AIxBio

Hackathon

Apr 24, 2026 — Apr 26, 2026

MIZIN IGOR

&

SVYATOSLAV PIZANOV

EPICURUS

AI

BlueDot Impact

CBAI CAMBRIDGE BOSTON ALIGNMENT INITIATIVE

Measuring AI Progress

Sentinel Bio

Fourth Eon Bio

Building on four billion years of innovation to develop adaptive safeguards for biology

VERSION 1.0.1

team

SVYATOSLAV PIZANOV

MIZIN IGOR

We are a team focused on epidemiological issues in the context of AI safety. Our main mission is to support the well-being of our society and advance the Sustainable Development Goals. We develop ML models based on open data and strive to ensure that our solutions are transparent, scalable, and aligned with public interest

Contact email: viktoriunlidor@gmail.com

EPICURUS AI: From Disease Forecasting to Pathogen Prediction

Developed at AIxBio Hackathon (April 24–26, 2026)

Organized by Apart Research, BlueDot Impact, and Cambridge Biosecurity Hub

Track: Pandemic Early Warning (sponsored by Measuring AI Progress)

Abstract

EPICURUS AI began as a deep learning-based system designed to forecast global disease outbreaks using neural sequence models and advanced machine learning ensembles. The original approach integrated three architectures — RNN, LSTM, and Bidirectional LSTM — to capture long-term temporal patterns in over forty years of global disease data. In parallel, a feature-rich ensemble model using XGBoost, LightGBM, Random Forests, and Gradient Boosting was developed to enhance predictive accuracy and stability.

During the AIxBio Hackathon, we significantly expanded the system's capabilities. While the original EPICURUS focused on forecasting known diseases from historical case data, the hackathon enabled us to build a bidirectional pathogen prediction module that bridges three domains: epidemiology, molecular biology, and vector ecology. This new module — the Extended Pathogen Predictor — can infer pathogen characteristics from epidemiological parameters and, conversely, predict epidemiological behavior from molecular and ecological features alone. This capability is critical for assessing risks from novel or AI-generated pathogens for which no historical case data exists.

The expanded system now integrates 1) official WHO GLASS antimicrobial resistance data for 10 priority bacteria, 2) SeqScreen molecular virulence annotations for 27 vector-borne viruses, and 3) a curated master table of 37 pathogens with 30+ features spanning R_0, incubation period, case fatality rate, environmental stability, AMR priority, zoonotic potential, and vector ecology.

The original forecasting models were trained on cleaned and imputed WHO-style datasets covering the period from 1980 to 2024, generating country-specific and disease-specific predictions for 2025. The system achieved strong performance, with ensemble methods outperforming individual neural networks in both R^2 and RMSE scores. The new pathogen prediction module extends this foundation by enabling bidirectional inference: epidemiological profiles predict molecular traits (pathogen type, genome structure, AMR profile, virulence factors), while molecular features predict epidemiological outcomes (R_0, incubation, CFR, environmental stability).

Outbreak risk levels in the original system were derived using predicted case trajectories, trend ratios, and statistical confidence intervals. The expanded system adds molecular risk assessment — evaluating pathogenic potential from genomic and ecological features regardless of prior observation. A comparative evaluation revealed that while sequence-based models effectively capture temporal dependencies, the hybrid ensemble approach provides superior robustness in handling irregular patterns and data variability.

Looking forward, EPICURUS AI represents a proactive defense framework against emerging biosecurity threats, including AI-generated pathogens. By integrating real-time data collection with adaptive modeling capabilities and bidirectional molecular-epidemiological inference, the system enables rapid identification of shifting pathogen and population determinants during ongoing outbreaks. This work demonstrates how deep learning, ML ensembles, and multi-domain data integration can support early warning systems, public health forecasting, and global disease surveillance in an era of evolving biological threats.

Keywords: Disease forecasting, LSTM, time series modeling, outbreak prediction, epidemiological intelligence, AI, global health analytics, biosecurity, hybrid modeling, pathogen prediction, bidirectional inference, AMR surveillance, virulence factors, SeqScreen

Materials and Methods

Hackathon Context

The AIxBio Hackathon (April 24–26, 2026), organized by Apart Research, BlueDot Impact, and Cambridge Biosecurity Hub, brought together researchers, engineers, and biosecurity professionals to work on urgent intersections between AI and biological risk. EPICURUS AI was developed within the Pandemic Early Warning track, sponsored by Measuring AI Progress. The original Epicurus AI system, developed prior to the hackathon, integrated three major methodological pipelines: a visualization and analytics module, neural sequence models, and a feature-engineered machine learning ensemble. During the hackathon, we built a fourth pipeline — the Extended Pathogen Predictor — and significantly enhanced the data foundation by integrating multiple authoritative sources.

Pipeline 1: Epidemiological Analytics and Visualization Module

This component was built using Streamlit, Plotly, and Seaborn to explore over four decades of disease data in an interactive and interpretable manner. It provides tools for temporal trend analysis, geographic disease mapping, correlation analysis, PCA-based dimensionality reduction, and disease clustering. These visual examinations uncover hidden structures such as synchronized epidemic years, regional outbreak behaviors, and disease-specific temporal signatures. By placing this pipeline first, the project ensures that analysts and researchers start with a strong contextual understanding of the dataset before modeling. This module also serves as a validation layer, enabling users to compare model outputs against historical patterns and real-world epidemiological behavior.

Pipeline 2: Neural Sequence Modeling Framework

The second pipeline consists of neural sequence models using RNN, LSTM, and Bidirectional LSTM architectures implemented in PyTorch. These models receive rolling ten-year sequences of historical case values for each disease and country and are trained to predict the following year's case count. All values undergo MinMax normalization to stabilize training. The networks are optimized using the Adam optimizer combined with mean squared error loss, early stopping, and adaptive learning rate reduction. The LSTM and Bidirectional LSTM models capture long-

term temporal dependencies, nonlinear outbreak patterns, lagged effects, and repeated seasonal behaviors that traditional models struggle to interpret. They provide deep sequence understanding and serve as the core predictive engine for temporal forecasting.

Pipeline 3: Feature-Engineered Machine Learning Ensemble

While the neural models learn directly from raw sequences, this pipeline transforms historical data into a rich suite of quantifiable indicators. These include statistical descriptors such as mean, median, variances, and interquartile ranges; dynamic measures such as momentum, volatility, exponential moving averages, and trend slopes; and comparative metrics such as year-over-year changes and seasonal ratios. More than twenty engineered features are computed for each sample. These features are then used to train models such as XGBoost, LightGBM, Random Forest, Gradient Boosting, and ExtraTrees. Five-fold time series cross-validation identifies the strongest performers, and their predictions are combined using weighted averaging based on R^2 scores. This pipeline complements the neural networks by capturing mathematical structure patterns and statistical behavior that neural architectures may not specialize in.

Pipeline 4: Extended Pathogen Predictor (Hackathon Addition)

The fourth pipeline, developed during the AIxBio Hackathon, enables bidirectional prediction between epidemiological parameters and molecular/ecological features. This module is critical for addressing the core biosecurity challenge: assessing risks from novel or AI-generated pathogens that lack historical case data.

Data Integration

In our work, we utilized data provided by organizations such as the UN and the WHO. We employed population characteristics from the period 1980 to 2024, including demographic parameters and per capita GDP, as well as parameters of the epidemiological process, such as: long-term dynamics of the epidemiological process across various countries **[20, 21, 22, 23].**

And now we compiled a master dataset of 37 priority pathogens (10 bacteria, 27 viruses) with 30+ features **[24, 25, 26, 27, 28, 29, 30]**

Bidirectional Prediction Architecture

The module implements two inference directions:

Forward Prediction (Epidemiology → Molecular): users input epidemiological parameters

1) R_0,

2) incubation period

3) case fatality rate

4) environmental stability

5) transmission type

6) zoonotic status

7) vaccine availability.

The model predicts:

1) pathogen type

2) genome type according to Baltimore classification

3) AMR priority and key resistance concerns for bacteria

4) GLASS median resistance percentages

5) virulence factors for viruses

Reverse Prediction (Molecular → Epidemiology): users input molecular features

1) genome size

2) GC content

3) structural traits (enveloped, segmented, cytoplasmic replication)

4) pathogenic factors from SeqScreen annotations.

The model predicts:

1) R_0

2) incubation period

3) case fatality rate

4) environmental stability.

Models are trained using Random Forest, Gradient Boosting, and linear methods (Logistic Regression, Linear Regression), with features selected based on SHAP importance scores and cross-validated on the curated 37-pathogen dataset.

