BioSecure: know-your-customer system for DNA synthesis companies[1]

Phil Palmer
University of Cambridge

Ryan Teo
University of Birmingham

Michal Karlubík
Sensible Biotechnologies

Jamie Harris
Imperial College London

With
Apart Research

Abstract

Artificial intelligence accelerates biological design capabilities whilst lowering barriers to hazardous information, creating biosecurity challenges at the critical interface between digital and physical biology: DNA synthesis. We present BioSecure, a modular customer screening framework built during a 48-hour hackathon to address emerging threats from AI-designed biological sequences. The system integrates government-backed identity verification with liveness detection through Veriff, sanctions and watchlist screening, institutional legitimacy checks against the Research Organization Registry (ROR) and biosafety level databases, publication history verification via ORCID, and sequence screening using the ESM-C 300M protein language model. BioSecure synthesises these signals into cumulative risk assessment rather than binary accept/reject logic, functioning as a decision-support tool that preserves human authority over order fulfilment. The framework addresses a critical gap: whilst sequence screening adoption has improved, customer screening remains patchy and manual. We demonstrate that AI-enabled threats require AI-enabled defences, and that automated customer verification can complement sequence screening to create defence-in-depth against misuse of synthetic biology.

[1] Research conducted at the Defensive Acceleration Hackathon, 2025

Keywords: Biosecurity, DNA synthesis screening, know-your-customer

1. Introduction

The convergence of artificial intelligence and synthetic biology has led to accelerated scientific discovery and breakthroughs in various sectors, including health, agriculture, and energy. However, it has also increased the risks associated with biological research and innovation, including accidental and deliberate releases. Large language models and other generative AI models trained on biological data (e.g. protein or genome language models) have been thought to both lower the barriers to hazardous information (e.g. reverse genetics protocols for creating recombinant infectious diseases) and to raise the ceiling of sophistication of biological designs. Currently, these risks primarily affect the digital space, where *in silico* designs have uncertain validity or viability in the physical world.

The critical "transition layer" between digital DNA sequences of concern and their physical counterparts lies in the process of gene synthesis, which are groups of methods used in synthetic biology to construct and assemble genes from nucleotides *de novo*. This could also be thought of as the ability to *write* DNA. Synthetic DNA is widely used in academic research and commercial applications in the medical, biotech, and agricultural industries as a means to create novel compounds and functions.

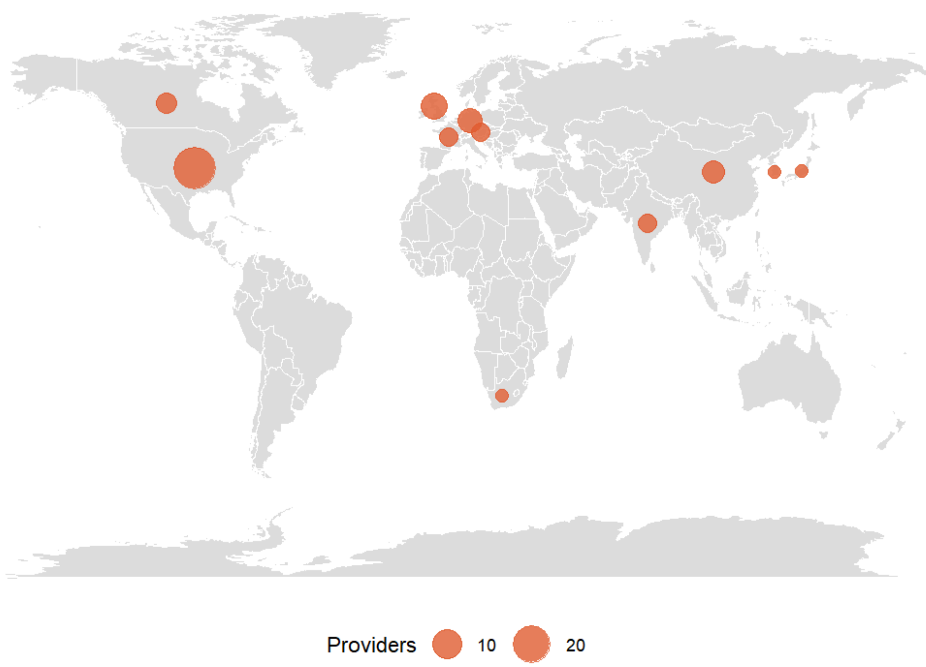

Figure 1 – Global geographic distribution of known synthetic DNA providers (non-exhaustive, [Pálya and Delaney 2022](#))

The global commercial synthetic DNA industry is composed of fewer than 100 providers, ranging from a few multinational corporations and many small-to-medium players offering specialized services. Although many commercial providers (including those in an industry-led consortium of gene synthesis companies) screen orders and customers to block sequences of concern, screening is not mandated by any global or national jurisdiction, and thus screening uptake and quality remains patchy. Furthermore, the increased likelihood of AI-designed sequences being able to evade detection by current screening methods (Wittmann et al., 2025) means that it is all the more important to screen the customers making the orders too.

