Commentary

A Blind Spot in AI Biosecurity: Physics-Based Sequence Analysis

Veljko Veljkovic[1]

1. Biomed Protection (United States), Galveston, United States

Current approaches to artificial intelligence (AI) biosecurity focus predominantly on the capabilities and accessibility of data-trained models. These approaches emphasize training-data restrictions, model-level safeguards, output monitoring, and controlled access to computational resources. However, this framework may overlook a distinct category of computational methods: analytical approaches based on physically defined molecular descriptors and mathematical transformations that do not require model training. The Informational Spectrum Method (ISM), based on the Electron–Ion Interaction Potential (EIIP), provides a concrete example. Generative AI introduces an additional consideration by enabling biologically knowledgeable users without programming expertise to implement computational tools that were previously beyond their technical capabilities. The resulting biosecurity challenge is not the creation of new scientific knowledge but the transformation of existing knowledge into accessible operational capabilities. Although EIIP/ISM provides a current example, the underlying issue is general and may become increasingly relevant as other physics-derived computational methods emerge.

Correspondence: papers@team.qeios.com — Qeios will forward to the authors

1. The Current Focus of AI Biosecurity

The rapid development of AI applications in biology has stimulated growing concern about their potential misuse. Much of the resulting discussion focuses on machine-learning systems trained on biological sequences, molecular structures, experimental measurements, and other large datasets. [1][2]

This focus is understandable. Data-trained models offer several identifiable points of intervention. Training datasets may be curated, model capabilities evaluated, access restricted, and potentially harmful

 doi.org/10.32388/38R683

outputs monitored. These measures can reduce the accessibility of certain capabilities through controlled AI services.

However, they do not address every computational approach capable of generating biologically relevant information. Some established methods rely on physically defined molecular properties and deterministic mathematical operations rather than statistical learning. Their underlying calculations can be implemented independently of model weights, training datasets, or centralized AI services.

This creates a distinction between controlling access to a trained AI model and controlling the independent implementation of a published scientific method. The distinction becomes particularly important when generative AI can assist users in translating published mathematical procedures into executable software.

2. EIIP and ISM: A Concrete Example

The Electron–Ion Interaction Potential (EIIP) is a molecular descriptor derived from a pseudopotential-based representation of valence-electron interactions. Unlike descriptors obtained through statistical training on biological datasets, EIIP values are calculated from atomic electronic properties. They provide a physically motivated numerical representation that can be applied to biological sequences and molecular structures. [3]

The Informational Spectrum Method uses EIIP values to transform a biological sequence into a numerical signal. Fourier analysis of that signal produces an informational spectrum containing frequency components that may be associated with biological recognition and function. The method has been investigated in studies involving proteins, nucleic acids, molecular interactions, and functional consequences of sequence variation. [3][4]

An important historical investigation by Lazović in 1996 compared 226 amino-acid parameters and identified EIIP as particularly suitable for informational spectral analysis. Subsequent research has explored its application across different biological systems. [3]

The significance of this approach lies in the distinction between the physical basis of the calculation and the biological interpretation of its results. The numerical transformation itself does not require a machine-learning model trained on biological examples. It can be performed directly from a sequence and a defined set of EIIP values.

However, assigning a particular biological function to a spectral feature may require previous experimental evidence, comparative analysis, or independent biological validation. Recognizing this distinction is essential. The method is not independent of biological knowledge in every aspect of its interpretation, but its computational foundation is independent of statistical model training.

3. The Biosecurity Relevance of Sequence-Based Physical Analysis

Computational methods capable of examining functional consequences of sequence variation have legitimate applications in medicine, biotechnology, and basic biological research. They can contribute to investigations of molecular recognition, disease-associated mutations, therapeutic targets, and protein interactions.

At the same time, methods that provide information about biologically significant sequence changes may have dual-use implications. Published ISM research has included analyses of viral proteins and mutations associated with biological recognition. Such applications illustrate why sequence-based functional analysis is relevant to biosecurity discussions. [4]

The central concern is not that a computational prediction automatically translates into a successful biological intervention. Biological activity depends on multiple molecular and cellular factors, and computational predictions require appropriate validation. [2]

Rather, the concern is that computational methods capable of generating potentially relevant biological hypotheses may become more accessible to users who already possess the expertise needed to interpret and act upon them. This distinction is crucial for understanding the role of generative AI.

