'dual-use pathogen data', moving away from static lists of agents to functional genomic signatures of virulence.
2. Community Biosecurity Literacy: Launch targeted initiatives to raise awareness among researchers of the dual-use potential of specific genomic datasets, and of their responsibilities as stewards of the data.
3. Tiered Access Architecture: Implement a risk-stratified access system utilising proportionate

controls ranging from open access for foundational data to credentialed verification for sensitive pathogenic sequences.

4. Integrate biosecurity reviews into the funding cycle and require robust risk-mitigation plans for any project intended to generate or curate potential dual-use genomic data.

4.2 The role of data governance in high-fidelity AI-assisted genomics

Data serves as the fundamental engine of genomic GenAI: the combination of raw sequences with functional labels determines the boundaries of the model's predictions and generation. As high-throughput experimental pipelines become more widespread, the volume of functionally annotated data, particularly with regard to protein toxicity, viral tropism and molecular virulence, continues to grow. While this increasing granularity enhances model performance, it also raises the risk of such data being used to generate sequences with harmful biological characteristics.

The technical feasibility of this risk is demonstrated by recent advances in antigenic escape modelling. For instance, deep mutational scanning (DMS) datasets for viral proteins can be employed to fine-tune biological foundation models, significantly enhancing their ability to predict mutations that evade immune detection. Although the automated design of enhanced human infectivity remains a theoretical concern rather than an established capability, the precedent set by antigenic escape shows how targeted fine-tuning can incorporate specialised knowledge into 'safe' base models.

This 'fine-tuning' vulnerability complicates current voluntary oversight. For example, although models such as Evo2 and ESM-3 exclude specific eukaryote-infecting viral data from their primary training sets, well-intentioned researchers can and do re-integrate this data during downstream fine-tuning. When these specialised models are subsequently published under open-science frameworks, they can inadvertently lower the barrier to designing sequences with optimised pathogenic traits [BHM+25a, BHM+25b].

4.3 Recommendation 1: Stimulate discussion towards expert consensus on the classification of dual-use pathogen data

We recommend the creation of a working group similar to the German 'Joint Committee of the German Research Foundation (DFG) and Leopoldina', the National Academies of Sciences' expert committees or the WHO's Advisory Group on Dual-Use Research of Concern should be formed to release a classification of potentially dual-use pathogen data that should be restricted, which can then be widely adopted globally. This group should include experts in molecular biology, AI, public health, national security, biosecurity, policy-making and data security.

Beyond classification, the discussion should also focus on organisational aspects, specifically who should control access and be responsible for classification. Possible stakeholders include data generators (e.g., wet-lab scientists or primary data analysts), database curators, database operators, institutional boards and national/international committees.

Consensus groups must regularly reconvene to update data classification, especially as capabilities advance. Reaching consensus will be difficult as stakeholders often have different views on data types. Explicitly identifying potentially problematic data carries risks: by delineating what is sensitive, we may inadvertently create a roadmap for malicious actors seeking out/exploiting valuable datasets.

4.4 Recommendation 2: Call for researchers to release certain dual-use pathogen data responsibly

As part of the responsible use of data and general data risk assessment, we recommend that research organisations develop internal guidelines and procedures for assessing data risks, and update these guidelines based on risk assessments enabled through GenAI methods. In order to encourage responsible self-governance among researchers, professional associations and funding agencies should incorporate these principles into their codes of conduct (for the latter, c.f., Recommendation 4).

In addition to the consensus group, which must comprise a broad range of relevant scientific expertise, it is essential that the 'regulated' community, that is to say researchers who generate the data under discussion, is given the opportunity to provide feedback on the guidance not to release concerning pathogen data, and ultimately endorses it.

This approach ensures that, in the absence of stricter measures, responsibility for considering responsible data release falls on lower organisational levels: research organisations, institutes or individual groups. However, relying on researchers alone could lead to inconsistent adherence, as they may lack awareness of which data is important or may be under pressure to release data due to grant requirements. Voluntary guidelines also risk allowing dual-use data to remain publicly accessible or to be acquired by malicious actors through deception.

4.5 Recommendation 3: Create a platform with tiered access and usage logging for dual-use pathogen data.

We propose creating a secure, tiered-access platform to host dual-use pathogen data. Such a platform would require user authentication and maintain detailed access logs, thereby enabling accountability and supporting investigations into any potentially problematic use of data. In addition to improving security, a shared repository could save researchers time and resources, facilitate collaboration, reduce duplication of effort and establish consistent standards for the storage and use of sensitive biological datasets.

However, this approach is not without risks. Centralising sensitive data could hinder research progress if access becomes overly bureaucratic or if datasets are insufficiently documented. A single, centralised repository would also present an attractive target for state-level adversaries, and identity theft or credential compromise could undermine the protection offered by access controls. Additionally, maintaining such infrastructure could divert funding away from research activities themselves. Nevertheless, if the alternative to creating and sharing the data on this secure, tiered-access platform is that the data is not generated at all due to dual-use concerns, this could enable important work in public health and pandemic preparedness.

This would not be a binary choice between access restrictions and no access restrictions, and it could include different tiers modelled on BSLs (Biological Safety Levels). To ensure the safety and intended usage of this platform, red teaming, audits and the four-eyes principle are recommended. Such a platform could facilitate standardised data annotation, enabling legitimate researchers to work with dual-use pathogen data more easily. As mentioned above this approach has recently been discussed for data infrastructure and sharing of DRUC relevant data [BHM+25b, BBC+26].