WHO GLASS Data Lookup

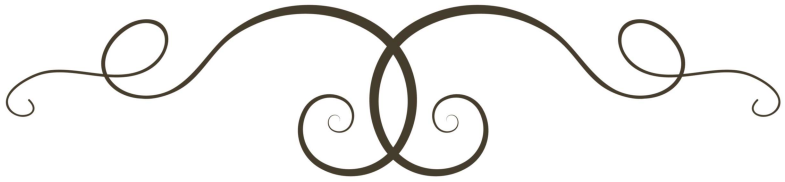
A dedicated lookup interface provides interactive access to WHO GLASS AMR data, including: 1) country-level surveillance indicators (94 enrolled, 78 reporting countries/territories/areas), 2) regional breakdowns (Africa, Americas, Eastern Mediterranean, Europe, South-East Asia, Western Pacific), 3) AMR trend graphs for critical priority pathogens, 4) SDG 3.d.2 indicators for bloodstream infections.

Integration of Pipelines

Together, these four pipelines provide a robust hybrid modeling ecosystem that spans temporal disease forecasting, statistical pattern analysis, and molecular-epidemiological inference. The

visualization module offers interpretability and exploratory insight; the neural sequence models capture temporal dynamics; the ensemble pipeline adds statistical precision and stability; and the Extended Pathogen Predictor bridges molecular biology with epidemiology for proactive biosecurity assessment. Combined, they enable EPICURUS AI to generate accurate, transparent, and epidemiologically meaningful predictions — whether for known diseases from historical data or for novel pathogens from molecular and ecological features alone.

CONTENT



Challenges

Math and AI in pathogen monitoring

Our solution

Conclusions

«If you live according to nature, you will never be poor. If you live according to people's opinions, you will never be rich.»

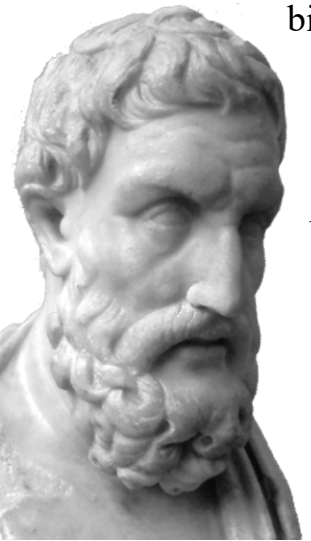
- Epicurus

Challenges

The COVID-19 pandemic has profoundly altered humanity's perception of biological threats. While this may seem a trivial observation to some, it is worth reiterating that in our modern world, the perceived utility of funding scientific research is largely dictated by the desires of the general public and large corporations, rather than by specialists and scientists. That said, epidemiology has undoubtedly fared better in this new climate than, for example, formal semantics. The heightened interest from states and major corporations in the field of biosecurity confirms that our research and development are more relevant than ever.
On one hand, the echoes of the COVID-19 pandemic have not yet faded. Some studies suggest that the pandemic period saw a rise in mental health disorders and eating pathologies **[1, 2]**. Conversely, other research indicates that the pandemic made people more disciplined and improved their attitude towards monitoring their own health **[3]**. This article is not intended to be a literary review on public attitudes and behaviors during the pandemic; however, one thing can be stated with certainty: the COVID-19 pandemic has left a lasting mark on humanity.

On the other hand, a new anxiety is emerging, driven by the rapid development of generative artificial intelligence. Society is now actively seeking new ways to regulate large language models and to quickly identify threats stemming from their potential use by malicious actors.

A particularly intriguing area, in our view, is the use of AI as a tool for generating a countless number of novel pathogens through molecular-genetic and structural modeling. This convergence of biology and advanced computation presents a paradigm shift in the nature of biological threats, moving the risk from the natural world to the digital realm. Consequently, our defensive strategies must evolve with the same speed and ingenuity as the offensive capabilities these technologies enable. The recent pandemic was a stark lesson in our vulnerability to natural pathogens; the next challenge may come from artificially engineered ones, designed not by nature, but by algorithms.



We are already witnessing the emergence of artificial intelligence models capable of drastically accelerating biological research. These systems integrate multiple levels of biological organization — from molecular to epidemiological — and fuse this multimodal understanding with the entire corpus of global scientific literature. A prime example is Evo 2, which leverages the novel StripedHyena 2 architecture, scaled to 40 billion parameters. This allows the model to identify long-range dependencies across a context window of 1 million tokens — a phenomenal leap from Evo's previous capacity of 131,072 tokens **[4]**.

This exponential increase in processing power, driven by architectural innovations **(fig. 1)**, represents a pivotal moment. In our view, it also marks a significant and dangerous escalation in the potential for creating sophisticated biological weapons. The very capability that allows for the rapid discovery of therapeutics — the deep integration of disparate biological data at scale—can be inverted to design pathogens with enhanced transmissibility or virulence. Therefore, the architectural breakthroughs that promise to unlock the secrets of life simultaneously lower the barrier to engineering its most potent threats.

The same architectural efficiency that allows StripedHyena to process vast scientific contexts also makes it a more scalable and potentially accessible tool for malicious actors. This creates a perverse incentive where progress in AI's biological understanding is inextricably linked to progress in its potential for misuse. Consequently, the scientific community must develop safeguards and governance frameworks that evolve in lockstep with these rapidly advancing capabilities.

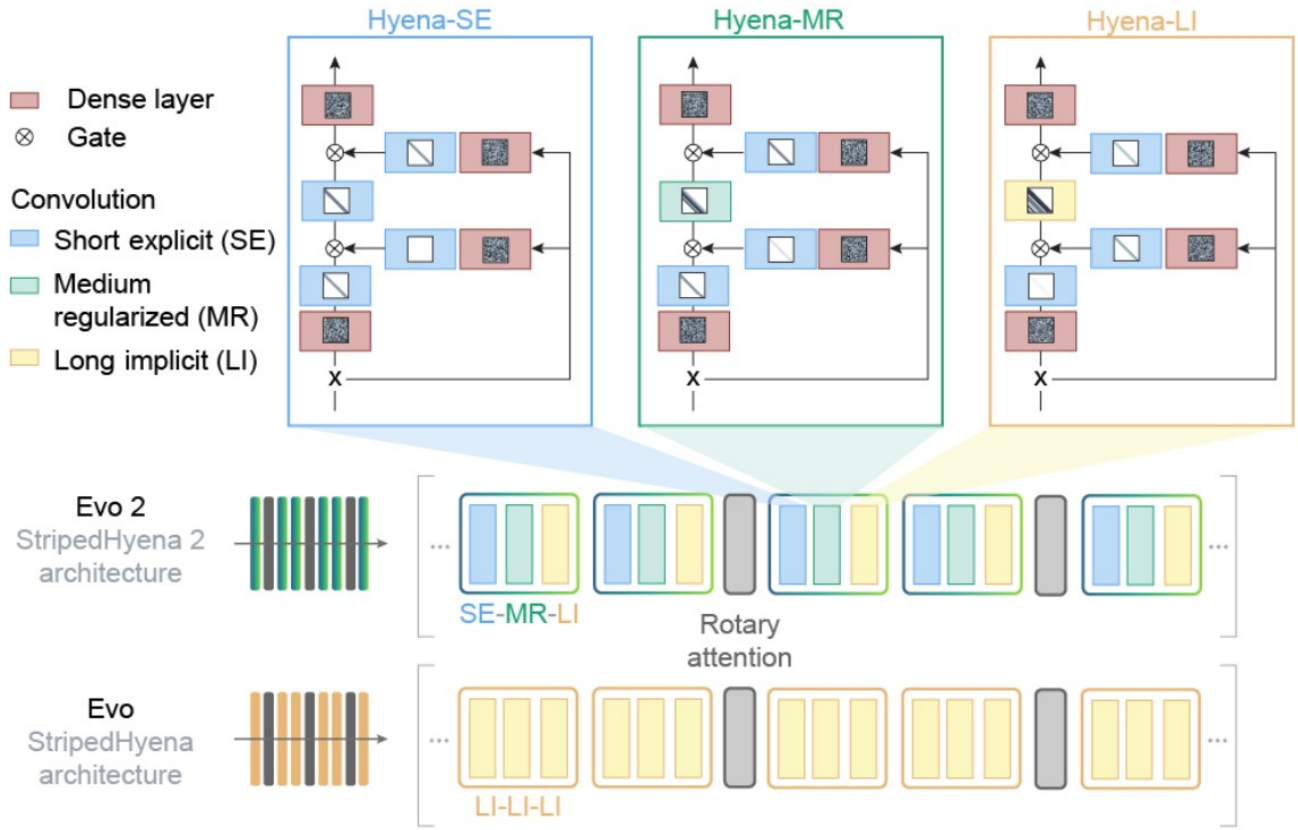


Figure 1. *Evo and Evo 2 AI model architecture. Tretina, K. (2025, February 19). Understanding the Language of Life's Biomolecules Across Evolution at a New Scale with Evo 2. NVIDIA Technical Blog.*

Another pressing question is the extent of our preparedness to counter such threats. Currently, it is difficult to specify the exact number of vaccines an individual's immune system can accommodate. The prevailing scientific view holds that the immune system is not a simple «vessel» to be filled. It is designed for constant engagement with a vast array of antigens; every day, we encounter thousands of new pathogens through food, air, and contact with other people. The primary challenge, therefore, lies not in overloading the immune system, but in the immense scientific and logistical hurdles of vaccine development and deployment, which include:

1. **The «Valley of Death»**: The transition from laboratory research to clinical trials is extremely difficult, hampered by challenges in manufacturing, vaccine stability, and developing assays to measure immune response.

