BMJ Global Health

The nexus of artificial intelligence and virology: transformative but with potential perils

Suresh V Kuchipudi [1], Douglas E Norris [2], Andrew Pekosz [2], James M. Pipas [3]

To cite: Kuchipudi SV, Norris DE, Pekosz A, *et al*. The nexus of artificial intelligence and virology: transformative but with potential perils. *BMJ Glob Health* 2025;**10**:e021140. doi:10.1136/bmjgh-2025-021140

Handling editor Seema Biswas

Received 7 July 2025
Accepted 5 November 2025

[1]Department of Infectious Diseases and Microbiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, Pennsylvania, USA
[2]Department of Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
[3]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Correspondence to
Dr James M. Pipas;
pipas@pitt.edu

EMERGING SCIENCE

The tools for advancing our understanding of basic biology have become increasingly capable of taking on large, multifactorial problems and deriving solutions with amazing speed (figure 1; Scientific Foundations).

Structures and interactomes. Upon emerging from the ribosome, newly synthesised polypeptide chains fold into three-dimensional protein structures. Each protein has a collection of specific shapes (conformations) that bring key amino acids together to create local chemistries that facilitate catalysis of specific reactions if the protein is an enzyme or interaction sites that govern binding to other proteins, nucleic acids or small molecule ligands if the protein function is mediated through intermolecular interactions. Thus, solving the structures of proteins is key to understanding how they function and to designing compounds (eg, drugs) that bind and modulate their activity, forming the basis of structure activity relationship studies.[1] Advances in high throughput X-ray crystallography, cryo-electron microscopy and nuclear magnetic resonance have resulted in hundreds of thousands of solved protein structures.[2] These structural insights, coupled with developments in mass spectrometry, have enabled the identification of thousands of protein-protein and protein-ligand interactions that regulate biological activity. Artificial intelligence (AI), particularly deep learning, has emerged as a powerful tool for modelling protein structure and interactions. The development of AlphaFold and other platforms that use computational strategies to predict the structure of individual proteins, as well as protein-protein and protein-ligand complexes, has already greatly accelerated research in these areas.[3 4] These approaches are in their infancy, and the application of AI to these strategies will lead to a rapid expansion of our understanding of protein biology and manipulation.

SUMMARY BOX

⇒ Advances in molecular and computational biology are rapidly converging.
⇒ This commentary discusses the nexus of virology and artificial intelligence.
⇒ The application of artificial intelligence to the biology of viral systems has significant benefits but also has the potential for misuse.

High throughput screening. New platforms such as droplet microfluidics, single-cell transcriptomics and digital PCR allow millions of reactions to be tested in parallel, providing the foundation for large-scale discovery and selection strategies. Droplet microfluidics allows the production of millions of pico-litre-sized droplets that function as mini-test tubes. The droplets can be manipulated to allow the insertion of reaction reagents and for the screening and isolation of specific products.

Guided evolution. Building on high throughput platforms, guided evolution enables the selection of enzymes with enhanced activity or viral variants that escape antibody inactivation.

Since millions of droplets can be created and tested in a few hours, the microfluidic platforms have been used to *guide* evolution in the test tube.[5] For example, microfluidics has been used to select and isolate enzymes with enhanced activity and to identify viral variants that escape inactivation by antibodies.[6 7] The application of AI to these approaches will accelerate the ability to identify and fine-tune the activity of drugs or toxins that improve, inhibit or alter protein function. Since many viruses have high mutation rates, it is also important to start to assess the mechanisms by which viruses can become resistant to therapies, as rapid emergence of resistance could

