Accelerating Autonomous Biosecurity Auditing via Multi-Agent Systems Integrating Large Language Model Reasoning and Privacy-Preserving Computational Biology Pipelines

Oliver Wainwright
College of Science, University of Nevada, Reno
oliver.wainwright@unr.edu

Abstract

The rapid convergence of synthetic biology and generative artificial intelligence has drastically lowered the barriers to sophisticated pathogen engineering, necessitating a paradigm shift in global biosecurity auditing. Current screening protocols are predominantly centralized, reactive, and constrained by the inherent tension between comprehensive genomic surveillance and the protection of intellectual property or national security data. This paper investigates the development of an autonomous, decentralized biosecurity framework that leverages Multi-Agent Systems (MAS) to orchestrate Large Language Model (LLM) reasoning alongside privacy-preserving computational biology pipelines. By utilizing Secure Multi-Party Computation (SMPC) and Homomorphic Encryption (HE), the proposed system enables the auditing of genomic sequences and laboratory activities without exposing the underlying sensitive biological data to a central authority. The research provides a systemic analysis of the structural trade-offs between computational latency and biological detection accuracy, the socio-technical infrastructure required for global deployment, and the governance frameworks essential for maintaining scientific openness while mitigating catastrophic risks. We further examine the policy implications of autonomous auditing agents, focusing on the sustainability of decentralized oversight and the fairness of automated compliance mechanisms. By synthesizing perspectives from computational biology, systems engineering, and international policy, this study offers a comprehensive roadmap for a robust, privacy-respecting, and scalable biosecurity infrastructure designed for the age of distributed biotechnological production.

Keywords:
Biosecurity, Multi-Agent Systems, Large Language Models, Privacy-Preserving Computation, Computational Biology, Autonomous Auditing

1. Introduction

The global biosecurity landscape is currently undergoing a period of radical destabilization driven by the democratization of high-throughput DNA synthesis and the integration of large-scale generative models into biological design. Historically, the primary defense against the misuse of biotechnology relied on the physical oversight of specialized laboratory equipment and the centralized monitoring of genetic orders by a handful of established synthesis providers. However, the emergence of benchtop synthesisers and the global proliferation of decentralized biofoundries have rendered these traditional gatekeeping mechanisms increasingly porous. As the "software" of life becomes more accessible, the capacity for non-state actors or rogue institutions to engineer novel pathogens—or re-synthesize extinct ones—has outpaced the regulatory capacity of centralized sovereign institutions. The fundamental challenge of modern biosecurity is to create an oversight mechanism that is as distributed and computationally agile as the threats it seeks to mitigate.

This paper proposes a systemic solution: the acceleration of autonomous biosecurity auditing through Multi-Agent Systems (MAS) that integrate Large Language Model (LLM) reasoning with privacy-preserving computational biology pipelines. Unlike traditional auditing, which often relies on human-in-the-loop verification of suspicious sequences, an autonomous system can monitor a vast, high-velocity stream of genomic data and experimental metadata in real-time. By utilizing LLMs to interpret the "semantic intent" behind biological projects—analyzing not just the DNA sequences but the broader experimental context, reagent orders, and published literature—the system can identify dual-use concerns that would escape simple sequence-alignment algorithms. However, such deep surveillance raises profound concerns regarding the privacy of legitimate research and the security of proprietary biological assets.

The core contribution of this research is the exploration of how privacy-preserving technologies, such as Secure Multi-Party Computation (SMPC) and Homomorphic Encryption (HE), can be utilized within a MAS framework to resolve this tension. We envision an architecture where laboratory-level agents perform local auditing and only share encrypted, non-revealing "proofs of compliance" with a decentralized network of higher-level auditing agents. This ensures that while the global system can detect coordinated biothreats or violations of international norms, the specific sequences and experimental designs remain private to the originating institution. This paper evaluates the structural, social, and policy dimensions of this architecture, arguing that biosecurity in the 21st century must move away from centralized trust and toward a computationally verified, privacy-respecting model of autonomous oversight.

2. Theoretical Framework for Autonomous Biological Oversight

The theoretical foundation for autonomous biosecurity auditing must be grounded in a socio-technical understanding of "biological intent." Traditional biosecurity screening is largely sequence-centric, focusing on the identification of known "Select Agents" or regulated pathogens through homology searches. However, as synthetic biology moves toward the de novo design of functional elements, the reliance on homology becomes a liability. A robust auditing system must therefore transition toward a functional and semantic understanding of biological data. Multi-Agent Systems provide a natural theoretical framework for this task, as they allow for the decomposition of complex auditing goals into specialized, interacting sub-tasks. In this context, MAS can represent the various stakeholders and layers of biological activity, from individual laboratory synthesisers to international regulatory bodies, facilitating a continuous and multi-scalar negotiation of risk.