2. **Stringent Safety Requirements**: Since vaccines are administered to healthy populations, their safety requirements are exceptionally high. Large-scale clinical trials involving tens of thousands of participants are necessary to detect even rare side effects.

3. **Complex Pathogens**: For diseases like tuberculosis, malaria, and HIV, creating an effective vaccine is incredibly challenging due to the biological complexity of the pathogens themselves, including their variability and ability to evade immune detection. The problem is not that the body cannot handle another vaccine, but that scientists have so far been unable to create a working vaccine that elicits the necessary immune response.

4. **Economics and Logistics**: Developing a single vaccine costs hundreds of millions of dollars. For diseases prevalent in poor countries, there is little commercial incentive for pharmaceutical companies. Even when a vaccine is successfully created, scaling up its production and delivering it to those in need remains a monumental task **[5]**.

Compounding these existing challenges is the accelerating pace of artificial intelligence. Recent statements by Sam Altman suggest that by 2028, AI could become an autonomous, independent researcher **[6]**. On one hand, this could unlock new horizons across all scientific fields; on the other, it could drastically simplify the process for malicious actors to create new and highly dangerous types of weaponry.

This prospect raises profound concerns for global and international security. There is a significant difference between such technologies falling into the hands of a small terrorist group and their acquisition by rogue regimes that have occupied entire nations and possess established laboratory infrastructures. The latter scenario would necessitate a truly coordinated response from both the political and scientific communities.

The potential dangers are amplified by the advanced capabilities scientists now possess for designing novel biological molecules **(fig. 2)**. Recent research conducted by Watson, J.L., Juergens, D., Bennett, N.R. et al. exemplifies a revolutionary breakthrough in computational protein design with the emergence of RFdiffusion. This generative model, based on diffusion networks, enables the creation of «de novo» protein structures with atomic-level accuracy,

significantly surpassing previous methods. Its versatility is demonstrated by its ability to solve diverse tasks—from unconditional monomer design and the creation of symmetric oligomers to scaffolding functional motifs and developing high-affinity binding proteins. Experimental validation confirmed exceptional precision; for instance, the cryo-EM structure of a designed influenza hemagglutinin binder is nearly identical to its computational model (RMSD 0.63 Å). The method demonstrates unprecedented efficiency, increasing the probability of experimental success in creating binding proteins by a factor of a hundred. RFdiffusion provides researchers with an intuitive tool for creating functional proteins based on minimal specifications, analogous to generating images from text prompts. While the open availability of such code promotes scientific progress, it justifiably intensifies concerns about potential misuse by malicious actors, despite the system's current limitations which preclude the direct design of pathogens or toxins **[7]**.

The integration of AI into this domain creates a perfect storm: it accelerates the design of novel bio-molecules while simultaneously lowering the expertise barrier for their creation. We are therefore approaching a critical juncture where our defensive capabilities, burdened by slow and costly traditional development, may be fundamentally outpaced by offensive possibilities. This asymmetry could become the single greatest threat to biosecurity in the coming decade, demanding pre-emptive and collaborative international governance.

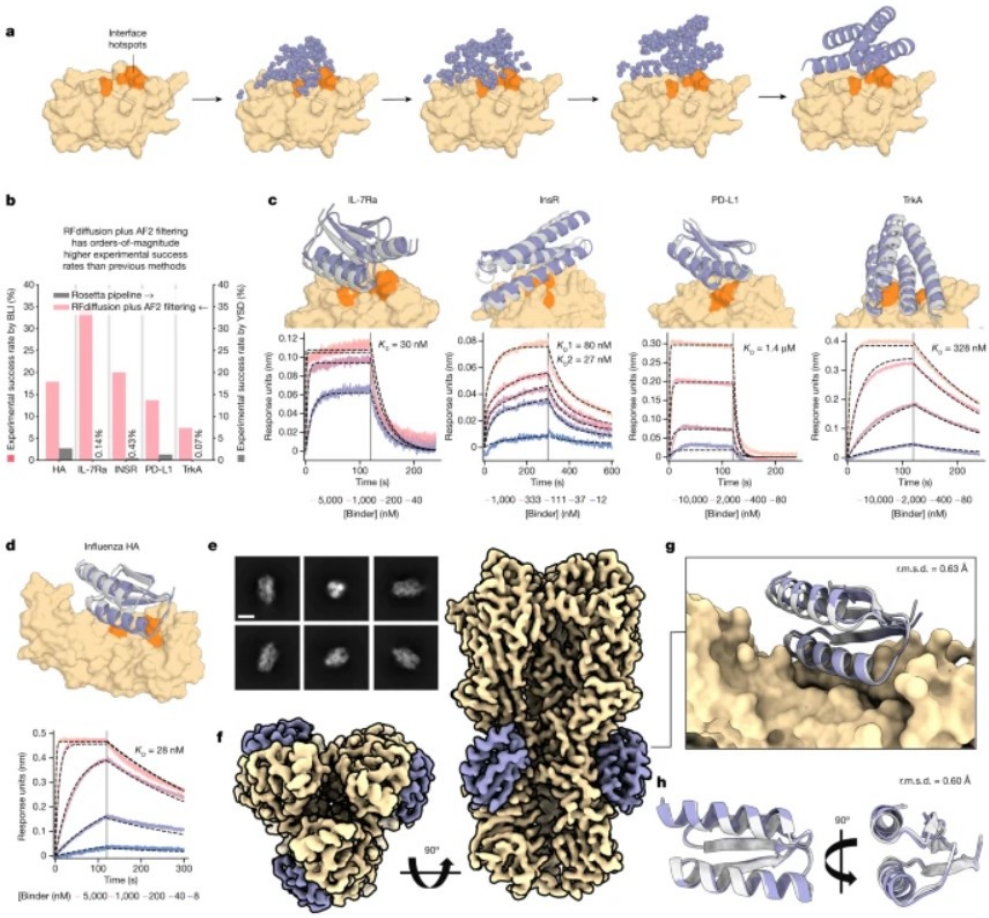


Figure 2. *De novo design of protein-binding proteins. Watson, J.L., Juergens, D., Bennett, N.R. et al. De novo design of protein structure and function with RFdiffusion.*

It is also striking how easily and in how many different ways genome editing technologies are developing in the 21st century within wet biology **(fig. 3)**. To date, there are numerous methods

for endowing microorganisms with new unique properties. And even if therapeutic developments today are unable to guarantee a 100% safety rate, one can only imagine the horrifying consequences that await us once malicious actors become aware of such technologies **[8]**. The democratization of these powerful tools, while fostering scientific progress, simultaneously lowers the barrier for their misuse. The same techniques that promise to cure genetic diseases could be twisted to engineer pathogens with enhanced virulence or novel transmission routes. Consequently, the scientific community faces a dual imperative: to accelerate the pace of discovery while erecting robust ethical and biosafety frameworks to prevent a catastrophe.