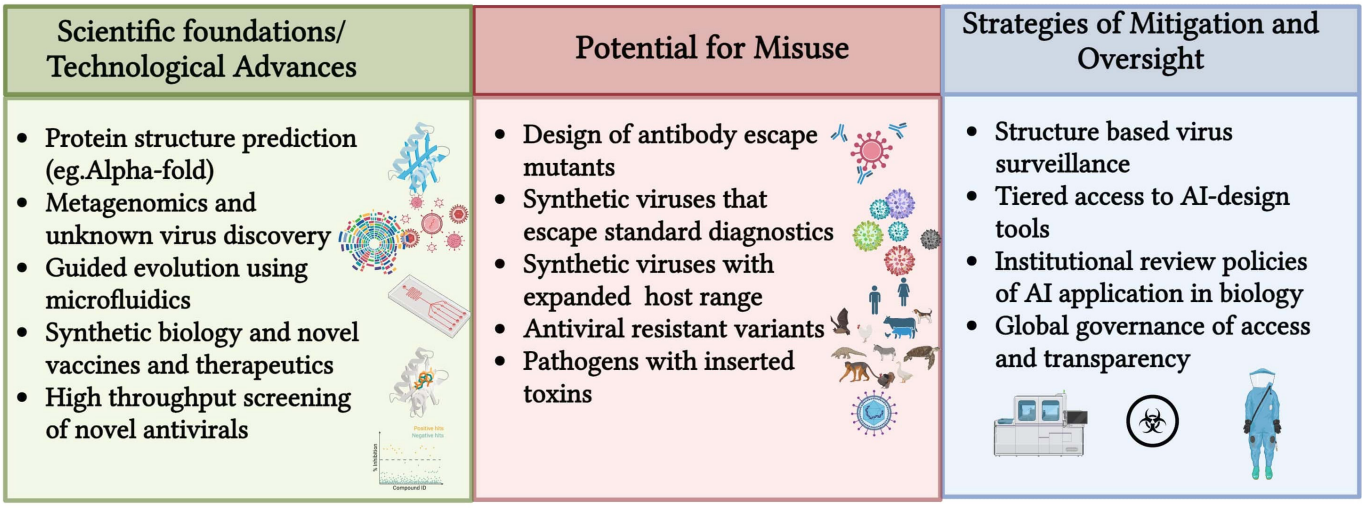

Figure 1 The emerging science of artificial intelligence-driven biology: promise, risk and responsibility.

dramatically shorten the effective lifetime of an antiviral therapy. In addition, AI-assisted directed evolution will enable high throughput selection strategies to alter virus yield, cell-type tropism, host range or virulence.

The unknown biosphere. The vast majority of organisms on planet Earth are unknown to science. Most mammals have been identified and perhaps most birds, fishes and reptiles; beyond that, coverage is fragmented and far from complete. Knowledge is especially sparse when it comes to microorganisms such as single cell eukaryotes, bacteria, fungi and viruses. It is likely that new medicines, new chemistries and entire new industries await discovery in this unexplored biological universe.[8] Metagenomics is a powerful tool for identifying known and unknown organisms present in tissues or environmental samples. Samples are subjected to unbiased deep sequencing, and resulting sequence reads are computationally compared with databases that contain sample sequences of all known species. For example, there are approximately 100 000 recognised viral species, of which about 219 are known to infect humans. These can be readily identified by metagenomics. Yet, metagenomic studies have shown that greater than 99.9% of viruses present in raw sewage are novel and it is not completely clear what their natural hosts are or what role they play in biological or ecological cycles.[9] Most of these are not a direct threat to humans or wildlife and therefore, the chances of a novel human pathogen arising from this viral mix are very low. However, it has been estimated that at any given moment, there are 10^{31} virus particles of all types on planet Earth, with about 10^{24} infections occurring every second, mostly involving cyanophages in the ocean.[10 11] These numbers suggest that novel pathogens are continuously emerging. AI has transformed the speed and accuracy at which metagenomic data can be acquired and analysed. The application of AI to metagenomics will lead to the identification of many of the currently unknown organisms, the assembly of their genomes, the structures and functions of their proteomes and the inference of their roles in nature.