4.6 Recommendation 4: Funding-stage review and potential requirements for DURC pathogen data generation

Funding bodies (e.g., the NIH, the DFG and the CEPI) should review grant proposals involving the generation of data on dual-use pathogens and impose risk mitigation requirements, such as tiered access, or prohibit data generation as a last resort. This approach is fairer to researchers, as it alerts them to potential restrictions before projects begin rather than after data has been generated. It requires the effort outlined in Recommendation 1 of establishing an expert consensus process. It also requires the implementation of Recommendation 3, which involves establishing a tiered-access platform.

The reason this should happen at the funding stage is that controlling dissemination at a later stage (e.g., the journal publication stage) is not feasible in an era of preprints and the rapid, instant dissemination of data. Potential funding requirements for data generation would most often include tiered access to dual-use pathogen data via a standardised trusted research platform, for example. These can include different levels of tiered access depending on how dual-use the data is. Ultimately, it could require that researchers do not generate dual-use pathogen data in the first place.

This would also mean that funding bodies would not require researchers to publish their data openly. Currently, even if researchers have DURC about publishing their data and results and would prefer not to, they are often required to do so under the terms of the grant. This recommendation is also discussed in the "Disseminating In Silico and Computational Biological Research: Navigating Benefits and Risks" workshop of the national academies of Science, Engineering and Medicine of the USA [oSEM+25].

Meta considerations

- These recommendations only apply to newly generated data. It is not possible to scrape existing data from the internet.
- Data controls are a new topic. The data guidelines will ramp up from a voluntary pilot scheme to potentially mandated regulations or comprehensive implementation at a later date.
- Data that could be used to generate pathogens with enhanced pandemic potential should be prioritised, as these could cause disproportionately large-scale disasters for society.
- Global harmonisation and implementation of all recommendations is beneficial.
- Data on dual-use pathogens of concern would only account for a small proportion of all existing data: over 99% of biological data would not be affected.
- Data-level controls form only part of a defence-in-depth approach, and should be considered alongside recommendations at the model and synthesis levels.

5 Model

We propose three mutually reinforcing measures for governing biological sequence models:

1. Lifecycle risk-benefit assessment: Implement structured validation frameworks for the continuous evaluation of model capabilities, from initial training to post-deployment, to identify any hazardous properties that emerge as the model's 'biological reasoning' evolves.
2. Validated safety benchmarking: Develop hybrid evaluation protocols that combine computational red-teaming (adversarial in silico testing) with controlled experimental validation, ensuring that model-generated sequences do not exceed safety thresholds for virulence or toxicity.
3. Tiered model deployment: Establish secure, hosted environments with API-based or 'gated' access for high-capability biological models. This allows for independent risk audits and prevents the unfettered distribution of model weights that could be used for malicious fine-tuning.

5.1 The role of model governance in high-fidelity AI-assisted genomics

The models are at the core of our discourse, as they are the tools that learn patterns in biological data and allow us to translate them into generative frameworks and potentially comprehend more of biology in the process. Their potential is significant, as they enable not only memorisation, but also the generation of new sequences by interpolating between known embeddings or by sampling from an embedding space that extends beyond the scope of existing sequences. This makes them a potentially transformative step in synthetic biology and a potential risk: new-to-nature sequences and purposefully designed biological agents (e.g., epidemic pathogens and AB toxins).

5.2 Recommendations

5.2.1 Recommendation 1: Lifecycle risk–benefit assessments across the model lifecycle

We propose assessing the risk of a bioAI project using a predefined framework that could be adopted more widely. This could be discussed in an expert consensus working group, as mentioned above. Such frameworks are common in other fields, such as wet-lab DURC. A risk–benefit analysis would need to be implemented from the outset, beginning with an assessment of the risks posed by the given dataset, and with the risks and benefits being re-evaluated during model development. For example, a model that generalises well might seem harmless initially when learning only from non-pathogenic sequences. However, if it can really generalise, it may be able to generate dangerous sequences regardless, or

humanise pathogens from other species. This necessitates continuous re-evaluation throughout the project [HAİ+26].

To comprehensively assess the risk of a model, both its input and output need to be considered. The above section covers input (known BSL organisms in the dataset, data modalities, scale and potential risk of sequences, both singularly and at scale); however, output sequences would share evaluation criteria, at least at the level of individual sequences. Additionally, available metadata may have implications for risk, such as connections between data points and patients, and the disease or health context in which they were generated.

A risk–benefit assessment framework would need to be informed by benchmarks that allow the evaluation of both model performance and risk (see the section below). Funding agencies could make such assessments mandatory for the implementation and funding of bioAI projects. The first steps towards risk-scoring biological AI tools have already been taken, such as CLTR's and RNAD's Global Risk Index for AI-enabled Biological Tools [WMDC+25].

One clear downside of this approach is that it places responsibility on model developers who may have a conflict of interest or not take risk seriously depending on their background or personal opinion. One possible solution to this issue is discussed below: the implementation of a tiered-access platform that would act as a neutral final risk assessment and serve as a regulatory force, emphasising the importance of conducting risk-benefit analyses at an early stage. Another potential issue is that it could result in poorer performance of certain models, as datasets or weights may be pruned to achieve lower risk scores. This could discourage model developers from adopting safe practices, as the current goal is to create the 'best-in-class' model.

5.2.2 Recommendation 2: Validated safety benchmarking

A general evaluation gap exists for DNA models, and the same is true of biosequence models in general. While some frameworks exist for evaluating models, these do not build on each other and depend on arbitrary biological downstream tasks. Moreover, the emphasis is on embedding models and the biological meaning of the representations learned by them. However, they are not well suited to evaluating generative power.