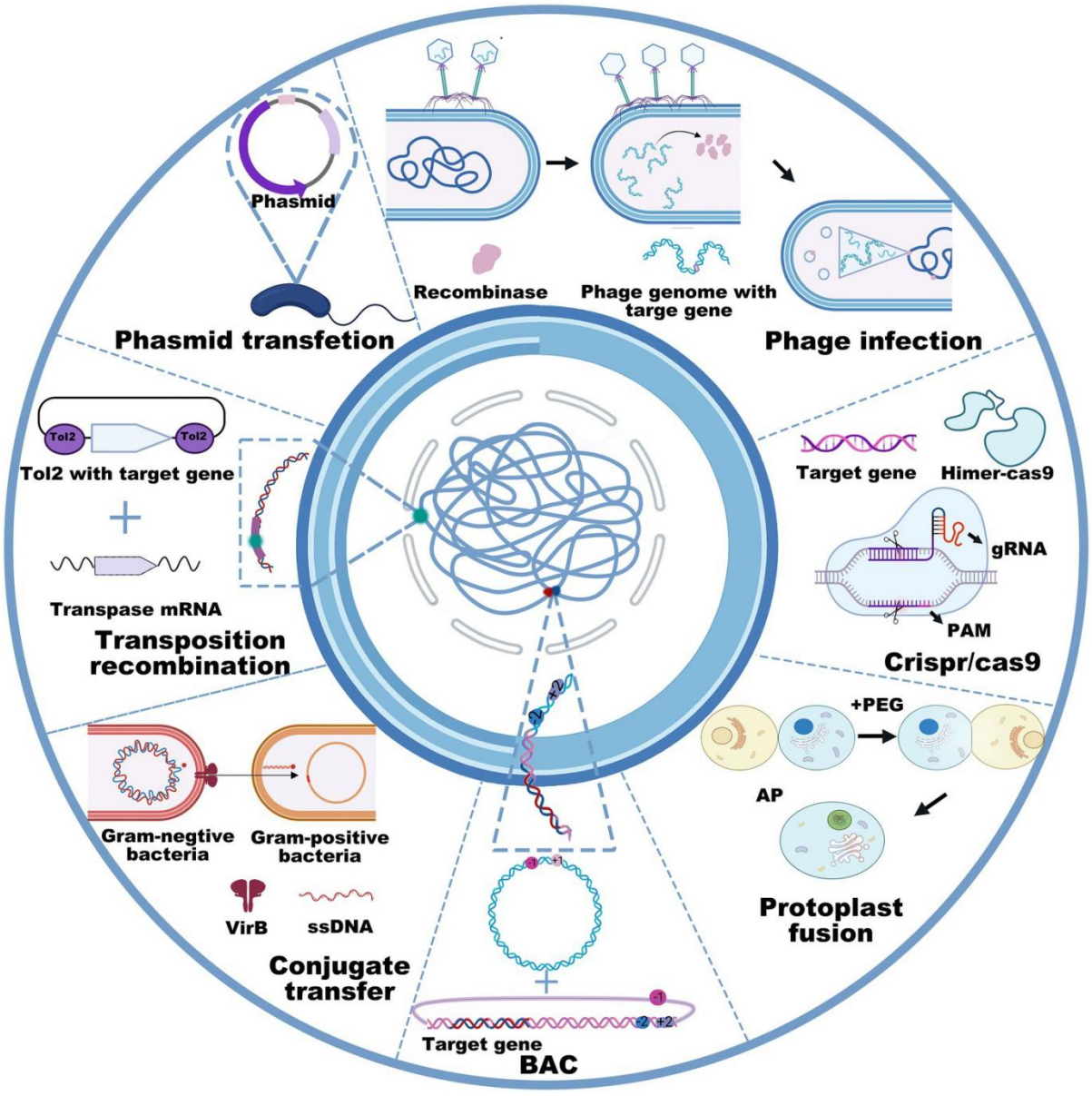


Figure 3. *General used recombination technologies, including phage infection, phasmid transfection, transposition recombination, conjugate transfer, BAC, protoplast fusion, and CRISPR-Cas9 in counter-clockwise order. Liu, Y., Feng, J., Pan, H., Zhang, X., & Zhang, Y. (2022). Genetically engineered bacterium: Principles, practices, and prospects. Frontiers in Microbiology, 13, 997587. https://doi.org/10.3389/fmicb.2022.997587.*

This presents us with an undeniable and pressing question: if malicious actors can rapidly generate hundreds of thousands of novel pathogens, is the global pharmaceutical industry, in its current state, prepared to mount a swift and effective response? Furthermore, can we possibly predict the collective impact of so many vaccines on the human organism?

While future research will undoubtedly need to address these questions from the perspectives of both toxicology and novel epidemiological methodologies, our immediate focus is on a more pragmatic challenge: how to effectively counter modern epidemiological threats using accessible, contemporary, and efficient tools. This is precisely the goal, in our view, pursued by leading AI safety companies and startups such as Anthropic, Apart Research, and Blue Dot Impact.

Consequently, we contend that this defines a new, critical mission for epidemiology: the development of cheap and effective mechanisms for the epidemiological control of pathogens designed by large language models. This mission is to create a strategic shield, a proactive defense system to protect humanity from the looming sword of this emerging danger. This necessitates a paradigm shift from reactive vaccine development to proactive pathogen-agnostic surveillance and containment. The very nature of the threat demands that our defensive tools be as scalable, adaptive, and computationally-driven as the offensive ones. Therefore, the next frontier in biosecurity lies not just in faster labs, but in predictive epidemiological networks powered by AI, capable of modeling and mitigating outbreaks before they even begin.

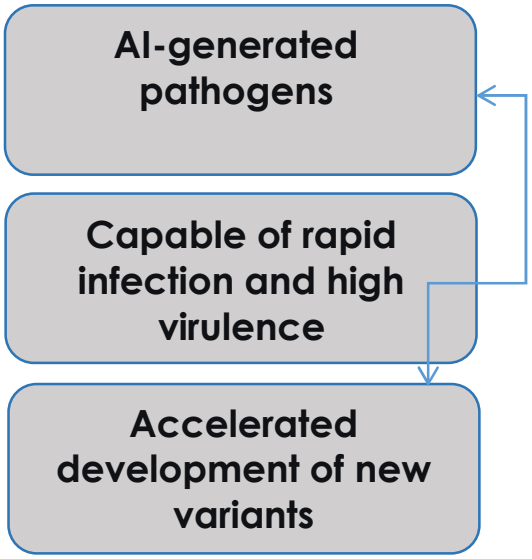


Figure 4. *Properties of pathogens created using AI.*

Math and AI in pathogen monitoring

"It is better to be miserable with reason than to be happy without it"

- Epicurus

Historical Development and Modern Approaches in Mathematical Epidemiology

The historical development of mathematical epidemiology is rooted in the foundational work of Daniel Bernoulli, who in 1760 evaluated the effectiveness of smallpox inoculation. Later, in 1840, William Farr described smallpox mortality data in England and Wales using the normal distribution curve. A statistical approach to epidemic theory was advanced by John Brownlee in his 1906 paper **[9]**. Subsequently, a cornerstone contribution was made by Kermack and McKendrick in 1927, who formulated the deterministic SIR model, introducing the "law of mass action" into epidemiology **[10]**.

The mid-20th century saw the rise of probabilistic approaches. McKendrick laid the groundwork with a stochastic SIR model in 1926. The Reed-Frost model, developed around 1928-1930 and published in 1952, along with Greenwood's work (1931), established chain-binomial models. Bartlett's research on the stochastic SIR model in continuous time, beginning in 1949, was pivotal for developing stochastic models of epidemic processes **[11, 12]**. Further significant contributions to the application of random process theory were made by Bailey and Whittle.

Spatial modeling was pioneered by Kendall, who formulated one of the first spatial models using partial differential equations in 1957 **[13]**. In the same year, Bartlett described epidemic spread on a spatial lattice using Monte Carlo methods, an early example of computational epidemiology **[12]**. The advent of individual-based modeling is marked by the work of Elveback, Ackerman, Gatewood, and Fox, who published the first disease spread model accounting for individual agents in 1971 **[14]**.

To this day, deterministic compartmental models, particularly the SIR framework, remain a fundamental tool. These models, described by systems of ordinary differential equations, simulate disease spread in large, homogeneous populations. Numerous modifications exist, such as SIRS (temporary immunity), SEIR (incubation period), and SIS (no immunity), to adapt to different infection types. The classical Kermack-McKendrick model is defined by the system:

$$\begin{aligned} dS/dt &= SI; \\ dI/dt &= \beta SI - \gamma I; \\ dR/dt &= \gamma I. \end{aligned} \qquad (1)$$

To this day, deterministic compartmental models, particularly the SIR framework, remain a fundamental tool **[10]**. These models, described by systems of ordinary differential equations, simulate disease spread in large, homogeneous populations. Numerous modifications exist, such as SIRS (temporary immunity), SEIR (incubation period), and SIS (no immunity), to adapt to different infection types.