Synthetic biology. For most of history, knowledge of biology was based on the study of proteins that nature has provided. The current pan-proteome of Earth is the result of billions of years of evolution, a process driven by chance and circumstances. Future studies will couple knowledge gained from the natural world with synthetic biology. Building on advances in structures, interactomes, screening and guided evolution, synthetic biology enables the design of proteins and pathways not found in nature. AI-driven approaches are beginning to enable the design of entirely new proteins and pathways with functions not found in nature.[12 13] This type of thinking is already being applied to biomedicine such as gene therapy and the development of oncolytic viruses.[14–16] Natural viral evolution can drive antibody escape or host-range expansion, while synthetic biology raises distinct possibilities for engineering viruses tailored to deliver genetic instructions or for specific therapeutic purposes. Such approaches could enable the design of novel viruses that are tailored to deliver genetic instructions to specific cells, or organisms for therapeutic purposes.

The convergence of protein structure, guided evolution, metagenomics and synthetic biology, accelerated by AI, will transform biological discovery, therapeutics and our understanding of the natural world. For example, AlphaFold2 has predicted over 200 million protein structures in just a few years, compared with ~200 000 structures solved experimentally over the previous six decades.[17] The recent use of AI-driven genome language models and synthetic biology has led to the discovery of a panel of phages with better bacterial killing properties, further cementing the potential for these technologies to bring dramatic advancements to virology.

THREATS

Advances in AI-assisted biology can be redirected by state-sponsored, terrorist organisations, small groups or even individual actors in ways that threaten society. This section describes some of the potential threats that are accelerated or enabled by the convergence of AI and biology (figure 1; Potential for Misuse). This threat analysis is limited to viruses, but most of the same principles apply to bacterial, fungal and protozoan pathogens.

Modification of known human pathogens to increase virulence and/or escape detection. The most straightforward path to creating a novel viral bioweapon is to modify a virus that is already a human pathogen.[18] The AI-biology convergence can be misused to generate variants of known viruses that escape neutralisation by antibodies or are resistant to existing antiviral therapies. For instance, computational approaches might help predict changes in viral surface proteins that reduce antibody binding while preserving receptor interaction. Escape from neutralising antibodies often comes at a fitness cost to the virus—perhaps the virus that escapes neutralising antibodies has a more unstable attachment protein or binds to its receptor less efficiently. AI approaches can work through those problems and provide roadmaps to how to make those escape mutations more tolerated by the virus.

The same approach could be used to generate viral variants that escape detection by diagnostic methods. Pathogens can be identified by their genomes. Two common ways of doing this are PCR and sequencing. Both methods require knowledge of the specific pathogen sequence. Pathogens can be manipulated to escape detection by utilising the redundancy of the genetic code. Thus, it is straightforward to generate viral variants with genome sequences that vary widely from the known pathogen but still encode the exact same proteome. Such agents would escape most detection schemes that depend on nucleic acid sequencing, despite retention of the infectivity of the original pathogen. Currently, these agents would be detected by computational approaches that perform six-frame translations of the genome to generate potential protein products coupled with protein database alignments of contigs assembled from sequence reads. However, inverse protein folding and other protein design strategies can be used to generate viral protein variants that retain their normal function but have a radically altered primary amino acid sequence, which will escape detection based on protein or nucleic acid sequences. The detection of such an agent requires that the structures of all viral proteins be computed and compared with structural databases.

Another form of misuse would be to insert a toxin or virulence gene into the genome of an agent that infects humans but is not normally pathogenic. For example, nearly all humans are infected with polyomaviruses. These viruses result in lifelong infections but seldom cause disease. Like many viruses, polyomavirus gene expression is marked by extensive splicing, the process that joins discontinuous protein coding sequences (exons) interrupted by non-coding sequences (introns). Consequently, some viral genes are expressed via overlapping genes. A toxin or virulence factor could be inserted as an overprinted gene with the coding sequence interrupted by multiple introns. Such an agent would escape detection because the protein-coding section of the gene is not contiguous in the genome. However, such an agent could be detected by analysing the structures of six-frame translation products from RNA-seq experiments.