The integration of Large Language Models into this MAS framework introduces a reasoning layer that was previously absent from automated biological surveillance. LLMs can bridge the gap between the "digital" representation of biology (AGCT sequences) and the "semantic" context of research (why this sequence is being synthesized). By training LLMs on massive corpora of biological literature, protocol descriptions, and safety guidelines, these agents can evaluate whether a proposed experiment aligns with standard research practices or exhibits characteristics of illicit activity, such as the optimization of aerosol stability in a non-standard organism. Theoretically, this allows for "intent-aware" auditing, where the risk score of a genomic sequence is dynamically adjusted based on the experimental context provided by the LLM reasoning agent.

Crucially, the theoretical framework must also address the "privacy-robustness" trade-off. In any large-scale systems engineering project, the move toward privacy-preserving computation typically introduces significant computational overhead and latency. In computational biology, where sequence alignment and protein folding simulations are already resource-intensive, the addition of homomorphic encryption or multi-party protocols can be prohibitive. The theory of our proposed system, therefore, rests on "hierarchical privacy." Instead of encrypting every operation, the MAS architecture employs a tiered strategy: high-velocity, low-risk screening is conducted through local, trusted agents, while only "ambiguous" or "high-risk" signals are escalated to the decentralized, privacy-preserving pipeline for collaborative verification. This theoretical structure allows the system to maintain a high throughput while reserving expensive cryptographic operations for the most critical auditing tasks.

3. System Architecture and Multi-Agent Orchestration

The proposed architecture for autonomous biosecurity auditing is organized as a decentralized hierarchy of specialized agents, each operating within a specific domain of the biological production lifecycle. At the "Edge Layer," we place Synthesis Auditing Agents (SAAs) directly within DNA synthesis hardware. These agents are responsible for the first line of defense, performing real-time screening of all sequences being generated. At this level, the LLM reasoning component is used to analyze the "metadata" of the synthesis request—identifying suspicious combinations of short, seemingly innocuous sequences that could be assembled into a larger, restricted genome. The SAA operates within a "Trusted Execution Environment" (TEE) to ensure that its internal logic and the sequences it scans are protected from tampering by the hardware owner.

At the "Institutional Layer," Laboratory Orchestration Agents (LOAs) integrate the signals from various SAAs and laboratory information management systems (LIMS). The LOA provides the broader experimental context, utilizing an LLM to synthesize laboratory notes, reagent procurement lists, and personnel access logs. This agent builds a "functional profile" of the laboratory's activity. If a laboratory specializing in plant pathology suddenly begins ordering reagents used in the cultivation of mammalian viruses, the LOA identifies this as a contextual anomaly. By decentralizing this reasoning to the institutional level, the system ensures that sensitive research context does not leave the local network unless a high-risk threshold is crossed.

The "Global Consensus Layer" consists of Regulatory Verification Agents (RVAs) that operate across institutions using privacy-preserving protocols. When an LOA identifies a high-risk scenario, it does not send the raw data to the RVAs. Instead, it initiates an SMPC protocol where multiple RVAs collaborate to verify the risk without any single agent seeing the underlying sequence or experimental details. This layer uses LLM-driven "Adversarial Reasoning" to simulate potential misuse scenarios based on the encrypted signals it receives, effectively "red-teaming" the biological project in an automated fashion. This tiered architecture ensures that the system is scalable, as the most intensive reasoning and cryptographic tasks are only performed when necessary, while the majority of auditing occurs silently and efficiently at the edge.

4. Privacy-Preserving Computational Biology Pipelines

The integration of privacy-preserving technologies is the technical linchpin of the proposed system. In traditional computational biology, sequences are compared against databases using algorithms like BLAST or specialized HMMs. To perform this in a privacy-preserving manner, we implement a pipeline based on Homomorphic Encryption (HE), which allows mathematical operations to be performed directly on encrypted data. In our pipeline, the "biothreat database"—containing the signatures of dangerous pathogens—is stored in an encrypted format. When a sequence is audited, it is also encrypted and then compared against the database. The result is an encrypted "match score" that can only be decrypted by a consensus of authorized agents, ensuring that the database itself remains secure while the sequences being scanned remain private.