Stochastic population models, utilizing Markov chains, birth-and-death processes, and systems of stochastic difference equations, are crucial for modeling small populations or early epidemic stages where randomness is significant **[11]**. A contemporary example is the stochastic SEIR model with discrete time presented by Lekone and Finkenständt for analyzing the 1995 Ebola outbreak **[15]**. In this model, new infections, transitions to the infectious state, and removals

(recoveries/deaths) are given by binomial distributions with probabilities derived from transmission rates, incubation period duration, and infectious period duration.

To account for spatial heterogeneity, metapopulation models link multiple SIR models for different regions. Cellular automata employ regular lattices with deterministic or stochastic transition rules for node states. Network models offer greater flexibility, simulating arbitrary social or spatial connections between individuals, enabling the study of population clustering and the role of "superspreaders." Individual-based and agent-based models allow each agent to be endowed with a unique set of parameters describing internal traits, spatial location, and social status **[14]**. A key characteristic of multi-agent models is that individuals act as independent entities with behavioral rules, leading to emergent system properties.

Multi-component models integrate different modeling approaches. An example is the BioWar model, designed to simulate an epidemic following a bioweapon attack in an urban setting, which requires extensive input data for calibration. Models are applied to three key tasks: investigating epidemiological processes (hypothesis testing), forecasting disease dynamics (resource planning), and comparing the effectiveness of control measures such as quarantine and vaccination.

Modern challenges include modeling the interaction between infectious and non-infectious diseases, for instance, the link between HPV and cervical cancer, necessitating new model classes. Another complex problem is modeling superinfection, which is crucial for diseases like HIV and tuberculosis and complicates the traditional model concept, requiring a revision of the basic reproduction number. The development of accurate deterministic approximations for stochastic models is an active area of research **[12]**. Nonlinearities in transmission dynamics necessitate accounting for covariances. Methods like the multivariate normal closure approximation (MVN) are used for approximations beyond the simple mean-field model.

SIR model:

$$\begin{aligned} dS(t)/dt &= S(t)I(t)/N; \\ dI(t)/dt &= \beta S(t)I(t) - \gamma I(t); \\ dR(t)/dt &= \gamma I(t). \end{aligned} \qquad (2)$$

$$S(t)+I(t)+R(t)=N \qquad (3)$$

S(t)	**Susceptible**	Number of individuals who are susceptible to the disease at time t.
I(t)	**Infectious**	Number of individuals who are currently infected and can spread the disease at time t.
R(t)	**Recovered**	Number of individuals who have recovered from the disease and gained immunity at time t.

N	**Total Population**	The total population size is constant.
β (beta)	**Transmission Rate**	Intensity of contacts leading to infection (contact frequency × transmission probability)
γ (gamma)	**Recovery Rate**	Intensity of recovery of infected individuals, γ = 1 / (Average Infectious Period)
R_0	**Basic Reproduction Number**	$R_0 = \beta / \gamma$ Average number of secondary infections from one infected individual in a fully susceptible population $R_0 > 1$ - disease spreads $R_0 < 1$ - disease dies out

***SIER model*:**

$$\begin{aligned} dS/dt &= \mu N[1-p] - S) - \beta ISN; \\ dE/dt &= \beta ISN - (\mu + \sigma)E; \\ dI/dt &= \sigma E - (\mu + \gamma)I; \\ dR/dt &= \gamma I - \mu R + \mu Np. \end{aligned} \quad (4)$$

S	**Susceptible**	Number of individuals who can be infected.
E	**Exposed**	Number of infected individuals in the incubation (latent) period. They are not yet infectious.
I	**Infectious**	Number of individuals who are actively sick and can spread the disease.
R	**Recovered**	Number of individuals who have recovered and gained immunity.
N	**Total Population**	$N = S + E + I + R$ (The total population size is constant). (5)

SIS model:

$$\begin{aligned} S &= bI - aSI; \\ I &= aSI - bI. \end{aligned} \quad (6)$$

S	**Susceptible individuals**
I	**Infected individuals**
a (or β)	**Transmission rate (infection rate)**
b (or ν, γ)	**Recovery rate**

Today, there exists a vast number of biological modeling methods, a complete analysis of which requires considerable time. It is important to note, however, that most of these models exhibit high sensitivity to changes in their initial parameters. A prime example of this characteristic is the classical Lotka-Volterra mathematical model.

$$\begin{aligned} dx/dt &= (\alpha - \beta y)x; \\ dy/dt &= (-\gamma + \delta x)y. \end{aligned} \tag{7}$$

α	**Prey population growth rate**
β	**Probability of prey being eaten upon contact with a predator**
γ	**Probability of predator dying from starvation**
δ	**Probability that predator has sufficient food for reproduction**

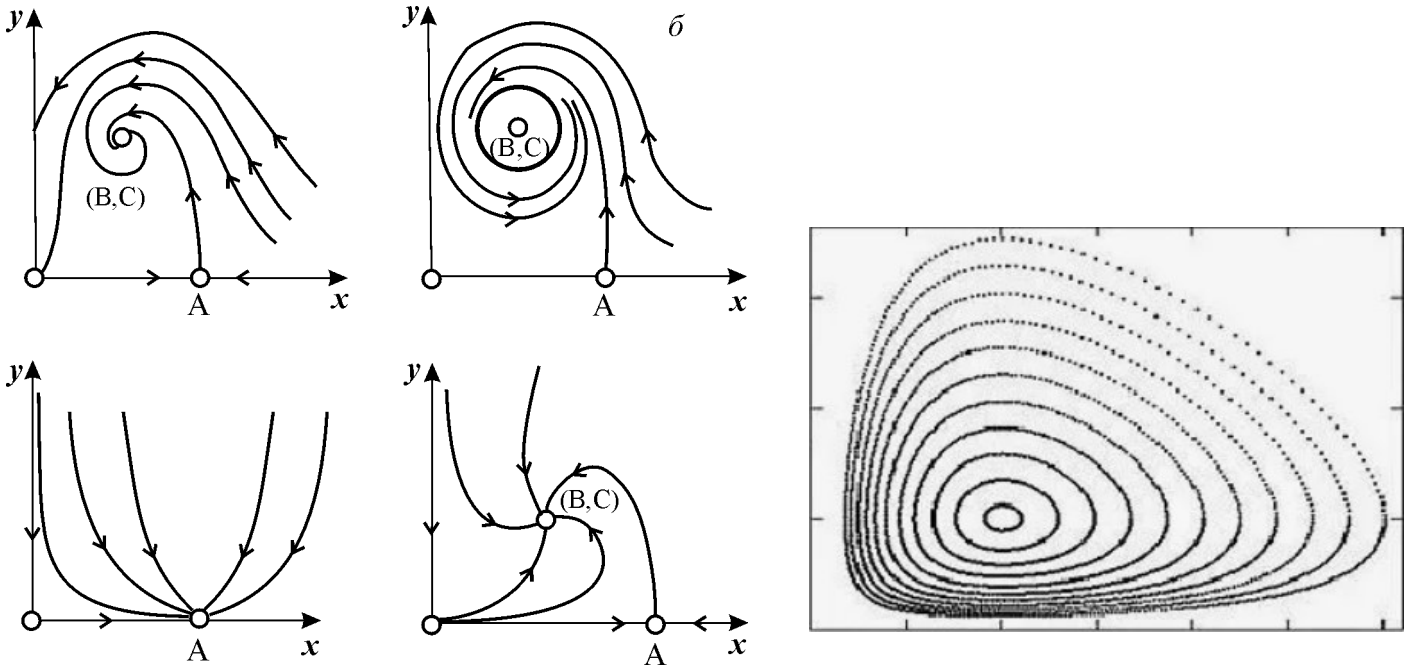


Figure 5. *Visual representation of the Lotka-Volterra equations.*

Contemporary artificial intelligence models for epidemic forecasting analyze large volumes of data to identify patterns and trends that may indicate potential disease outbreaks. This enables governments and health institutions to respond in a timely manner by deploying resources and implementing preventive measures.