Testing the perplexity (a measure of how well a model predicts a given sequence, where lower values indicate greater familiarity) of viral sequences can give a first impression of a model's ability to generalise to unseen data, but is insufficient for the assessment of risk. Security benchmarks and red-teaming are already established in other fields applying AI models, especially in cybersecurity. Paired with interpretability methods, which demonstrate early promise in biological models, for example through the application of sparse autoencoders to extract biologically meaningful features from genomic language model representations [MJD+25], these benchmarks would be able to point out what is risky, as well as why.

Risk-focused benchmarking. Inclusion of risk-scoring at the model and sequence level is an essential component of a risk-focused benchmark. It must be reproducible and model-agnostic, producing datasets primarily for training and testing purposes, with the focus of the output scoring on the post-processed sequence rather than token representations. Ideally, sequence-level risk scores would be combined with model-level performance on risk-related tasks to produce a comprehensive risk score.

Our interest lies not only in the biological performance of representation learning, but also in the generation of sequences, for which the necessity of tasks is essential. Can generated sequences be aligned with known pathogens? For example, does the model have the ability to complete the sequence of a given pathogen? Ideally, there would be a good way to test a model's ability to generate out-of-distribution (OOD) sequences, which could be achieved to some extent by training from scratch on a dataset withheld from known pathogens. However, such a dataset would need to be very large, and this approach would ignore potential risks within the training corpus of a given model. More tractable ways to build adaptable benchmarks for multiple model types and test their ability to generate *de novo* sequences is to build orthogonal models to score the output sequences. These models could use sequence-to-function prediction tasks or functional alignment based on embeddings to see if sequences co-localise with known pathogens in the latent space [WAB+25, WMDC+25].

Interpretability. In this context, interpretability methods would also play a key role. To ensure model alignment, it is necessary to understand what is happening inside the 'black box'. Interpretabil-

ity could be used to identify high-risk sequences by locating pathogenic features, and it could also facilitate the identification of novel sequences that would not be detected by homology or sequence alignment tests. However, it is difficult to quantify interpretability methods, as DNA is not human-readable. Even in NLP, they have relied on large-scale auto-interpretability or manual qualitative evaluation [BTB+23].

We propose mimicking NLP methods that use concept probes to interrogate model internals. A probe is a classifier trained to predict human-level concepts. One advantage of this approach is that probes are straightforward to implement and enable the automated evaluation of model interpretability. For instance, a probe that predicts the probability of a human host could serve as a risk score for a generated sequence. However, the most significant caveat is that the concepts must be chosen carefully to avoid assigning risk (causation) to a correlating signal.

It should be noted that interpretability models can pose a risk themselves. Identifying functional features facilitates the controlled generation of DNA sequences. For example, model outputs could be steered towards viral tropism according to host probabilities. Given the current performance of *de novo* sequence generation, one could argue that model steering poses a more immediate risk.

Wet-lab validation. Although computational benchmarks provide an initial screening, definitive assessment of the hazard posed by a *de novo* generated sequence lies in wet-lab validation. However, this 'final frontier' is fraught with risk and expense, as testing potentially hazardous sequences requires maximum bio-containment protocols. Aside from the physical risks, such testing can inadvertently produce 'misuse blueprints' that provide detailed experimental data proving a sequence's lethality that could be exploited by malicious actors. To mitigate this risk, institutions must implement a tiered pipeline in which only the highest-risk sequences proceed to the laboratory. Even then, researchers should prioritise the use of safe biological proxies. Using attenuated or non-pathogenic organisms representative of the target task enables the biological risk to be assessed without creating a direct roadmap for high-consequence misuse.

Red-teaming. In addition to a general benchmark, red-teaming of some models is needed to identify high-risk capabilities and establish whether a secured model can be compromised by malicious actors.

As this work is labour-intensive and requires manual interaction with the model and code, it cannot be carried out for every published model (the number of which is expected to increase over time). The aim is to guide a model towards a higher risk level and record the barriers overcome in order to improve future versions of these models. This information would inform the final risk assessment and would only be applied to models that already have risk scores in the automated benchmarks. Furthermore, this information should be exchanged between the testing institution or platform and the developers, rather than being published.

An overarching strategy and underlying structure, such as a tiered-access platform, should be implemented for a benchmarking and red-teaming project, as discussed below. A consortium dedicated to this purpose should be set up to discuss and develop an ontology of benchmark types and red-teaming goals. A precedent for this can be found in Transluce's Network of Evaluators forum. Such an ontology would also include if–then commitments: if a model is capable of achieving high-risk levels in the benchmark or being easily steered towards dangerous outputs by red-teaming, it should be assigned to a higher tier. By clarifying this before the stakes are clear for the developers, we reduce the risk of both false positives and risky behaviour going unnoticed.

5.2.3 Recommendation 3: Tiered model deployment

In order to mitigate the risks posed by certain models, such as the generation of new pandemic pathogens, while still realising their potential, we propose establishing a tiered-access platform similar to that outlined in the data section. This platform would host biological AI models, simplifying access for users while implementing risk-dependent access controls and a Know-Your-Customer approach. Models with low risk scores would be fully open source, while those with higher scores would require institutional email addresses and, in stricter cases, institutional verification. Models with the highest associated risk could only be run within the platform as a trusted research environment (TRE). Outputs could be provided upon a reasonable request presented as a proposal.

The risk of a given model would be determined by points (1) and (2) of this section: an in-platform risk–benefit analysis based on automated benchmarking, including red-teaming in edge cases. The

platform would encourage self-testing before upload and perform a final test to determine the model's tier.