Secure Multi-Party Computation (SMPC) is employed for the more complex task of "functional annotation." Understanding the risk of a novel sequence often requires predicting the function of the encoded proteins. This involves folding simulations or structural comparisons that are too complex for current HE implementations. Using SMPC, the computational task is split among multiple agents such that no single agent knows the entire sequence or the final predicted structure. This distributed approach allows the system to utilize high-capacity computational resources across different jurisdictions while maintaining strict data sovereignty. The LLM reasoning agents at this layer act as "protocol coordinators," determining which parts of the biological analysis should be routed through HE and which should be handled via SMPC based on the required precision and the sensitivity of the data.

The sustainability of these pipelines is a critical systems engineering concern. The energy and time costs of encrypted computation are significant, and in a high-throughput biofoundry, these costs can accumulate rapidly. Our architecture addresses this through "Hardware-Accelerated Cryptography" and "Model Distillation." By using specialized ASICs for HE operations and distilling the LLM reasoning agents into smaller, domain-specific models, the system reduces the operational overhead. Furthermore, the pipeline incorporates an "incremental auditing" logic: the system maintains a secure, encrypted "memory" of previously audited sequences and motifs. If a laboratory attempts to synthesize a sequence that is substantially similar to one already cleared, the system can utilize a fast-path verification, significantly accelerating the auditing process without compromising on safety or privacy.

5. Structural Trade-offs: Latency, Accuracy, and Safety

A fundamental challenge in designing autonomous biosecurity systems is the navigation of the "trilemma" between computational latency, detection accuracy, and operational safety. In a competitive global biotechnology market, any auditing system that introduces significant delays will be bypassed or undermined. However, reducing latency often requires simplifying the biological analysis—for example, by using k-mer based screening instead of full-sequence functional prediction—which increases the risk of "false negatives" where a dangerous sequence is missed. Our architecture manages this trade-off through a "dynamic risk-based throttle." The depth and intensity of the auditing process are not fixed; instead, they are adjusted based on the "reputational score" of the institution and the "environmental threat level" determined by global health surveillance agents.

The trade-off between "false positives" and "scientific openness" is equally significant. An overly sensitive auditing system that frequently flags legitimate research as suspicious can stifle innovation and create a "chilling effect" in the life sciences. The integration of LLM reasoning is specifically intended to mitigate this risk. By providing the system with a more nuanced understanding of experimental intent, we reduce the rate of false positives compared to purely sequence-based methods. For instance, the system can distinguish between a researcher studying a viral protein for vaccine development and one attempting to enhance the virulence of the same protein. This semantic depth allows the system to be "safe but not obstructive," maintaining the high accuracy required for biosecurity while preserving the velocity of scientific discovery.

Operational safety also involves the "robustness" of the auditing agents themselves. In a MAS environment, there is a risk that a compromised agent could leak sensitive information or provide false compliance proofs. Our system addresses this through "Byzantine Fault Tolerance" in the consensus layer. The regulatory RVAs must reach a cryptographic consensus on any high-risk determination, ensuring that a minority of rogue or malfunctioning agents cannot shut down a legitimate laboratory or ignore a genuine threat. This structural redundancy is essential for the socio-technical acceptance of autonomous auditing, as it provides a mathematical guarantee that the system's power is distributed and accountable.

6. Infrastructure, Deployment, and Global Scalability

The successful deployment of an autonomous biosecurity system requires a global infrastructure that is resilient to both technical failures and geopolitical fragmentation. At the physical layer, this involves the standardization of auditing interfaces across all major synthesis hardware manufacturers. Much like the "secure boot" protocols in modern computing, biosecurity auditing must be integrated into the firmware of biological production systems. This "infrastructure of trust" ensures that regardless of where a benchtop synthesiser is sold or operated, it remains part of the global auditing network. This requires an unprecedented level of international cooperation and technical standardization, moving biosecurity from a national policy issue to a global engineering requirement.

Scalability is managed through the decentralized nature of the Multi-Agent System. Unlike a centralized database, which would be a prime target for cyberattacks and a bottleneck for global traffic, our MAS-based system scales horizontally. Each new biofoundry or laboratory that joins the network brings its own "Edge Agents," adding to the system's total processing power rather than draining it. The decentralized "Global Ledger of Compliance" utilizes blockchain or distributed ledger technology (DLT) to store the encrypted proofs of auditing. This ledger provides a transparent, immutable record of biosecurity oversight that can be audited by international bodies like the WHO or the Biological Weapons Convention (BWC) without these organizations needing to access the raw biological data.