The epidemiological forecasting model PandemicLLM, developed by Hongru Du, Yang Zhao, Jianan Zhao, Shaochong Xu, Xihong Lin, Yiran Chen, Lauren M. Gardner & Hao 'Frank' Yang, represents an innovative system that converts multimodal data into textual formats for processing by large language models. This approach enables the integration of heterogeneous information, including textual descriptions of public health measures, genomic surveillance, and spatiotemporal epidemiological data. Having demonstrated superiority over traditional methods when tested on COVID-19 pandemic data across all 50 U.S. states, the model opens new possibilities for real-time disease spread forecasting using heterogeneous data. **[16]**
Another major project, Outbreaks Near Me (formerly Flu Near You), is a crowdsourced epidemiological monitoring system that has been operating in the United States and Canada since 2011. The project enables volunteers to report symptoms of influenza-like illnesses weekly through an online platform and mobile application. The collected data show a high correlation with the official CDC ILINet surveillance system in terms of the timing and intensity of infection spread. As part of the project's development, specialized dashboards were created for public health authorities, and international EpiHack hackathons were held to integrate the data into government monitoring systems. The project has confirmed the effectiveness of the crowdsourcing approach for operational tracking of epidemiological trends. **[17]**

Thus, most existing epidemiological forecasting models today represent several distinct types of models:

Real-time outbreak detection models:
Systems for early identification of epidemic outbreaks using streaming data sources and anomaly detection algorithms.

Machine learning methods for infection spread prediction:
Computational approaches leveraging statistical learning and AI to forecast disease transmission dynamics and outbreak trajectories.

Intervention-aware epidemiological models:
Mathematical frameworks that simulate and predict infection spread under different public health measures and control strategies.

Contemporary research confirms that machine learning is effective for discovering new antibiotics and predicting the antimicrobial activity of compounds, yet shows limitations in predicting their specific mechanisms of action. Artificial intelligence algorithms can model virus evolution and identify potential targets for vaccine development, though the accuracy of these predictions directly depends on the completeness and representativeness of source data. In diagnostics, AI shows promising results, but most existing models still fall short of the accuracy required for routine clinical application. Key unresolved challenges include limited model generalizability to new pathogen classes and persistent risks of bias associated with insufficient diversity in training data **[18, 19]**.

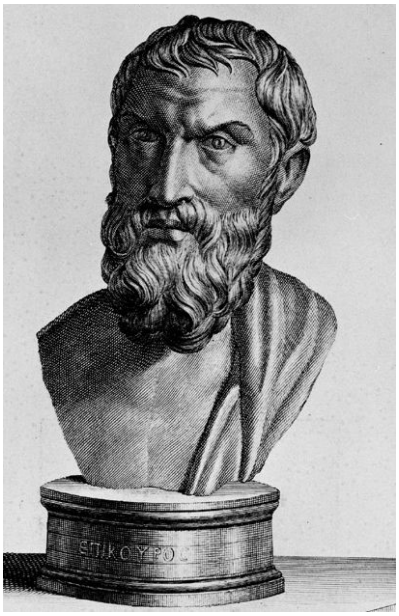
In our view, the fundamental problem of contemporary epidemiological modeling using artificial intelligence lies in its insufficient integration with fundamental biological research and its inadequate treatment of populations as distinct levels of biological organization requiring systematic investigation. While these models perform adequately during stable periods and for managing well-studied infections with established prevention methods, they would prove insufficiently responsive and effective during biological terrorist attacks employing AI-generated pathogens. Furthermore, we must immediately initiate AI-driven development of countermeasures against biological weapons utilizing nucleotide sequences and structural organizations generated by models such as Evo.

These advanced countermeasures should incorporate real-time biosurveillance systems capable of detecting novel pathogen signatures through integrated genomic and environmental monitoring. Additionally, developing adaptive response frameworks that can rapidly simulate outbreak scenarios and optimize resource allocation will be crucial for mitigating the impact of engineered biological threats.

Our solution

EPICURUS

Pathogen and Population Characteristic Forecasting System



«Necessity is a disaster, but there is no necessity to live with necessity.»

-Epicurus

The 21st century presents new biosecurity threats emerging from the capabilities of artificial intelligence. As discussed throughout this work, models like Evo 2 and RFdiffusion now enable the rapid design of novel biomolecules, lowering the expertise barrier for pathogen engineering. This is a complex problem that requires a comprehensive solution from policymakers, legal experts, ML specialists, doctors, and biologists.
Our solution represents an innovation in safeguarding biosecurity against biological threats, whether natural or engineered. We have developed the first prototype of a "shield" designed to provide a foundation for fundamental population-level research and proactive pathogen assessment. Our concept is based on several key premises:

AI will enable malicious actors to rapidly generate vast quantities of new pathogens with high degrees of virulence and pathogenicity.

The current state of "emergency" and "military" epidemiology is insufficient. For a long time, we have dealt with established pathogens and have not yet achieved decisive victories. What will happen when we face mass-scale terrorist biological attacks?

The vaccinology and pharmaceutical industries are not prepared to rapidly scale up the production of new vaccines and drugs. Ensuring vaccine safety requires lengthy clinical trials. Furthermore, even a single adverse effect could lead to millions of casualties and fuel anti-vaccination sentiments.

The earlier an epidemic is combated, the more effective the response will be.

Any epidemiological process is highly sensitive to the properties of both the susceptible population and the infectious agent.

Different pathogens exhibit different epidemiological curves.

While current mathematical models are imperfect, their refinement is only possible in close collaboration with AI capable of analyzing both localized outbreaks and multi-year epidemic dynamics.

EPICURUS AI 1. 0. 1

To address the growing challenges in epidemiological forecasting and biosecurity, we developed EPICURUS AI — a comprehensive, web-based analytical platform that integrates classical epidemiological modeling with state-of-the-art machine learning, molecular trait analysis, and ecological data. Originally built for global disease forecasting, the system was significantly expanded during the AIxBio Hackathon (organized by Apart Research, BlueDot Impact, and Cambridge Biosecurity Hub) to include bidirectional pathogen prediction capabilities. This solution provides public health officials, researchers, and biosecurity analysts with an intuitive yet powerful tool for understanding disease dynamics, predicting outbreak trajectories, assessing pathogen similarity, and evaluating risks from novel or AI-generated pathogens.

Core Architecture

The platform is built as a unified Streamlit application, centered around a Python-based analytics engine. This architecture ensures modularity, scalability, and ease of maintenance.

Key Technical Components

Frontend: An interactive web interface built with Streamlit, enabling real-time parameter adjustment and visualization. The interface supports both forward prediction (epidemiology → molecular) and reverse prediction (molecular → epidemiology) workflows.

Backend & Analytics: A Python engine handling data processing, model training, simulation, and bidirectional inference. Models are trained on curated datasets and persisted for instant loading.

Machine Learning: Ensemble algorithms from Scikit-learn, XGBoost, and LightGBM for robust pattern recognition. Multiple model types — Random Forest, Gradient Boosting, Logistic Regression, and Linear Regression — are switchable at runtime.

Epidemiological Modeling: Implementation of foundational compartmental models (SIR, SEIR) using SciPy for outbreak simulation and R_0 estimation.

Molecular Trait Analysis: Integration of SeqScreen functional annotations — viral adhesion, host xenophagy, immune evasion, viral invasion, and host cell death — as predictive features for virulence assessment.

Data Visualization: Dynamic charts generated with Plotly and Matplotlib to illustrate model outputs, AMR trends, and confidence intervals.

Key Functionalities

1. Multi-Disease Analytical Engine

The system integrates curated data for 37 priority pathogens (10 bacteria, 27 viruses) sourced from WHO GLASS, WHO BPPL 2022, NCBI Virus, and peer-reviewed literature. It enables comparative analysis across pathogens, regions, and molecular profiles, identifying patterns and anomalies through advanced analytical techniques. The dataset includes R_0, incubation period, case fatality rate, transmission route, vector type, zoonotic status, environmental stability, genome type, AMR profiles, and SeqScreen virulence annotations.

2. Hybrid Predictive Modeling

At the core of EPICURUS AI is a machine-learning pipeline that combines multiple modeling approaches with bidirectional inference capabilities:
Feature Engineering: The system extracts features from epidemiological, molecular, and ecological data: 1) statistical descriptors (mean, standard deviation, variance), 2) trend indicators (momentum, exponential moving averages), 3) pattern metrics (seasonal ratios, year-over-year changes), 4) distribution characteristics (volatility profiles), 5) molecular descriptors (genome size, GC content, structural traits), 6) categorical encodings (transmission type, genome classification, AMR priority).
Ensemble Learning: A diverse set of algorithms is employed: 1) tree-based methods (Random Forest, XGBoost, LightGBM), 2) neural architectures (LSTM, BiLSTM for temporal patterns), 3) Gradient Boosting machines, 4) linear models (Logistic Regression, Linear Regression).
Bidirectional Prediction: The platform supports two inference directions. Forward prediction: users input epidemiological parameters (R_0, incubation, CFR, transmission, zoonotic status, vaccine availability) and receive predicted pathogen type, genome classification, AMR priority and resistance percentages, and virulence factors. Reverse prediction: users input molecular and ecological features (genome size, GC content, structural traits, pathogenic factors from SeqScreen) and receive predicted R_0, incubation period, CFR, and environmental stability.
Performance Optimization: Models are validated using time-series cross-validation and ranked by R^2 scores to ensure forecasting accuracy. Features are selected based on SHAP importance scores.