Discovery or synthesis of novel human pathogens. The discovery of new viral species is accelerating, and this is expected to continue. Each new viral species holds the potential to unlock new biological insights and applications. Therefore, each new virus also has the potential to be misused. Despite well-developed methods for identifying and propagating known human viral pathogens, new human pathogens are hard to create de novo. The factors that restrict a virus from becoming a human pathogen can be quite varied: inability to replicate well at human core body temperature, inability to recognise the human version of the virus's cell surface receptor and ineffectiveness of the viral proteins to interact with human proteins are just a few examples of biological restrictions a potential human pathogen must overcome. AI-driven viral genomics will enhance the ability to identify novel viruses that can potentially infect and cause disease in humans. It may also provide blueprints or pathways by which those novel, potential pathogens can be engineered to overcome some of the restrictions needed to replicate well in humans. The coupling of metagenomics, computation and microfluidic screens that can select for viruses that infect human cells and then modify them to be better at infecting humans can be misused to discover new bioweapons.

Advances in knowledge of the virus-host interactome and how these interactions contribute to each stage of the virus life cycle will ultimately lead to the AI-assisted design of new viruses de novo. Such viruses would be truly novel in that they have no counterpart in nature. Advances towards this goal are currently driven by research in gene therapy and oncolytic viruses. The potential for misuse of designer viruses is all too clear.

STRATEGIES FOR MITIGATION

The threats emerging from the intersection of biology and AI, though still largely theoretical, demand careful and proactive attention (figure 1; Mitigation and Oversight). One important step is to integrate safety measures directly into the technologies being developed. As computational platforms become increasingly capable of designing or modifying biological sequences, it is essential that those systems include embedded filters to screen for potentially dangerous outputs. These safeguards could rely on databases of known pathogens, toxins or immune evasion markers, flagging concerning designs before they are ever synthesised in a laboratory. As tools

evolve, so must their ability to self-regulate and support responsible use.

The majority of investigators using AI in virology may not fully realise the aspect of their research that could be used to engineer viruses that are increased biothreats and recognise the need for some level of guidance or oversight.[18] Developing standards that help researchers, institutions and policymakers recognise red flags early is critical. This includes establishing criteria for identifying when a proposed design, such as a viral sequence or synthetic protein, poses a dual-use concern, meaning it could be used for both beneficial and harmful purposes. These efforts should build on the regulatory frameworks that already govern biological research, such as biosafety and biosecurity protocols while expanding them to address the unique challenges posed by AI-assisted design. Creating systems that support expert oversight, encourage internal review and promote transparency can help ensure that risky work is identified and discussed.

Alongside prevention, detection capabilities must also evolve. Traditional sequence-based surveillance methods are powerful but can be evaded through clever genetic changes that preserve protein function while altering the underlying code. To counter this, there is a growing need for surveillance systems that incorporate protein structure analysis and functional prediction to recognise threats even when they do not resemble known pathogens at the sequence level.

Much like cybersecurity, where defences evolve in step with new attacks, ideas from adversarial AI may offer a useful parallel for thinking about faster and more adaptive safeguards in biology.

Just as access to high-risk biological agents is limited to labs with the right capacity and safeguards, access to advanced biological design tools should also be managed with care. A tiered system, based on the research purpose, institutional biosafety capabilities and oversight, can ensure that these tools are used only by those prepared to handle them responsibly. None of this will be effective without global cooperation. Science and technology move swiftly across borders, and the systems that govern them must do the same. Governments, academic institutions and international organisations must work together to create shared expectations for how high-risk biological research is reviewed, regulated, conducted and communicated. A shared commitment to transparency and accountability is essential to prevent misuse.

In the long run, we must also foster a broader shift in the culture of science itself. The creativity and drive that fuel discovery must be accompanied by an equal sense of responsibility. Researchers working at the frontiers of biology need not only resources and tools but also support in thinking critically about the impact of their work. Ethics, foresight and care must be as much a part of the research process as curiosity and ambition. Scientific progress and public safety are not in conflict; they depend on each other, and both are strengthened when we approach them with humility and intention.