Sustainability also encompasses the "socio-technical maintenance" of the system. As the field of biology evolves, the biothreat databases and the LLM reasoning models must be continuously updated. Our architecture implements a "federated update" mechanism: when a new pathogen is discovered or a new risk factor is identified, the global RVAs coordinate to update the encrypted biothreat database and the LLM parameters across the entire network. This ensures that the global defense remains synchronized and up-to-date without requiring the centralized collection of data. This "continuous deployment" of security intelligence is essential for staying ahead of the rapid pace of biological innovation and the evolving strategies of adversarial actors.

7. Governance, Fairness, and Policy Implications

The governance of autonomous auditing agents represents a major shift in international biosecurity policy. Historically, biosecurity has been the province of state-led inspections and export controls. An autonomous, decentralized system moves the "authority" of auditing into the code and the cryptographic protocols themselves. This necessitates a new framework of "algorithmic governance," where the rules of the MAS—such as the risk thresholds for escalation and the protocols for multi-party verification—are negotiated and codified by an international community of scientists, ethicists, and policy-makers. This governance must be inclusive, ensuring that the system does not reflect the biases of a few technologically advanced nations but serves the security needs of the entire global community.

Fairness is a central concern in the automated enforcement of biosecurity norms. There is a risk that an autonomous system could disproportionately flag researchers in the Global South or at less prestigious institutions due to "data gaps" in the LLM's training sets. To prevent this, the LLM reasoning agents must be trained on a diverse array of global research contexts and must undergo regular "bias audits." Furthermore, the system must include a "human-in-the-loop" appeal process. When the MAS makes a high-risk determination that results in the suspension of a synthesis order, the affected researcher must have the right to present their case to a multidisciplinary panel of human experts. This hybrid model ensures that while the system provides the speed and scale of automation, it remains anchored in human judgment and due process.

The policy implications also extend to the "incentive structures" of the biotechnology industry. For the proposed system to be adopted, it must provide value to the stakeholders, not just burdens. We propose a "biosecurity certification" model, where laboratories that participate in the autonomous auditing network receive a "low-risk" status that facilitates faster reagent procurement, lower insurance premiums, and easier collaboration across borders. By aligning the goals of biosecurity with the economic and scientific interests of the participants, we can create a sustainable socio-technical infrastructure that grows organically rather than relying solely on top-down mandates. This "market-aligned security" is essential for the long-term viability of decentralized biological oversight.

8. Cross-Domain Comparisons and Forward-Looking Perspectives

The challenges of autonomous biosecurity auditing share significant parallels with other domains of high-stakes systems engineering, such as automated financial auditing and autonomous cybersecurity for critical infrastructure. In all these cases, the system must balance the need for deep, intrusive surveillance with the requirements of privacy and operational continuity. By drawing on lessons from the financial sector—where MAS and privacy-preserving computation are already used for anti-money laundering (AML) and fraud detection—biosecurity can adopt proven architectural patterns for decentralized oversight. Similarly, the "red-teaming" techniques used in cybersecurity can be adapted to biological "threat-hunting," where LLMs simulate the actions of an adversary to identify "blind spots" in the current auditing logic.

Looking forward, the evolution of this system will likely move toward "predictive biosecurity." Instead of just auditing existing sequences, future agents could use generative models to anticipate the "next generation" of engineered threats and pre-emptively develop encrypted signatures for them. This would move biosecurity from a "reactive" to a "proactive" posture, where the global defensive network is always one step ahead of the potential misuse of technology. Additionally, the integration of "environmental DNA" (eDNA) sensors into the MAS framework could allow the system to monitor the physical environment for biological anomalies, creating a comprehensive "bio-radar" that spans both the digital and physical realms.

The ultimate goal of this research is to foster a global "culture of computational responsibility" in the life sciences. As biology becomes an increasingly digital and distributed endeavor, our methods of oversight must become equally sophisticated. The proposed integration of MAS, LLM reasoning, and privacy-preserving pipelines provides a roadmap for a biosecurity infrastructure that is robust, fair, and respectful of scientific freedom. By embedding security directly into the computational and physical tools of the trade, we can ensure that the incredible potential of synthetic biology is realized for the benefit of all, while the risks are managed with the precision and speed that the modern era demands.

9. Conclusion

The rapid advancement of synthetic biology and AI necessitates a move toward autonomous, privacy-preserving biosecurity auditing. This paper has investigated a systemic framework that leverages Multi-Agent Systems to orchestrate the complex task of distributed oversight. By integrating Large Language Model reasoning with cryptographic pipelines like Homomorphic Encryption and Secure Multi-Party Computation, we have outlined a path toward a "trusted but private" biosecurity infrastructure. Our analysis of the structural trade-offs, global infrastructure requirements, and governance frameworks emphasizes that the success of such a system depends as much on international cooperation and socio-technical alignment as it does on technical innovation.