3. Interactive Epidemiological Simulation
The platform incorporates multiple compartmental models, allowing users to simulate outbreak scenarios in real-time:
Mathematical Foundation: The system implements SIR, SEIR, and SIS models through systems of ordinary differential equations.
Parameter Analysis: Users can adjust critical parameters including infection rate (β), recovery rate (γ), and initial population values. The system calculates key metrics such as Basic Reproduction Number ($R_0 = \beta/\gamma$).

4. Intelligent Outbreak Forecasting
EPICURUS AI provides predictive capabilities for disease outbreaks:
Temporal Forecasting: LSTM and BiLSTM models capture long-range dependencies and seasonal patterns in historical case data spanning 1980–2024.
Similarity Analysis: Identifies nearest known pathogen matches using combined confidence scores across epidemiological, molecular, and ecological feature spaces.
Risk Assessment: Generates outbreak risk classifications based on predicted case trajectories, statistical confidence intervals, and molecular virulence signatures.
Novel Pathogen Assessment: For pathogens without historical data (including AI-designed organisms), the system predicts epidemiological behavior from molecular and ecological features alone.

5. WHO GLASS Data Integration
A dedicated lookup interface provides interactive access to official WHO GLASS antimicrobial resistance data: 1) country-level surveillance indicators (94 enrolled, 78 reporting countries/territories/areas), 2) regional breakdowns across all six WHO regions, 3) AMR trend

graphs for critical priority pathogens, 4) SDG 3.d.2 indicators for bloodstream infections, 5) median resistance percentages for priority bacteria.

6. Data Processing and Robustness

The platform is supported by a comprehensive data processing pipeline:
Automated Data Cleaning: Handles missing values, reporting inconsistencies, and data normalization.
Quality Validation: Implements minimum data requirements, outlier detection, and cross-source verification against WHO official statistics.
Error Handling: Robust exception management throughout the data lifecycle.
Real-time Processing: Supports streaming data integration for continuous model updating.
7. Analytical Outputs
EPICURUS AI delivers actionable insights through:
Visual Analytics: Interactive dashboards showing disease trends, geographic distributions, model predictions, and AMR trends.
Comparative Metrics: Side-by-side analysis of disease patterns across pathogens, countries, and time periods.
Forecast Reports: 30-day and 90-day outbreak projections with confidence intervals.
Confidence Scoring: All predictions include uncertainty estimates, enabling risk-calibrated decision-making.
Export Capabilities: Complete results export in multiple formats for further analysis.
8. Strategic Advantage
EPICURUS AI represents a paradigm shift in epidemiological intelligence by:
Bridging Traditional and AI Methods: Combining compartmental models with machine learning ensembles and bidirectional inference.

Integrating Three Domains: Unifying epidemiology, molecular biology, and vector ecology in a single analytical framework.

Enabling Proactive Response: Providing early warning signals for emerging outbreaks through bidirectional prediction — assessing threats from molecular features before cases appear.
Supporting Resource Allocation: Informing public health decisions through data-driven forecasts and risk assessments grounded in WHO statistics.

Addressing Modern Biosecurity Threats: Incorporating capabilities to analyze AI-generated pathogen risks by predicting epidemiological behavior from molecular and ecological features alone — closing the gap between pathogen design and epidemiological assessment.
Building on Authoritative Data: All predictions and lookups are grounded in WHO GLASS statistics, WHO BPPL classifications, SeqScreen functional annotations, and peer-reviewed molecular trait databases.

In our work, we utilized data provided by organizations such as the UN and the WHO. We employed population characteristics from the period 1980 to 2024, including demographic parameters and per capita GDP, as well as parameters of the epidemiological process, such as

long-term dynamics of the epidemiological process across various countries. The expanded dataset now includes molecular traits from NCBI Virus, vector ecology from global arbovirus surveys, and AMR surveillance data from the WHO GLASS dashboard. Through the AIxBio Hackathon, the system evolved from a disease forecasting tool into a comprehensive biosecurity platform capable of bidirectional pathogen prediction — bridging the critical gap between molecular design and epidemiological impact assessment.

Conclusions

«The ultimate goal of a blissful life is bodily health and mental tranquility.»

-Epicurus

The findings of this study demonstrate that global infectious disease dynamics exhibit pronounced temporal complexity, geographic heterogeneity, and disease-specific structural patterns. Extensive visual analysis confirms that outbreak trajectories are shaped by long-term cycles, abrupt transitions, and regionally concentrated disease burdens. While illnesses such as measles and pertussis follow identifiable oscillatory patterns, others display sporadic outbreaks that challenge traditional linear models. The moderate year-to-year correlations and diverse distributional characteristics underscore the necessity for modeling approaches capable of handling non-stationarity, non-linearity, and variable data density across different diseases and countries.

A comparative evaluation of forecasting models reveals clear performance distinctions. Although sequence-based deep learning models, particularly the BiLSTM, effectively capture temporal dependencies and cyclical epidemic behavior, the advanced ensemble method consistently demonstrates superior predictive accuracy and robustness. This is attributable to its incorporation of engineered epidemiological features combined with gradient-boosting and tree-based methods, which excel at modeling irregularities, volatility, and non-linear interactions that neural architectures may not fully capture. The ensemble's stability across

diverse diseases and regions suggests that hybrid modeling approaches offer tangible advantages for real-world epidemiological forecasting.
The Epicurus AI forecasts for 2025 align with historical epidemiological patterns, projecting a higher incidence for diseases with established global burdens while maintaining coherence with regions historically prone to outbreaks. The derived outbreak risk classifications provide actionable insights by highlighting areas where predicted increases deviate significantly from recent trends. This capability is invaluable for resource planning, vaccination strategy optimization, and public health surveillance.

Looking Ahead

Looking ahead, EPICURUS AI is positioned to address the emerging biosecurity threats fueled by artificial intelligence. As discussed throughout this work, contemporary AI models like Evo 2 and RFdiffusion possess the potential to rapidly generate vast numbers of novel pathogens with enhanced virulence. This creates a dangerous asymmetry, where traditional vaccine and drug development systems, burdened by lengthy clinical trial cycles and logistical complexities, may be outpaced by the speed of threat generation.

The bidirectional prediction capabilities developed during the AIxBio Hackathon represent a first step toward closing this gap. By training the system on data from localized outbreaks — integrating molecular features, vector ecology, and epidemiological dynamics — and implementing real-time automated data collection from WHO GLASS, NCBI Virus, and global arbovirus surveys, EPICURUS AI will enable rapid identification and prediction of shifts in both pathogen and host population determinants as an epidemic unfolds.

By integrating the mathematical modeling capabilities of EPICURUS AI with microbiological and molecular methodologies — including SeqScreen functional annotations, Baltimore genome classification, and vector-host ecology — we can leverage AI to develop complex, adaptive models in real-time. The system's ability to predict epidemiological behavior from molecular features alone is particularly critical for assessing AI-generated pathogens before they circulate. Systematizing this information across the three domains — epidemiology, molecular biology, and ecology — will allow us to construct a proactive shield, bolstering global health resilience against both natural and engineered threats.

Future development will focus on several key areas: 1) incorporating mobility data and environmental factors to refine transmission predictions, 2) expanding the curated pathogen database beyond the current 37 priority organisms, 3) integrating real-time health surveillance streams from national and international sources, 4) implementing probabilistic uncertainty modeling with confidence intervals for all predictions, 5) developing nearest-neighbor pathogen matching to identify similarities between novel sequences and known threats, and 6) building automated data pipelines to continuously update molecular, ecological, and AMR datasets.
We envision the EPICURUS AI system being integrated into the operational frameworks of global epidemiological services, such as the WHO, and national centers for disease control and prevention. The platform's WHO GLASS data lookup already demonstrates this integration potential, providing interactive access to official AMR surveillance statistics across 94 enrolled countries. Only through a united front — combining the efforts of epidemiologists, data

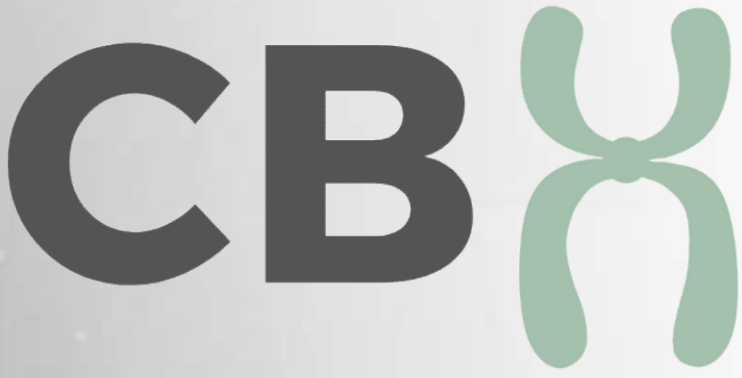
scientists, biologists, and policymakers in a coordinated international response — can we effectively counter the challenges of the 21st century, including the risks of AI-enabled bioterrorism.

