Our approach to biosecurity for AlphaFold 3

Authors: Conor Griffin, Heidi Howard, Antonia Paterson, Nick Swanson, Dawn Bloxwich, John Jumper, Pushmeet Kohli, Nicklas Lundblad

We believe that AlphaFold 3 will help to accelerate progress on tackling priority biosecurity challenges, and that it can be safely released. Below, we explain how we think about biosecurity, how we approached it for AlphaFold 3, and our plans for future work.

Today we announced AlphaFold 3, a revolutionary model which can predict the structures and interactions of all life's molecules with state-of-the-art accuracy, providing valuable insights into how proteins interact with DNA, RNA and other biomolecules. Scientists can access the majority of its capabilities, for free, through our newly launched AlphaFold Server, an easy-to-use research tool.

We began exploring the potential effects of our research on biosecurity in the early days of AlphaFold, in 2020. Back then, discussions about 'AI *and* biosecurity' were rare and often limited to a handful of individuals. Today the topic has become much more prominent given the recent acceleration of progress in AI and biology. Last year, biosecurity was mentioned in both the White House Commitments and the US Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. The UK & US AI Safety Institutes and other government bodies are also considering how to best assess AI biosecurity risks, while civil society actors, such as the Nuclear Threat Initiative (NTI), are establishing dedicated forums to identify and share AI and biosecurity best practices. At Google DeepMind, we actively contribute to and participate in these deliberations.

Based on consultations with a broad community of experts, and early evidence on how AlphaFold 2 is being used, **we believe that AlphaFold 3 will help researchers make progress on priority biosecurity challenges**, **such as pandemic preparedness and neglected tropical diseases.** We have also concluded that AlphaFold 3 can be safely released and does not lead to a material uplift in risk when compared to other readily available state-of-the-art structure prediction tools.

To contribute towards ongoing AI and biosecurity discussions, we believe it is important to outline how we reached this conclusion. In this blog, we first explain how we define biosecurity,

and how we think AI - more broadly - could affect it. We then walk through how we assessed the potential impact of AlphaFold 3. Finally, we conclude with some future work priorities for ourselves and the broader AI and biosecurity community. This includes a new exploration into how biosecurity screening processes could work. We will trial this for the launch of AlphaFold Server, in a narrow, targeted manner to avoid impinging on beneficial use cases. Our goal with this exploration is to inform potential screening systems for future, more powerful, AI models and tools.

What do we mean by 'biosecurity'?

We define biosecurity as protection against high-impact biological risks, including (a) those that arise naturally, (b) via unintentional accidents[1]; and (c) via the intentional development and/or deployment of biological weapons.[2] Throughout history, most biosecurity harms have come from pandemics, including the estimated 27 million excess deaths from COVID-19. Mortality was particularly high before the discovery of germ theory, when the world lacked good knowledge of pathogens and how to protect ourselves. Fatal accidents from high-risk biology research are rarer, but do sometimes occur. The intentional development and deployment of bioweapons dates back centuries, but has also been extremely rare since the Biological and Toxin Weapons Convention (BWC) came into force in 1975, and larger state-sponsored programmes ended.

As outlined in the UK's recently-updated Biological Security Strategy, today's priority biosecurity risks include the "reasonable likelihood" that another serious pandemic could occur soon, increased antimicrobial resistance (AMR), and the prospect of a deliberate biological attack by state or non-state actors.[3] However, biosecurity priorities look different across the world. For example, neglected tropical diseases, such as dengue and Chagas disease, are a diverse group of often debilitating diseases caused by a variety of pathogens. Outside of a small number of organisations, these diseases receive relatively little global funding, despite being an immediate, everyday concern for many people, particularly those living in poverty and in tropical and subtropical conditions.

[1] Organisations define biosecurity in different ways. Historically, the term was often used to refer to protection against the intentional misuse of biology, for example via bioterrorism. More recently, the concept has broadened to also include protection against biological accidents - traditionally referred to as 'biosafety' - and societal protection and resilience against natural outbreaks.
[2] The three biosecurity categories of natural, accidental, and intentional misuse can sometimes be hard to distinguish. For example, historically some naturally-occurring outbreaks were exacerbated by intentional misuse.
[3] A range of underlying risk factors are increasing the likelihood of naturally-occuring epidemics and pandemics, including the effects of climate change; encroachment on animal habitats, for example via deforestation; urbanisation and migration flows; inadequate public health infrastructure; and growing vaccine hesitancy among certain population groups.