4. AI Removes a Technical Barrier, Not a Knowledge Barrier

One of the most consequential effects of generative AI is its ability to translate scientific descriptions, mathematical equations, and computational procedures into executable software. Traditionally, implementing a specialized analytical method required a combination of domain expertise, mathematical understanding, and programming skills. These capabilities are not always present in the same individual or research group. [1]

A biological laboratory may possess extensive knowledge of molecular mechanisms, experimental techniques, and biological systems while lacking the programming expertise needed to implement a particular computational method. Generative AI can substantially reduce this technical limitation.

The relevant user is therefore not necessarily someone without biological training. The more consequential case is a user who already possesses specialized biological knowledge but lacks the computational skills required to develop an analytical tool independently.

Such a user may already understand which biological questions to investigate, which molecular systems are relevant, and how computational results could inform subsequent research. AI does not need to provide that biological knowledge. It only needs to facilitate implementation of the computational component.

Consequently, the relevant capability emerges from the combination of three elements: existing scientific knowledge, AI-assisted computational implementation, and specialized biological expertise.

This combination can make previously difficult computational workflows practically accessible without requiring the development of a new scientific theory or a new trained AI model. The biosecurity implications therefore extend beyond the direct outputs of biological foundation models. They also involve the capacity of general-purpose AI systems to operationalize existing scientific methods.

5. Why Conventional Model-Level Controls Are Not Sufficient

Data-trained AI systems and deterministic physics-based analytical methods differ in their computational dependencies. A trained model relies on learned parameters and a particular implementation. Its availability can be influenced through model distribution, access policies, deployment restrictions, and monitoring. [1]

A published mathematical method has a different status. Its underlying equations and physical definitions can be independently implemented. Restricting access to a particular trained model does not eliminate the possibility of applying those equations through conventional software. Similarly, modifying or restricting a model's training data does not alter the mathematical foundations of an independently implemented method.

This does not mean that physics-based methods are entirely beyond governance. Software distribution, institutional policies, access to computational services, and responsible AI deployment can still influence how such methods are implemented and used. However, these mechanisms differ fundamentally from controls directed at the learned parameters of biological AI models.

The distinction is especially relevant when general-purpose AI systems can assist with software development. Even if a particular biological model is restricted, the scientific literature may continue to

contain the mathematical descriptions needed to implement an alternative analytical approach. Thus, biosecurity strategies focused exclusively on trained biological models may overlook computational capabilities that do not depend on those models.

6. A General Problem Illustrated by an Existing Method

EIIP/ISM provides a concrete example of a physics-derived computational method with a substantial history of biological applications. Its relevance to the present discussion does not depend on demonstrating that numerous equivalent methods already exist. Indeed, identifying another approach with the same combination of physical foundation, broad biological applicability, and documented practical use would require a separate comparative assessment.

The central argument is more fundamental. A single established example is sufficient to demonstrate that computational biological capabilities are not necessarily dependent on data-trained AI models.

This observation has implications for the design of biosecurity frameworks. A framework that assumes biologically relevant computational capabilities arise primarily through trained models may be incomplete even if relatively few alternative methods currently exist.

Moreover, future scientific developments may produce additional analytical approaches based on physical principles, mathematical transformations, or other computational foundations that do not require conventional model training. The present example therefore illustrates a potentially general problem rather than an isolated property of EIIP/ISM. Its importance lies not only in the capabilities of an existing method but also in the possibility that future methods could create similar challenges.

7. Implications for AI Biosecurity Governance

The distinction between trained biological models and independently implementable scientific methods suggests that AI biosecurity assessment should consider more than the capabilities embedded in a particular model. It should also examine how general-purpose AI assistance changes the practical accessibility of established scientific knowledge.