Acknowledgements
The authors would like to thank Apart Research, Cambridge Biosecurity Hub and Blue Dot Impact for organizing and hosting the Defensive Acceleration Hackathon (Apr 24 - 26), where this work was conducted.

References
1. Exploring the COVID-19 pandemic's influence on eating disorders: insights from an Italian residential center* // Front. Psychol. – 2025. – Vol. 16. – URL: https://doi.org/10.3389/fpsyg.2025.1522599
2. Pieh C, Budimir S, Probst T. The effect of age, gender, income, work, and physical activity on mental health during coronavirus disease (COVID-19) lockdown in Austria. J Psychosom Res. 2020 Sep;136:110186. doi: 10.1016/j.jpsychores.2020.110186. Epub 2020 Jul 3. Erratum in: J Psychosom Res. 2020 Dec;139:110278. doi: 10.1016/j.jpsychores.2020.110278. PMID: 32682159; PMCID: PMC7832650.
3. The Importance of Childhood for Adult Health and Development—Study Protocol of the Zurich Longitudinal Studies / F. M. Wehrle, J. Caflisch, D. A. Eichelberger [et al.] // Frontiers in Human Neuroscience. – 2020. – Vol. 14. – Article 612453. – DOI: 10.3389/fnhum.2020.612453.
4. Tretina, K. (2025, February 19). *Understanding the Language of Life's Biomolecules Across Evolution at a New Scale with Evo 2*. NVIDIA Technical Blog.
URL: https://developer.nvidia.com/blog/understanding-the-language-of-lifes-biomolecules-across-evolution-at-a-new-scale-with-evo-2/
5. Front. Immunol., 14 October 2020, Sec. Vaccines and Molecular Therapeutics, Volume 11 - 2020 | https://doi.org/10.3389/fimmu.2020.517290
6. Callegari A. OpenAI Roadmap: AI as Autonomous Researchers by 2028 // WebProNews. 2025. 29 October. URL: https://www.webpronews.com/openai-roadmap-ai-as-autonomous-researchers-by-2028/
7. Watson, J.L., Juergens, D., Bennett, N.R. et al. De novo design of protein structure and function with RFdiffusion. Nature 620, 1089–1100 (2023). https://doi.org/10.1038/s41586-023-06415-8
8. Liu Y., Feng J., Pan H., Zhang X., Zhang Y. Genetically engineered bacterium: Principles, practices, and prospects // Frontiers in Microbiology. – 2022. – Vol. 13. – Article 997587. – DOI: 10.3389/fmicb.2022.997587.
9. Brownlee, John. "Statistical studies in immunity: the theory of an epidemic." Proceedings of the Royal Society of Edinburgh 26.1 (1906): 484–521.
10. Kermack W. O., McKendrick A. G. A contribution to the mathematical theory of epidemics. Proc. R. Soc. Lond. A 1927 115 700–721.
11. Bartlett M. S. An Introduction to Stochastic Processes, with special reference to methods and applications. Third edition. Cambridge University Press, 1978.
12. Bartlett M. S. Stochastic Population Models in Ecology and Epidemiology. Methuen, 1960.
13. Kendall D. G. Discussion of 'Measles periodicity and community size' by M. S. Bartlett // J. Roy. Stat. Soc., Ser. A. 1957. No. 120. P. 64–76.

14. Elveback L., Ackerman E., Gatewood L., Fox J. Stochastic two-agent epidemic simulation models for a community of families // American Journal of Epidemiology. 1971. No. 93. P. 267–280.
15. Lekone P. E., Finkenstadt B. F. Statistical Inference in a Stochastic Epidemic SEIR Model with Control Intervention: Ebola as a Case Study // Biometrics. 2006. Vol. 62. P. 1170–1177.
16. Du, H., Zhao, Y., Zhao, J. et al. Advancing real-time infectious disease forecasting using large language models. Nat Comput Sci 5, 467–480 (2025). https://doi.org/10.1038/s43588-025-00798-6
17. Ending Pandemics. (2011). Outbreaks Near Me (formerly Flu Near You): A participatory surveillance system for influenza-like illness. Ending Pandemics. Retrieved from: https://endingpandemics.org/projects/flu-near-you/
18. Mendes, V. I. S., Mendes, B. M. F., Moura, R. P., Lourenço, I. M., Oliveira, M. F. A., Ng, K. L., & Pinto, C. S. (2025). Harnessing artificial intelligence for enhanced public health surveillance: a narrative review. Frontiers in Public Health, 13. https://doi.org/10.3389/fpubh.2025.1601151
19. Wong F, de la Fuente-Nunez C, Collins JJ. Leveraging artificial intelligence in the fight against infectious diseases. Science. 2023 Jul 14;381(6654):164-170. doi: 10.1126/science.adh1114. Epub 2023 Jul 13. PMID: 37440620; PMCID: PMC10663167.
20. World Health Organization. (2024). Immunization dashboard: Reported cases of vaccine-preventable diseases (VPDs). World Health Organization. Retrieved [November 21, 2025], from https://immunizationdata.who.int/
21. World Health Organization. (2024). Global Health Observatory (GHO) data: Indicators index. World Health Organization. Retrieved [November 21, 2025], from https://www.who.int/data/gho/data/indicators/indicators-index
22. United Nations, Department of Economic and Social Affairs, Population Division. (n.d.). Home page. Retrieved November 21, 2025, from https://www.un.org/development/desa/pd/
23. Edouard Mathieu, Hannah Ritchie, Lucas Rodés-Guirao, Cameron Appel, Daniel Gavrilov, Charlie Giattino, Joe Hasell, Bobbie Macdonald, Saloni Dattani, Diana Beltekian, Esteban Ortiz-Ospina, and Max Roser (2020) - “Coronavirus (COVID-19) Cases” Published online at OurWorldinData.org. Retrieved from: 'https://ourworldindata.org/covid-cases' [Online Resource]
24. World Health Organization. Global antimicrobial resistance and use surveillance system (GLASS) report 2022. Geneva: WHO; 2022. ISBN 978-92-4-006270-2. Available from: https://www.who.int/publications/i/item/9789240062702
25. World Health Organization. Global Antimicrobial Resistance and Use Surveillance System (GLASS). WHO; 2024. Available from: https://www.who.int/initiatives/glass
26. World Health Organization. WHO Bacterial Priority Pathogens List, 2022. Geneva: WHO; 2022. Available from: https://www.who.int/publications/i/item/9789240062702
27. Balaji A, Kille B, Kappell AD, Godbold GD, Diep M, Elworth RAL, Qian Z, Albin D, Nasko DJ, Shah N, Pop M, Segarra S, Ternus KL, Treangen TJ. SeqScreen: accurate and sensitive functional screening of pathogenic sequences via ensemble learning. Genome Biol. 2022;23:133. doi: 10.1186/s13059-022-02695-x
28. NCBI Virus. National Center for Biotechnology Information. Available from: https://www.ncbi.nlm.nih.gov/labs/virus/
29. Huang Y, Wang S, Liu H, Atoni E, Wang F, Chen W, Li Z, Rodriguez S, Yuan Z, Ming Z, Xia H. A global dataset of sequence, diversity and biosafety recommendation of arbovirus and

arthropod-specific virus. Sci Data. 2023 May 19;10(1):305. doi: 10.1038/s41597-023-02226-8. PMID: 37208388; PMCID: PMC10199016
30. Hu H, Zhao C, Jin M, Chen J, Liu X, Shi H, Guo J, Wang C, Chen Y. Ensemble Learning: Predicting Human Pathogenicity of Hematophagous Arthropod Vector-Borne Viruses. bioRxiv. 2023. doi: 10.1101/2023.12.30.23300660


CBX


BlueDot
Impact


A*PART