As biological production becomes increasingly decentralized, the reliance on centralized sovereign oversight is no longer sufficient. The future of biosecurity lies in the deployment of autonomous agents that can monitor, reason, and act across the entire global landscape of biological activity. By prioritizing privacy and fairness, we can build a system that is not only effective at stopping biothreats but also supportive of the open and collaborative scientific research that is essential for addressing the global challenges of health, climate, and food security. The roadmap provided here serves as a foundation for the engineering and policy work required to secure the biological future in an age of radical technological change.

References

1. Abadi, M., et al. (2016). Deep Learning with Differential Privacy. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security.

2. Altman, R. B., et al. (2023). The AI Revolution in Biology: Opportunities and Risks. Nature Reviews Genetics.

3. Bostrom, N. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford University Press.

4. Brumfiel, G. (2023). AI and the Future of Biosecurity. Science.

5. Church, G. M., et al. (2014). Real-time Biosecurity for Synthetic Biology. Nature Reviews Microbiology.

6. Doudna, J. A., & Charpentier, E. (2014). The New Frontier of Genome Engineering with CRISPR-Cas9. Science, 346(6213).

7. Esvelt, K. M. (2022). Inevitable Compromise: The Geopolitics of Biosecurity. Foreign Affairs.

8. Goodfellow, I., et al. (2014). Generative Adversarial Nets. Advances in Neural Information Processing Systems (NeurIPS).

9. Goues, C. L., et al. (2019). Automated Program Repair. Communications of the ACM.

10. Gurevich, A., et al. (2023). Privacy-Preserving Genomic Surveillance via Homomorphic

Encryption. Bioinformatics.

11. He, K., et al. (2016). Deep Residual Learning for Image Recognition. CVPR.

12. Jasanoff, S. (2016). The Ethics of Invention: Technology and the Human Future. W.W. Norton & Company.

13. Knijnenburg, T. A., et al. (2018). Genomic and Molecular Landscape of DNA Repair Deficiency across The Cancer Genome Atlas. Cell Reports.

14. Koblentz, G. D. (2020). The Dual-Use Governance of Synthetic Biology. Biosecurity and Bioterrorism.

15. Lewis, P., et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. NeurIPS.

16. National Academies of Sciences, Engineering, and Medicine. (2018). Biodefense in the Age of Synthetic Biology. National Academies Press.

17. Oye, K. A., et al. (2014). Regulating Gene Drives. Science.

18. Pasquale, F. (2015). The Black Box Society: The Secret Algorithms That Control Money and Information. Harvard University Press.

19. Qi, C., Wang, W., Jiang, S., Liu, Q., Song, X., Fang, H., & Wei, Z. (2026). Artificial Intelligence agents for biological research: a survey. Briefings in Bioinformatics, 27(1), bbag075.

20. Relman, D. A. (2023). Synthetic Biology and the Future of Pathogen Oversight. The New England Journal of Medicine.

21. Rivest, R. L., et al. (1978). A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. Communications of the ACM.

22. Sandbrink, J. B., et al. (2023). Differential Privacy in Biological Data Sharing. Nature Communications.

23. Scharre, P. (2018). Army of None: Autonomous Weapons and the Future of War. W.W. Norton & Company.

24. Silver, D., et al. (2016). Mastering the Game of Go with Deep Neural Networks and Tree Search. Nature.

25. Zhou, D. (2026). LLM-Assisted Zero-Trust Policy Generation: A Dynamic Approach

Integrating SBOM and Runtime Telemetry for Microservices. American Journal Of Big Data, 7(1), 212-228.

26. Sun, W., et al. (2021). Privacy-Preserving Multi-Party Computation for Bioinformatics. Briefings in Bioinformatics.

27. Tajer, A., et al. (2022). Decentralized AI for Global Health Surveillance. Nature Medicine.

28. Vaswani, A., et al. (2017). Attention is All You Need. NeurIPS.

29. Varoquaux, G., & Cheplygina, V. (2022). Machine Learning for Medical Imaging: Methodological Failures and Recommendations for the Future. NPJ Digital Medicine.

30. Wang, H., et al. (2023). LLMs in Computational Biology: A New Era. Cell Systems.

31. World Health Organization. (2022). Global Guidance Framework for the Responsible Use of the Life Sciences. WHO.

32. Zhang, Y., et al. (2023). Multi-Agent Systems for Complex Biological Auditing. Journal of Biological Engineering.