Researchers Would benefit from this platform: it would be easy to use and provide access to high-tier models, with even the most high-level models running on secure servers. This would be subsidised by the developers. Researchers could still provide their models to the community. At the same time, the platform would provide publicity, which is currently lacking. Only institutes with large social media and press coverage can push their models, while small labs must rely on adoption. Within the platform, models could also be easily compared, and a central API could allow multiple models for a shared goal. Critically, as emphasised by Carter and Butchello [CB26], such a platform must be designed with equitable access as a core principle, ensuring that verification criteria do not systematically disadvantage researchers at under-resourced institutions or in the Global South. This is a good way to increase the security of AI agents and allow fast innovation, but it is limited by traction and early adoption. There would be associated costs, especially in maintenance and computing. If the project is initially funded by state grants, a plan is necessary to hand over the platform to industry or research partners. Potential partners could advocate open source, such as Hugging Face. Access could be slower, but the benefits outweigh the waiting time. A centralized repository is vulnerable to hacking, but this risk is not lower than the potential benefits.

In conclusion, we recommend combining our suggestions with those on the data side: an expert consensus is needed to create a risk assessment benchmark and framework. Responsible model release must be encouraged, which could be facilitated by a semi-automated, tiered-access platform specialising in biological AI models. Additionally, a review of risk at the funding stage could encourage researchers to assess the risks of their work more effectively.

Table 3: Summary of the tiered access for the data and the model. The data access strategy is adopted from Bloomfield et al. 2026 [BBC+26]. The tiered access for the model encompasses different levels of access for the weights and sometimes also the architecture and output of the model.

Tier 0 **unrestricted**	**Tier 1** **Soft wall**	**Tier 2** **Institutional access**	**Tier 3** **Restricted access**
Open access model, all part of the model are shared	Access is only possible via a secure institutional domain. Weights and architecture can be shared after approval of an institutional email address.	Access is only granted via a proposal from the institution. This must be signed by the institution's director, after which the model weights and architecture can be shared with a verified, secure location such as an institutional cluster with secure compute nodes.	Permission is requested only. A proposal needs to be written to run inference and retrieve the model's outputs. None of the model can be in the public domain. Inference is performed on servers that are specifically installed for the purpose of sharing the model, and proposals must justify the use of the model and the intended generation of data/sequences.

6 Synthesis

We have the following recommendations for biosequence synthesis:

1. Strengthen screening, detection and registries across nucleic acid synthesis providers.
2. Implement KYC rules and RFID tracking for critical components in the supply chain.
3. Establish or reinforce clear governance for nucleic acid synthesis screening.
4. Set up a sector-level insurance fund to finance research into countermeasures and risks.

6.1 Why synthesis governance matters

Although data and models form the computational basis of modern biotechnology, AI-generated sequences only pose a tangible threat once they have been synthesised and introduced into a physical medium. Therefore, synthesis serves as the final barrier against the misuse of digital designs. However, the risk is not confined to functional expression within a living organism or a cell-free system. The production of genetic intermediates, such as a plasmid containing a potent toxin gene, poses a significant dual-use risk even before activation in a host. As AI capabilities outpace current laboratory regulations, malicious actors may find new ways to circumvent established screening protocols for these precursors. These rapid advancements necessitate more agile governance across the entire synthesis and GMO domain to ensure oversight covers both the final biological products and the foundational genetic components that enable them.

6.2 Recommendations

6.2.1 Recommendation 1: Strengthen screening, detection, and registries across molecular synthesis providers

Participants agreed that synthesis screening systems must integrate homology, structural analysis and AI-based annotations, as well as taking the user's context into account. This would require a database of species-specific harmful bio-sequences. This would enable synthesisers to progress beyond sequence-only filters, which are inadequate for AI-driven synthetic biology, to a system that integrates various methodologies to build a more realistic risk profile. This could potentially allow them to identify dangerous sequences that are heavily modified or even designed *de novo*. Although ongoing work in this area has recently been published, it has focused solely on toxic protein sequences and limited viral protein sequences. No DNA was considered, and the scope was limited due to biosecurity constraints [WAB+25]. In addition to strengthening screening processes, it would be beneficial to register ongoing work requiring synthesis. In order to achieve this without overstepping privacy boundaries while still allowing research to operate efficiently, two considerations must be taken into account.

Firstly, user-based authentication would allow scores to be modulated based on individuals or organisations. For example, a sequence synthesised for a team of oncolytic virus researchers should be treated differently to one requested by verified members of the public. Even if the sequence shares structural motifs with toxins, rather than assuming weaponisation, the therapeutic intent of the user should be acknowledged, synthesis should be progressed and the risk noted for all involved. The ability of those involved to handle dual-use risks safely should be verified during the registration process.

Secondly, biological context is paramount. For instance, a viral population in a particular animal reservoir (e.g., cattle) could be harmless or even beneficial. However, if that same virus were engineered to cross species barriers into humans, it could become pathogenic. Therefore, filters must generate species-specific risk scores that evaluate the potential harm to humans, agriculture and the wider ecology independently [ZUPH+20, LKRS+22].

Combining strengthened technical detection mechanisms with the registration of harmful sequences creates safety guardrails that enable scientific progress while mitigating the probability of accidental release or malicious misuse.

6.2.2 Recommendation 2: Implement KYC rules and RFID tracking for critical components in the supply chain.

For filters and screening to have a meaningful impact, it is crucial to understand who is synthesising what and where. In order to reduce the risk of malicious actors using synthesis services to hide the dangers of their sequences. For example, by shuffling the sequence or with the help of GenAI, regulators would need to implement Know Your Customer (KYC) rules. This would involve extending identity verification and legitimate-use checks, as well as introducing documentation obligations, all of which would be proportionate to the level of risk. This could be implemented alongside a usage registry, as suggested above.

Some elements of this already exist in the form of serial numbers, end-use logging and sales controls for certain high-end synthesis equipment. Participants viewed these as useful foundations that require significant strengthening and expansion in light of GenAI.