This requires distinguishing several levels of capability. First, a scientific method may be publicly documented and mathematically reproducible. Second, AI assistance may reduce the technical expertise required to implement that method. Third, a user with relevant domain knowledge may be capable of interpreting the resulting analysis and incorporating it into a broader experimental workflow.

These levels should not be treated as equivalent. The availability of a computational method does not guarantee the accuracy of its predictions or the feasibility of a particular experimental application. Nevertheless, reducing technical barriers can change the range of activities accessible to scientifically trained users.

Effective governance should therefore consider the interaction between user expertise, computational assistance, and downstream experimental capabilities. Measures such as sequence screening by synthesis providers, institutional biosafety oversight, responsible software deployment, and evaluation of AI-assisted scientific workflows remain important. No single measure can address every aspect of this problem. [1][2]

In particular, restrictions on biological foundation models cannot substitute for a broader assessment of computational capabilities that may be assembled from publicly available scientific methods. The objective should not be to restrict legitimate scientific knowledge or mathematical research. It should be to recognize how emerging AI capabilities alter the practical accessibility of that knowledge and to develop proportionate safeguards where justified by demonstrated risks.

8. Conclusion

AI biosecurity discussions have understandably concentrated on data-trained biological models and the risks associated with their capabilities. However, this focus may leave an important category of computational methods insufficiently examined.

Physics-based approaches such as EIIP/ISM demonstrate that biologically relevant sequence analysis can be performed using physically defined descriptors and mathematical operations without training a statistical model. Generative AI changes the practical significance of this distinction by making the implementation of established methods more accessible to users who possess biological expertise but lack specialized programming skills.

The resulting challenge is not simply the dissemination of scientific information. It is the conversion of existing knowledge into operational computational capability. EIIP/ISM provides a concrete current example, but the underlying issue is broader and may become more important as additional physics-derived analytical methods are developed.

The central question for AI biosecurity is therefore not only what biological knowledge an AI model has learned, but also what established scientific capabilities it enables knowledgeable users to implement.

Addressing this distinction would help make future biosecurity frameworks more comprehensive, technologically realistic, and adaptable to scientific progress.

Statements and Declarations

Funding

No specific funding was received for this work.

Potential Competing Interests

The author is a co-founder of Biomed Protection, LLC, which has commercial interests in applications of the EIIP/ISM methodology discussed in this manuscript. This affiliation and interest are disclosed for transparency.

Data Availability

Data sharing is not applicable to this commentary, as no datasets were generated or analysed.

Author Contributions

V.V. was the sole author and is responsible for all aspects of the manuscript.

References

1. [a], [b], [c], [d] *Wang M, Zhang Z, Bedi AS, et al. (2025). "A Call for Built-in Biosecurity Safeguards for Generative AI Tools." Nat Biotechnol. **43**:845–847. doi:10.1038/s41587-025-02650-8.*
2. [a], [b], [c] *National Academies of Sciences, Engineering, and Medicine (2025). The Age of AI in the Life Sciences: Benefits and Biosecurity Considerations. Washington, DC: National Academies Press. Chapter 3: AI-Enabled Biological Design and the Risks of Synthetic Biology. https://www.ncbi.nlm.nih.gov/books/NBK614591/.*
3. [a], [b], [c] *Lazović J (1996). "Selection of Amino Acid Parameters for Fourier Transform-Based Analysis of Protein s." Comput Appl Biosci. **12**(6):553–562. doi:10.1093/bioinformatics/12.6.553.*
4. [a], [b] *Schmier S, Mostafa A, Haarmann T, Bannert N, Ziebuhr J, Veljkovic V, Dietrich U, Pleschka S (2015). "In Sil ico Prediction and Experimental Confirmation of HA Residues Conferring Enhanced Human Receptor Speci ficity of H5N1 Influenza A Viruses." Sci Rep. **5**:11434. doi:10.1038/srep11434.*

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: The author is a co-founder of Biomed Protection, LLC, which has commercial interests in applications of the EIIP/ISM methodology discussed in this manuscript. This affiliation and interest are disclosed for transparency.