How might AI benefit biosecurity?

Recently, a number of organisations have published research outlining how AI could affect biosecurity, some of which we contributed to.[4] Most publications focus on risks from AI to biosecurity, which is a very important conversation. There has been less focus on potential benefits from AI to biosecurity, an imbalance that we would like to help address.

We believe that AI could help us to better *understand*, *prevent, detect,* and *respond* to future outbreaks. For example, genome sequencing played a crucial role in the response to COVID-19, with the COVID-19 Genomics UK Consortium using it to rapidly assess viral transmission patterns, new strains and to determine the potential implications for vaccines. Looking ahead, practitioners hope to use metagenomic sequencing to detect and respond to new outbreaks, by rapidly collecting more diverse types of genetic material from waterways and wastewater. AI could be critical to analysing this huge volume of data. Practitioners could also use AI to help design new antibodies and vaccines, both for known and unknown future pathogens. For instance, we already see early examples of how AI could support efforts to develop a new generation of mRNA vaccines, building on the striking success of these vaccines during COVID-19.

We also see promising early evidence for how scientists are using AlphaFold to better understand the biology underpinning pathogens, and in doing so contributing to efforts to develop new diagnostics, therapeutics and vaccines. For example, cholera is an acute diarrheal illness that leads to an estimated 20-140,000 deaths every year, primarily in low-and-middle-income countries. Combating it is difficult because the *Vibrio cholerae* bacteria can enter into a protective "survival" state. Researchers recently used AlphaFold to validate experimental structures and predict interactions for two proteins that are critical to this transition, opening up new ways to target the bacteria.

In 2022, an estimated 1.3 million people died as a result of tuberculosis (TB), partly because the *Mycobacterium tuberculosis* has a complex life cycle, and biology, that makes creating new drugs difficult. For nearly 10 years, a team at Bhabha + Ekiert Labs has studied a family of proteins that help these bacteria to scavenge nutrients and stay alive. In a recent Nature paper, the team documented how they used AlphaFold, alongside experimental imaging techniques and biochemical analysis, to understand how these proteins fit together, a critical step in better understanding TB and potentially developing new treatments.

AlphaFold has also directly supported new vaccine efforts. Malaria accounts for approximately 610,000 deaths every year. 95% of these deaths occur in Africa, where 80% of those who die are children under five years of age. Malaria is caused by single-celled parasites, of which

[4] Recent publications on AI and biosecurity include those from the Nuclear Threat Initiative, the Centre for Long-Term Resilience, the Engineering Biology Research Consortium, and the US National Security Commission on Emerging Biotechnology.

Plasmodium falciparum is the most dangerous. This parasite contains a gamete surface protein (Pfs 48/45), which is one of the most promising candidates for inclusion in a potential transmission-blocking malaria vaccine. In a 2022 paper, Professor Matthew Higgins and his team at the University of Oxford determined the first full-length structure of the protein, using crystallography, guided by AlphaFold. Using this structure, they were able to show where transmission-blocking antibodies bind to the protein, supporting new vaccine efforts.

With the launch of AlphaFold 3, we expect to see new opportunities to benefit biosecurity. For example, the ability to predict how DNA interacts with other molecules, such as helicase and topoisomerase enzymes, could support efforts to develop new antiviral drugs and antibiotics to target these essential enzymes, helping to tackle antimicrobial resistance. Similarly, AlphaFold 3 can predict RNA structures, but also complexes of RNA, proteins, ligands and ions, and biologically meaningful interactions. This will open up new therapeutic opportunities. For example, AlphaFold 3 can computationally reproduce the recent experimentally-derived structure 8AW3, a complex formed by an enzyme protein and an RNA molecule, and its interactions with ions, with high accuracy. This serves as a proof-of-principle that AlphaFold 3 could deliver sophisticated structures more quickly and cheaply than experimental techniques, with comparable quality. The molecules that form the 8AW3 complex come from the parasite that causes East African Trypanosomiasis - a neglected tropical disease spread by the Tsetse Fly. Such structures have been a longstanding challenge to model, impeding drug development efforts.

How might AI pose risks to biosecurity?