They also emphasised the importance of incident response capacity and human oversight to ensure that automated synthesis processes remain within the scope of responsible governance, and that any suspicious or concerning synthesis activities prompt a timely review and response. There are ongoing efforts in this direction, by the Gene Synthesis Consortium.

6.2.3 Recommendation 3: Establish or reinforce clear governance for nucleic acid synthesis screening.

Good governance of synthesis needs technical/procedural measures, as well as robust legal, normative and educational frameworks. Guidelines can be updated more often to allow governance structures to adapt to technological and threat developments. These frameworks must be designed for regular revision. Training and education are needed to raise awareness of the risks associated with synthetic sequences and generative design among researchers, laboratory personnel, institutional leaders and industry stakeholders. Institutional by-laws and manufacturer rules are important layers of governance that can complement national legislation and international agreements. All of these measures are essential for establishing a governance system capable of addressing both centralised industrial synthesis and benchtop or decentralised systems.

6.2.4 Recommendation 4: Create Sector-Level Insurance to Fund Research on Countermeasures and Risks

As a final point, the countermeasures recommended in this work, especially those in the synthesis section, can be costly and research-intensive. Therefore, participants proposed the establishment of sector-level insurance and pooled funding mechanisms to support sustained investment in preparedness, risk mitigation, and countermeasure research related to synthetic biology and biological GenAI. Such a fund could also be used to finance research into mitigation techniques and governance guidelines. There are precedents for sector-level pooled insurance financing prevention research in other high-risk industries. The most notable example is the German statutory accident insurance system (DGUV), in which employer contributions organised by industry sector explicitly fund occupational safety research [FMM+18].

7 Discussion

Biological sequence design is experiencing a significant transformation driven by advances in precision physical technologies and engineering systems. This convergence has led to a significant drop in sequencing costs and the generation of vast quantities of data, which traditional analytical methods cannot process [Wet22, HAİ+26]. The EU Biotech Act recognises the importance of these developments for the Union's competitiveness and strategic autonomy. However, the 'digital deluge' presents a challenge for the scientific community, as they must balance the potential of GenAI with its risks to societal safety and scientific integrity [HAİ+26]. Responsible handling and regulation are essential to ensure progress and security, as historical precedents of dual-use wet-lab research demonstrate.

The effectiveness of these generative systems is limited by barriers to translation within the data domain, where intense competition and significant bottlenecks have been created by the 'extreme hunger' for information. Rigorous analysis of biological datasets reveals they do not provide neutral descriptions of reality. Rather, they reflect historical, cultural and biological inequalities, particularly the under-representation of non-European populations in genomic studies [MR20]. These inequalities lead to ethical failures and hinder scientific progress by producing models that lack universal applicability. The proposed Act seeks to address these deficiencies by creating 'data quality accelerators' and 'high-impact strategic projects', as outlined in Chapter VI. These projects aim to harmonise technical fragmentation. Compartmentalised information within organisations, coupled with academia's 'publish or perish' culture, hinders the dissemination of vital annotated datasets [BBC+26, GFG+24] (also see Section subsection 2.1). These barriers, compounded by the heterogeneity of the software used to store biological information, mean that even when relevant data exists, it often remains inaccessible. The Commission intends to mitigate this challenge through the 'Networks of Health Biotechnology Clusters', as described in Article 15. Active projects are working towards a shared format for data and models for biological AI models, such as bioAIrepo [TBB+26].

Regarding translational barriers originating from model development, the main issue is the 'semantic gap' between the opaque statistical nature of these 'black box' systems and the requirement for intelligible and contestable outputs in high-stakes clinical settings. These models' lack of interpretability hinders their adoption, as statistical predictions without causal explanations are often insufficient for medical or strategic decision-making [HAİ+26]. Additionally, there is a critical 'verification bottleneck', whereby researchers are currently generating biological entities at a rate that far exceeds the capacity for validation. The EU Biotech Act provides a direct structural response to this via Article 31 and the broader mandate for 'trusted AI testing environments'. By providing the physical and digital infrastructure for in silico validation, these provisions are meant to enable the industry to bridge the gap between digital design and clinical (or physical) reality. Through the development of such validation tools the industry can demonstrate its safety and obtain authorisation to operate at higher velocities, thereby transforming verification from an obstacle into a strategic project for Union-wide expansion.

Implementation is the final and perhaps most critical stage in the biological risk chain, as it is where digital designs are brought into the physical world. The rapid advancement of AI-driven design currently outpaces the development of regulatory frameworks in the wet-lab domain, creating a mismatch that could allow security safeguards to be circumvented. For instance, malicious actors could design 'functional homologues' — sequences that perform the same function as a known pathogen, but differ enough to evade detection. To mitigate these risks, Article 52 of the Act establishes an 'Advisory Group on Biosecurity' to monitor 'biological systemic risk' and ensure robust screening protocols for molecular synthesis against such evasion [WMDC+25]. Article 53 reinforces this 'defence in depth' strategy by mandating the tracking of critical supply chains and the application of accountability measures [HAİ+26]. By embedding these protections into the research lifecycle, the scientific community can ensure that the transition from digital code to physical life remains a controlled and beneficial endeavour for society.

The key change in thinking about GenAI for designing biological sequences is moving from a zero-sum view to a model of radical interdependence, which is a shift in mindset that has significant implications for the field. Biosecurity is seen as a brake on discovery, not even as a necessary evil nor as the chance for growth it actually can represent. But a modern biosecurity regulatory framework is key to the bioeconomy. Setting clear rules and safety benchmarks within regulatory sandboxes enables governance to reduce existential risk and encourage investment. Security should be an integral part of technology, providing the social licence to operate and enabling the scientific community to keep pace with digital technology while ensuring safety. This shift towards interdependence is in line with the EU's strategic autonomy and economic security goals. The proposed EU Biotech Act acknowledges risks such as the unlawful transfer of design protocols and foreign dependencies. By presenting security-by-design as a distinctive European value proposition for gene and protein engineering, breakthroughs in sequence synthesis and functional genomics can be brought to market safely. Therefore, a security-first approach empowers the scientific community to innovate at the pace of digital AI while ensuring safety and cutting-edge progress.