CONCLUSIONS

The convergence of AI-assisted advances in protein structure and biochemistry, guided evolution, metagenomics and synthetic biology will facilitate and accelerate the generation of better therapies for a host of human diseases but could also be used to enhance the effectiveness or lower the sensitivity for the detection of bioweapons. More importantly, the convergence could greatly accelerate the generation of new classes of bioweapons with novel targets and mechanisms. These advances could fundamentally alter the biosecurity landscape and thus require novel modes of detection and mitigation. The dual use research conundrum—research that can benefit human health might also be co-opted to create viruses that pose even more of a risk to human health—is quickly becoming a clear and imminent problem that needs to be addressed to maintain the clear benefits of AI-driven research while minimising the potential risks. Harnessing the power of AI to help detect research that is designed with malicious intent can be a promising strategy to maintain the benefits of AI-driven research while minimising the chance of it being used for nefarious purposes.

Contributors All authors, SVK, DEN, AP and JMP, contributed equally to this article. JMP is the guarantor.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer-reviewed.

Data availability statement There are no data in this work.

ORCID iDs
Suresh V Kuchipudi https://orcid.org/0000-0002-8640-254X
Douglas E Norris https://orcid.org/0000-0002-4631-2777
Andrew Pekosz https://orcid.org/0000-0003-3248-1761
James M. Pipas https://orcid.org/0000-0003-1253-300X

REFERENCES

1 Ocana A, Pandiella A, Privat C, *et al*. Integrating artificial intelligence in drug discovery and early drug development: a transformative approach. *Biomark Res* 2025;13:45.
2 Shoemaker SC, Ando N. X-rays in the Cryo-Electron Microscopy Era: Structural Biology's Dynamic Future. *Biochemistry* 2018;57:277–85.
3 Jumper J, *et al*. *Nature New Biol* 2021;596:583–9.
4 van Kempen M, Kim SS, Tumescheit C, *et al*. Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* 2024;42:243–6.
5 Stucki A, Vallapurackal J, Ward TR, *et al*. Droplet Microfluidics and Directed Evolution of Enzymes: An Intertwined Journey. *Angew Chem Int Ed Engl* 2021;60:24368–87.
6 Agresti et al. Proc. *Natl Acad USA* 2010;107:4004–9.
7 Rotem A, Serohijos AWR, Chang CB, *et al*. Evolution on the Biophysical Fitness Landscape of an RNA Virus. *Mol Biol Evol* 2018;35:2390–400.
8 Pavlopoulos et al. *Nature New Biol* 2023;622:594–602.

9 Cantalupo PG, Calgua B, Zhao G, *et al*. Raw sewage harbors diverse viral populations. *MBio* 2011;2:e00180-11.
10 Mushegian AR. Are There 1031 Virus Particles on Earth, or More, or Fewer? *J Bacteriol* 2020;202:e00052–20.
11 Suttle CA. Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol* 2007;5:801–12.
12 Kuhlman B, Dantas G, Ireton GC, *et al*. Design of a novel globular protein fold with atomic-level accuracy. *Science* 2003;302:1364–8.
13 Kortemme T. De novo protein design-From new structures to programmable functions. *Cell* 2024;187:526–44.
14 Bulcha JT, Wang Y, Ma H, *et al*. Viral vector platforms within the gene therapy landscape. *Signal Transduct Target Ther* 2021;6:53.
15 Shalhout SZ, Miller DM, Emerick KS, *et al*. Therapy with oncolytic viruses: progress and challenges. *Nat Rev Clin Oncol* 2023;20:160–77.
16 Martínez-Espinoza I, Bungwon AD, Guerrero-Plata A. Human Metapneumovirus-Induced Host microRNA Expression Impairs the Interferon Response in Macrophages and Epithelial Cells. *Viruses* 2023;15:2272.
17 Yang L, Li A, Wang Y, *et al*. Intratumoral microbiota: roles in cancer initiation, development and therapeutic efficacy. *Signal Transduct Target Ther* 2023;8:35.
18 Walsh ME. Risk Anal. 2025.