Most analyses on AI and biosecurity focus on two broad categories of risk.[5] First, there is a concern that threat actors may use large language models (LLMs) to obtain information about how to cause harm. In the same way as practitioners are excited by the potential of LLMs to help with scientific research tasks, some are concerned that threat actors may use LLMs to obtain information that is relevant to carrying out an attack, from basic information about sourcing toxins, through to more targeted troubleshooting for experiments. In response, other practitioners typically note that a large amount of information, for example on toxins and pathogens, is already freely available online, and it has not led to an increase in bioweapon attacks. In a recent study involving several LLMs, RAND researchers also concluded that weapon attack planning lies beyond these LLMs' current capabilities.[6] While AlphaFold is not an LLM, Google DeepMind is undertaking research on safety evaluations for LLMs and we look forward to sharing more on this work in due course.

[5] The potential benefits and risks from AI to biosecurity that we discuss in this blog are not exhaustive. For example, other potential benefits range from AI-enabled biosensors to help detect new outbreaks to the use of AI to predict supply chain issues. Other potential risks include an increase in mis- or disinformation about vaccines. For a more comprehensive review of potential benefits and risks, see the 2023 paper The Convergence of Artificial Intelligence and the Life Sciences, by the Nuclear Threat Initiative (NTI).
[6] A 2024 report from the US National Security Commission on Emerging Biotechnology also found that: "*At this time [LLMs do] not significantly increase the risk of the creation of a bioweapon*."

A second category of risk focuses on narrow AI biology models.[7] Unlike large language models, these models usually have a small number of targeted functions. They are typically trained exclusively on biological data, and individuals must have scientific expertise to use them effectively. Within this category, biosecurity practitioners have focussed, in particular, on AI biodesign models, such as for protein design. They have expressed concerns that these models, or future iterations of these models, may lower the barrier for threat actors and enable them, in concert with other technologies, to design and engineer pathogens and toxins that are more transmissible or harmful. In response, other practitioners have noted the technical complexity of the tasks, and the existence of non-AI approaches that appear much more tractable, such as making use of more readily-available toxins. While biodesign models are a fast-moving area requiring vigilance, we are not aware of a credible report of a threat actor using, or trying to use, an AI model to cause harm in this way.[8]

Our approach to biosecurity for AlphaFold 3

AlphaFold is first and foremost a tool to equip scientific researchers with the information they need to design and conduct experiments and research. AlphaFold is not a large language model, nor is it a biodesign tool that can be used to design a new protein or other biomolecule. AlphaFold does have a shared intellectual lineage with biodesign models, such as for protein design, as the breakthrough has supported progress in that field. But the tool itself is substantively different, and requires a distinct biosecurity approach.

A typical AlphaFold workflow includes analysing the structure for the protein of interest to identify functionally important regions like the active site of an enzyme or the binding site of a receptor, to guide further biochemical experiments. Biologists also use the structures to analyse the variability of a particular protein in different organisms, in order to better understand its function, or to identify similar proteins and analyse their function experimentally.[9] Prior to AlphaFold, researchers relied principally on the Protein Data Bank (PDB) which contains around 200,000 experimentally-determined structures of biomolecules. The release of 200 million protein structures via the AlphaFold Database, in partnership with EMBL's European Bioinformatics Institute (EMBL-EBI), has already enabled many experiments which were not previously possible.

Our work on biosecurity began with the first version of AlphaFold and has continued to evolve as we developed AlphaFold 3. We have undertaken a series of measures, informed by our

[7] The distinction between LLMs and AI biology models is not a perfect one. For example, some practitioners also train LLMs on biological data, and there are various other ways the two approaches may intersect.
[8] A recent report by Moulange et. al (2023) also concluded that *"there have been no publicly recorded attempts of biological design tools being used to produce biological weapons or otherwise cause harm."*
[9] Another common use of AlphaFold is to help interpret structural data from experimental techniques, such cryogenic electron microscopy or crystallography. In many cases, this experimental data might be incomplete, of low resolution, or suffer from other imperfections. AlphaFold predictions can complement such data and accelerate their interpretation.

broader approach to responsibility and safety, and by emerging governance practices. Concretely, we did the following:

- **Research:** We carried out targeted research, both internally, and with external partners, to understand the potential future trajectories of AI biology tools, how they might be used, and how this might affect biosecurity, positively and negatively.
- **Ethics and safety assessment:** Our Responsible Development and Innovation team worked with our AlphaFold team and external experts (see below) to carry out an Ethics and Safety Assessment to identify and analyse potential risks and benefits from AlphaFold 3, including their potential likelihood and impact. This assessment was subsequently reviewed by Google DeepMind's Responsibility & Safety Council, who provided further feedback on the release and have worked closely with our research teams since before AlphaFold was initially released in 2021
- **External expert consultation:** Over the course of the AlphaFold development life cycle, we have consulted more than 50 experts in domains ranging from DNA synthesis, to virology, to national security, to understand their view on potential benefits and risks.
- **Technical grounding:** We grounded all of our assessments in the specific technical capacities of what we are developing and deploying. As part of this, we have compared AlphaFold 3 to other resources, such as the Protein Data Bank, which contains a number of openly-available protein structures for viruses and other pathogens, and other available AI biology tools.
- **Security:** As with all Google DeepMind models, AlphaFold 3 is built on an industry-leading approach to general and infrastructure security. We developed, trained, store and serve AlphaFold 3 within Google's infrastructure, supported by central security teams and engineers and researchers with world-class expertise.
- **Accelerating beneficial applications:** We use quantitative and qualitative techniques to monitor AlphaFold adoption and impact. These efforts have highlighted many beneficial applications, such as those highlighted above, as well as obstacles to wider use. In response, we recently partnered with EMBL-EBI to launch a free, comprehensive collection of tutorials and explainers for scientists to help them use AlphaFold. To date, more than 10,000 users have benefited from the course, almost a quarter of whom are located in low- and middle income countries that are most affected by neglected tropical diseases and emerging infectious diseases. Together with EMBL-EBI, we will be expanding the AlphaFold course and partnering with local capacity builders to accelerate the equitable adoption of AlphaFold 3.
- **External community and policy engagement:** Our approach to AI and biosecurity does not start and end with the models and tools that we develop. We recognise that we also need to support community best practices and effective public policy. To that end, we have contributed to civil society and industry efforts in this space, such as NTI's newly-proposed AI Bio Forum, and the Frontier Model Forum. We are also engaging

with government bodies on AI and biosecurity, such as the UK's AI Safety Institute and the UK Biosecurity Leadership Council - which we are a member of. We have learnt a lot from these initiatives, and we look forward to continuing these important conversations.

Future work

Based on the measures described above, we believe that AlphaFold 3 will help researchers to make progress on priority biosecurity challenges, such as pandemic preparedness and neglected tropical diseases. We also believe that it can be safely released and does not lead to a material uplift in risk when compared to other readily available state-of-the-art structure prediction tools.

In addition, we do see AlphaFold 3 as an opportunity to start exploring possible mitigations for future, more powerful AI biology models that may pose a higher risk. One mitigation proposed in the community is to screen and filter the requests that users can make. The question of how to best screen requests for biological information is an open research problem that multiple actors are working on. For example, the International Gene Synthesis Consortium (IGSC), a consortium of gene synthesis companies, is committed to screening customer orders for DNA sequences, a distinct but related challenge to screening uses of AI biology tools. The IGSC, and other initiatives, such as the international Common Mechanism for DNA synthesis screening, are exploring how to adapt approaches to DNA order screening, to account for potential challenges posed by fast-evolving technologies, such as benchtop DNA synthesis devices, CRISPR and AI biodesign tools - which could potentially enable individuals to design sequences that evade current DNA screening protocols.

We want to support these efforts, and recognise that there are several open questions to resolve. In that spirit, we will use the launch of AlphaFold Server as an opportunity to explore the efficacy of screening and filtering processes for AI biology tools. Using the Server as a testbed, we will start our exploration by blocking a small number of viral protein sequences and take steps to minimise false positives, inviting feedback and engaging in dialogue with the AI safety and biosecurity community. We hope to share lessons, and collaboration ideas, with other actors working on this challenge.

We view this as one component of a broader future work programme on AI and biosecurity, to ensure that we advance the safety frontier in line with the capability frontier. This could include other important challenges, such as better aligning on biological sequences of concern; modelling potential threat actors; moving from evaluating risks posed by individual biotechnologies to evaluating risks posed by suites of biotechnologies; and reaching community alignment on how to best use AI to support pandemic preparedness and priority biosecurity challenges.