International frameworks like the EU's planned biotech legislation are vital, but governments must also take action. Currently, the security of AI-driven biological research often falls between departments. In Germany, the DFG and the Leopoldina are suitable to lead this debate, using their Joint Committee on Security-related Research. By extending this focus to AI-driven design and perhaps forming groups between health, defence and research, countries can develop a joined-up strategy. Local approaches can complement the EU's oversight. Ultimately, responsible governance does not stop progress. Demonstrating management of these technologies can build public trust and stability.

8 Conclusion

Artificial intelligence for biological sequence design holds transformative promise for health and agriculture. Yet, its full deployment in high-stakes domains remains constrained by translational barriers in data and model verification. A central tension identified by this Sandpit is that removing these barriers will inevitably increase model capabilities, which in turn amplifies dual-use potential. Governance must therefore evolve in step with capability. Importantly, this relationship is not purely adversarial: many of the safeguards proposed in this work, particularly advances in model interpretability, uncertainty quantification, and structured risk assessment, are likely to improve the scientific utility of these

models as well, placing biosecurity and performance as converging objectives. The most productive path forward is to embed biosecurity and ethical considerations into the development of biological AI from the outset, treating responsible governance not as a constraint on progress but as an integral driver of it.

9 Ethical and Societal Impact

This Perspective is motivated by the recognition that generative AI for biological sequence design is advancing faster than the governance frameworks needed to ensure its responsible use. While our recommendations are intended to reduce biosecurity risks and promote equitable access to biological AI tools, we acknowledge that governance itself carries risks. Overly restrictive data or model access controls could create a two-tiered research community, disadvantaging researchers at under-resourced institutions and in the Global South. Risk classification systems, if poorly designed, could inadvertently signal which datasets or models are most valuable for misuse. Tiered access platforms, if centralised, could become targets for state-level adversaries. We have attempted to address these tensions throughout the framework by proposing proportionate, risk-stratified controls rather than blanket restrictions, by emphasising equitable access as a design principle, and by recommending that classification efforts involve broad, interdisciplinary expert consensus rather than unilateral decisions. We also note that interpretability methods, while valuable for safety, could themselves be misused to steer models toward harmful outputs. The defence-in-depth approach we advocate is designed to ensure that no single point of failure determines the safety of the system. Ultimately, the societal benefit of this work lies in enabling the life sciences community to realise the transformative potential of biological AI while maintaining the public trust necessary for sustained investment and adoption.

Author Contributions

M. Sprang and H.A. Hassan wrote the main draft. K. Ladewig contributed to the conceptualisation of the manuscript and was responsible for the critical review. G. Hattab contributed to the conceptualisation, writing, critically reviewing and proofreading the manuscript. M. Sprang, M. Schäfer, H.A. Hassan, K. Ladewig, and G. Hattab. discussed, designed, and implemented the figures and tables. All authors contributed to the development of the ideas and contributed to the discussions at the Sandpit. M. Sprang and K. Ladewig designed the Sandpit programme. H.A. Hassan designed the outreach material and did literature research and data mining in preparation of the event.

Acknowledgments

This work is the result of a Sandpit Conference funded by the Wübben-Stiftung für Wissenschaft. G.H., K.L., H.A.H. and M. Sprang acknowledge the support of the Wübben Foundation through the Sandpit Funding Scheme. M. Sprang was supported by funding from the Einstein ECR Award, the Rise Up! programme of the Boehringer Ingelheim Foundation (BIS), the ReALity Initiative of Johannes Gutenbergalek University Mainz, and the Forschungsinitiative des Landes Rheinland-Pfalz. M. Schäfer was supported by the German Federal Ministry of Health (BMG) under grant No. 2523DAT400, project "AI-assisted analysis and visualization of pandemic situations" (AI-DAVis-PANDEMICS). M. Sprang and H.A.H. acknowledge funding from the Carl Zeiss Foundation through the Seed Funding Programme of the Centre for Synthetic Genomics.

References

[AL23] Sage Albright and Stilianos Louca. Trait biases in microbial reference genomes. *Scientific Data*, 10(1):84, 2023.

[BBC+26] Doni Bloomfield, James RM Black, Oliver Crook, Nadav Brandes, Moritz S Hanke, Thomas V Inglesby, Anita Cicero, Robert Pollack, Tina Hernandez-Boussard, Michael J Imperiale, et al. Biological data governance in an age of ai. *Science*, 391(6785):558–561, 2026.

[BDCuR+13] Alfredo Benso, Stefano Di Carlo, Hafeez ur Rehman, Gianfranco Politano, Alessandro Savino, and Prashanth Suravajhala. A combined approach for genome wide protein function annotation/prediction. *Proteome science*, 11(Suppl 1):S1, 2013.

[BHM+25a] James RM Black, Moritz S Hanke, Aaron Maiwald, Tina Hernandez-Boussard, Oliver M Crook, and Jassi Pannu. Open-weight genome language model safeguards: Assessing robustness via adversarial fine-tuning. In *NeurIPS 2025 Workshop on Biosecurity Safeguards for Generative AI*, 2025.

[BHM+25b] Doni Bloomfield, Moritz S. Hanke, Aaron Maiwald, James R. M. Black, Toby Webster, Tina Hernandez-Boussard, Allison Berke, Oliver M Crook, and Jassi Pannu. Securing dual-use pathogen data of concern. In *NeurIPS 2025 Workshop on Biosecurity Safeguards for Generative AI*, 2025.

[BTB+23] Trenton Bricken, Adly Templeton, Joshua Batson, Brian Chen, Adam Jermyn, Tom Conerly, Nick Turner, Cem Anil, Carson Denison, Amanda Askell, Robert Lasenby, Yifan Wu, Shauna Kravec, Nicholas Schiefer, Tim Maxwell, Nicholas Joseph, Zac Hatfield-Dodds, Alex Tamkin, Karina Nguyen, Brayden McLean, Josiah E Burke, Tristan Hume, Shan Carter, Tom Henighan, and Christopher Olah. Towards monosemanticity: Decomposing language models with dictionary learning. *Transformer Circuits Thread*, 2023. https://transformer-circuits.pub/2023/monosemantic-features/index.html.

[BTWH18] Stephen Baker, Nicholas Thomson, François-Xavier Weill, and Kathryn E Holt. Genomic insights into the emergence and spread of antimicrobial-resistant bacterial pathogens. *Science*, 360(6390):733–738, 2018.

[CB26] Sarah R. Carter and Greg Butchello. A framework for managed access to biological AI tools. Technical report, Nuclear Threat Initiative, January 2026.

[Cho24] Samuel D Chorlton. Ten common issues with reference sequence databases and how to mitigate them. *Frontiers in Bioinformatics*, 4:1278228, 2024.

[DAİH26] Akshat Dubey, Aleksandar Anžel, Bahar İlgen, and Georges Hattab. Ubiqtree: Uncertainty quantification in xai with tree ensembles. *Patterns*, 2026.

[dARDT+25] Bernardo P de Almeida, Guillaume Richard, Hugo Dalla-Torre, Christopher Blum, Lorenz Hexemer, Priyanka Pandey, Stefan Laurent, Chandana Rajesh, Marie Lopez, Alexandre Laterre, et al. A multimodal conversational agent for dna, rna and protein tasks. *Nature Machine Intelligence*, 7(6):928–941, 2025.

[DS24] Frances Ding and Jacob Steinhardt. Protein language models are biased by unequal sequence sampling across the tree of life. In *ICLR 2024 Workshop on Generative and Experimental Perspectives for Biomolecular Design*, 2024.

[FMM+18] PÜTZ Fabian, Finbarr Murphy, Martin Mullins, Karl Maier, Raymond Friel, and Torsten Rohlfs. Reasonable, adequate and efficient allocation of liability costs for automated vehicles: a case study of the german liability and insurance framework. *European Journal of Risk Regulation*, 9(3):548–563, 2018.

[GFG+24] Anthony Gitter, James S Fraser, Tamir Gonen, Robert Patro, Hannah K Wayment-Steele, Alicia Williams, Benjamin Haibe-Kains, Roland L Dunbrack10, Cameron Cook11, Anshul Kundaje12, et al. A renewed call for open artificial intelligence in biomedicine. *Preprint at https://doi. org/10.31219/osf. io/2xh3w*, 2024.

[GYW22] Wilson Wen Bin Goh, Chern Han Yong, and Limsoon Wong. Are batch effects still relevant in the age of big data? *Trends in biotechnology*, 40(9):1029–1040, 2022.

[HAİ+26] Hassan Ahmed Hassan, Aleksandar Anžel, Bahar İlgen, Marina Luchner, Patrick Schramowski, Anja Blasse, Johannes U Mayer, Katharina Ladewig, Georges Hattab, and Maximilian Sprang. A safety-centric perspective on innovation and risk in the use of artificial intelligence in genomics. *Trends in Genetics*, 2026.

[Hat26] Georges Hattab. Human agency, causality, and the human-computer interface in high-stakes artificial intelligence. In TBA, editor, *Proceedings of the 2026 CHI Workshop on Human-AI Interaction Alignment: Designing, Evaluating, and Evolving Value-Centered AI for Reciprocal Human-AI Futures (CHI EA '26)*, ACM Workshop Proceedings, page 6, New York, NY, USA, 2026. ACM.

[IDH25] Bahar Ilgen, Akshat Dubey, and Georges Hattab. PHAX: A structured argumentation framework for user-centered explainable AI in public health and biomedical sciences. In Timotheus Kampik, Antonio Rago, Kristijonas Cyras, and Oana Cocarascu, editors, *Proceedings of the 3rd International Workshop on Argumentation for eXplainable AI (ArgXAI 2025) co-located with the 28th European Conference on Artificial Intelligence (ECAI 2025), Bologna, Italy, October 26, 2025*, volume 4066 of *CEUR Workshop Proceedings*, pages 78–89. CEUR-WS.org, 2025.

[KDL+25] 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*, pages 2025–09, 2025.

[LKRS+22] Mats Leifels, Omar Khalilur Rahman, I-Ching Sam, Dan Cheng, Feng Jun Desmond Chua, Dhiraj Nainani, Se Yeon Kim, Wei Jie Ng, Wee Chiew Kwok, Kwanrawee Sirikanchana, Stefan Wuertz, Janelle Thompson, and Yoke Fun Chan. The one health perspective to improve environmental surveillance of zoonotic viruses: lessons from covid-19 and outlook beyond. *ISME Communications*, 2(1):107, 10 2022.

[MJD+25] Aaron Maiwald, Piotr Jedryszek, Florent Draye, Bernhard Schölkopf, Garrett M Morris, and Oliver M Crook. Decode-glm: Tools to interpret, audit, and steer genomic language models. *bioRxiv*, pages 2025–10, 2025.

[MR20] Melinda C Mills and Charles Rahal. The gwas diversity monitor tracks diversity by disease in real time. *Nature genetics*, 52(3):242–243, 2020.

[MRPCC+25] Gemma I Martínez-Redondo, Francisco M Perez-Canales, Belén Carbonetto, José M Fernández, Israel Barrios-Núñez, Marçal Vázquez-Valls, Ildefonso Cases, Ana M Rojas, and Rosa Fernández. Fantasia leverages language models to decode the functional dark proteome across the animal tree of life. *Communications Biology*, 8(1):1227, 2025.

[MTH+23] Frederikke Isa Marin, Felix Teufel, Marc Horlacher, Dennis Madsen, Dennis Pultz, Ole Winther, and Wouter Boomsma. Bend: Benchmarking dna language models on biologically meaningful tasks. *arXiv preprint arXiv:2311.12570*, 2023.

[NMRK26] Masayuki Nagai, Alan E Murphy, Kaeli Rizzo, and Peter K Koo. Toward interpretable and generalizable ai in regulatory genomics. *arXiv preprint arXiv:2602.01230*, 2026.

[oGSoA20] NIHR Global Health Research Unit on Genomic Surveillance of AMR. Whole-genome sequencing as part of national and international surveillance programmes for antimicrobial resistance: a roadmap. *BMJ global health*, 5(11):e002244, 2020.

[oSEM+25] National Academies of Sciences Engineering, Medicine, et al. Disseminating in silico and computational biological research: Navigating benefits and risks.proceedings of a workshop. 2025.

[PSW+22] Jaspreet Pannu, Jonas B Sandbrink, Matthew Watson, Megan J Palmer, and David A Relman. Protocols and risks: when less is more. *Nature protocols*, 17(1):1–2, 2022.

[PSW+24] Aman Patel, Arpita Singhal, Austin Wang, Anusri Pampari, Maya Kasowski, and Anshul Kundaje. Dart-eval: A comprehensive dna language model evaluation benchmark on regulatory dna. *Advances in Neural Information Processing Systems*, 37:62024–62061, 2024.

[QHY+25] Yuanhao Qu, Kaixuan Huang, Ming Yin, Kanghong Zhan, Dyllan Liu, Di Yin, Henry C Cousins, William A Johnson, Xiaotong Wang, Mihir Shah, et al. Crispr-gpt for agentic automation of gene-editing experiments. *Nature Biomedical Engineering*, pages 1–14, 2025.

[RHC+] Sebastian Rivera, Moritz S Hanke, Samuel Curtis, Neil Cherian, Anthony Gitter, Jeffrey J Gray, Andrew Hebbeler, Stephen McCarthy, David Nannemann, Claire Qureshi, et al. Responsible biodesign workshop: Ai, protein design, and the biosecurity landscape–recommended actions.

[SANF22] Maximilian Sprang, Miguel A Andrade-Navarro, and Jean-Fred Fontaine. Batch effect detection and correction in rna-seq data using machine-learning-based automated assessment of quality. *BMC bioinformatics*, 23(Suppl 6):279, 2022.

[SMANF24] Maximilian Sprang, Jannik Möllmann, Miguel A Andrade-Navarro, and Jean-Fred Fontaine. Overlooked poor-quality patient samples in sequencing data impair reproducibility of published clinically relevant datasets. *Genome Biology*, 25(1):222, 2024.

[TBB+26] Matthew Thakur, Nicolas Bosc, Cath Brooksbank, Christina Ernst, Mallory A Freeberg, Kim T Gurwitz, Henning Hermjakob, David G Hulcoop, Maria J Martin, Ellen M McDonagh, et al. Embl's european bioinformatics institute (embl-ebi) in 2025. *Nucleic acids research*, 54(D1):D10–D19, 2026.

[TGB+17] Julien Troudet, Philippe Grandcolas, Amandine Blin, Régine Vignes-Lebbe, and Frédéric Legendre. Taxonomic bias in biodiversity data and societal preferences. *Scientific reports*, 7(1):9132, 2017.

[WAB+25] Bruce J Wittmann, Tessa Alexanian, Craig Bartling, Jacob Beal, Adam Clore, James Diggans, Kevin Flyangolts, Bryan T Gemler, Tom Mitchell, Steven T Murphy, et al. Strengthening nucleic acid biosecurity screening against generative protein design tools. *Science*, 390(6768):82–87, 2025.

[Wet22] Kris A. Wetterstrand. Dna sequencing costs: Data from the NHGRI genome sequencing program (GSP), 2022.

[WHZ+26] Dianzhuo Wang, Marian Huot, Zechen Zhang, Kaiyi Jiang, Eugene I Shakhnovich, and Kevin M Esvelt. Without safeguards, ai-biology integration risks accelerating future pandemics. *Frontiers in Microbiology*, 16:1734561, 2026.

[WMDC+25] T Webster, R Moulange, B Del Castello, J Walker, S Zakaria, and C Nelson. Global risk index for ai-enabled biological tools. *The Centre for Long-Term Resilience & RAND Europe*, 2025.

[ZCQ+25] Qiang Zhang, Wanyi Chen, Ming Qin, Yuhao Wang, Zhongji Pu, Keyan Ding, Yuyue Liu, Qunfeng Zhang, Dongfang Li, Xinjia Li, et al. Integrating protein language models and automatic biofoundry for enhanced protein evolution. *Nature Communications*, 16(1):1553, 2025.

[ZUPH+20] Jakob Zinsstag, Jürg Utzinger, Nicole Probst-Hensch, Lv Shan, and Xiao-Nong Zhou. Towards integrated surveillance-response systems for the prevention of future pandemics. *Infectious Diseases of Poverty*, 9(1):140, 10 2020.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- nrreportingsummary.pdf
- rs.pdf