2. Methods

BioSecure is modular and allows users to incorporate their existing sequence screening tool of choice. To illustrate this, we showed how BioSecure was able to integrate with the biological AI model ESM-C.

To conduct customer screening, BioSecure largely follows best practices and recommendations by industry discussion groups.[2] These involve assessing "legitimacy" at the customer level and the institution they are affiliated with.

The verification framework employs a cumulative risk assessment model rather than a rigid, binary checklist. Individual indicators, such as the absence of specific government identity documentation, are treated as weighted risk signals rather than absolute grounds for denial. Consequently, BioSecure is meant to function strictly as a decision-support tool which synthesises these signals into a comprehensive report and provides a recommendation. This ensures the final authority on order fulfilment remains exclusively with the provider.

At the individual level, BioSecure checks for customer legitimacy with a combination of government ID verification and watchlist screening through Veriff. These indicators help to ensure that there are no red flags associated with the customer's identity. There is also a need to establish that the customer has a legitimate reason to use the synthetic DNA ordered. To this end, checks on an individual's publication history (from their ORCID) are conducted to ensure that they have a history of working on similar use cases.

All customers should be ideally a part of a legitimate institution, which is understood to be an organisation of legal standing, a focus that includes the life sciences, and has sufficient biosafety oversight to work with sequences that are ordered (Alexanian & Carter, 2024). BioSecure checks the listed organisation against trusted registries of organisations, including the Research Organization Registry (ROR), a global, community-led registry of open persistent identifiers for research and funding organizations. A further check on whether an institute has the correct biosafety containment level or relevant certification to handle the organisms identified via sequence screening is done, against established lists of laboratories (Lentzos et al., 2023; Muluneh et al., 2025).

[2] The International Biosecurity and Biosafety Initiative for Science (IBBIS) hosted a Customer Screening Working Group and shared their findings here.

3. Results

We successfully implemented and deployed BioSecure within the 48-hour hackathon timeframe. Figure 2 shows the application interface, which guides users through the multi-stage verification process. The system accepts customer information, processes them through verification layers, and generates comprehensive screening reports.

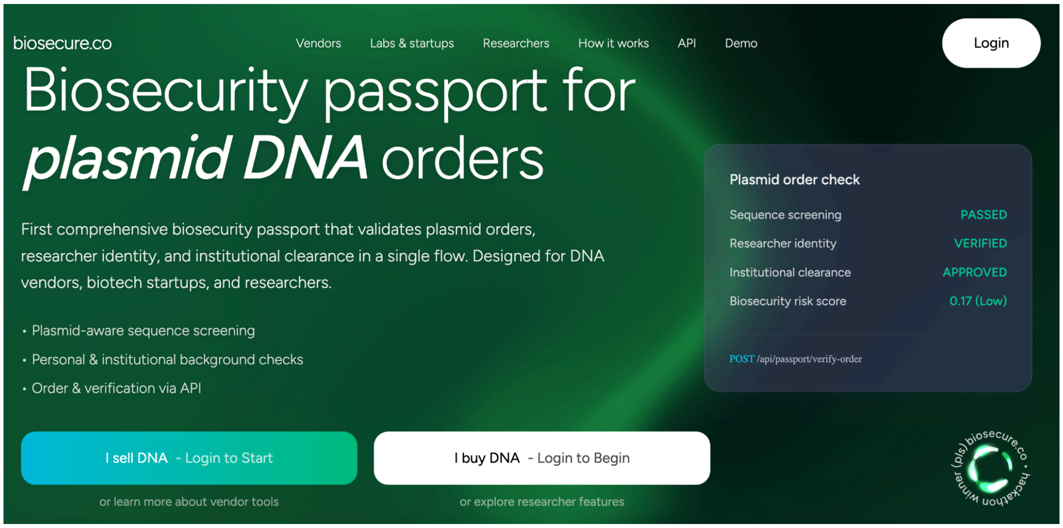

Figure 2 – BioSecure application screenshot

Figure 3 details the complete screening decision tree. Each node represents a verification checkpoint, with colour coding indicating the stage type: blue for identity verification, yellow for legitimacy assessment, and red/green for final decisions. The flowchart demonstrates how BioSecure routes orders through increasingly stringent checks as risk signals accumulate.

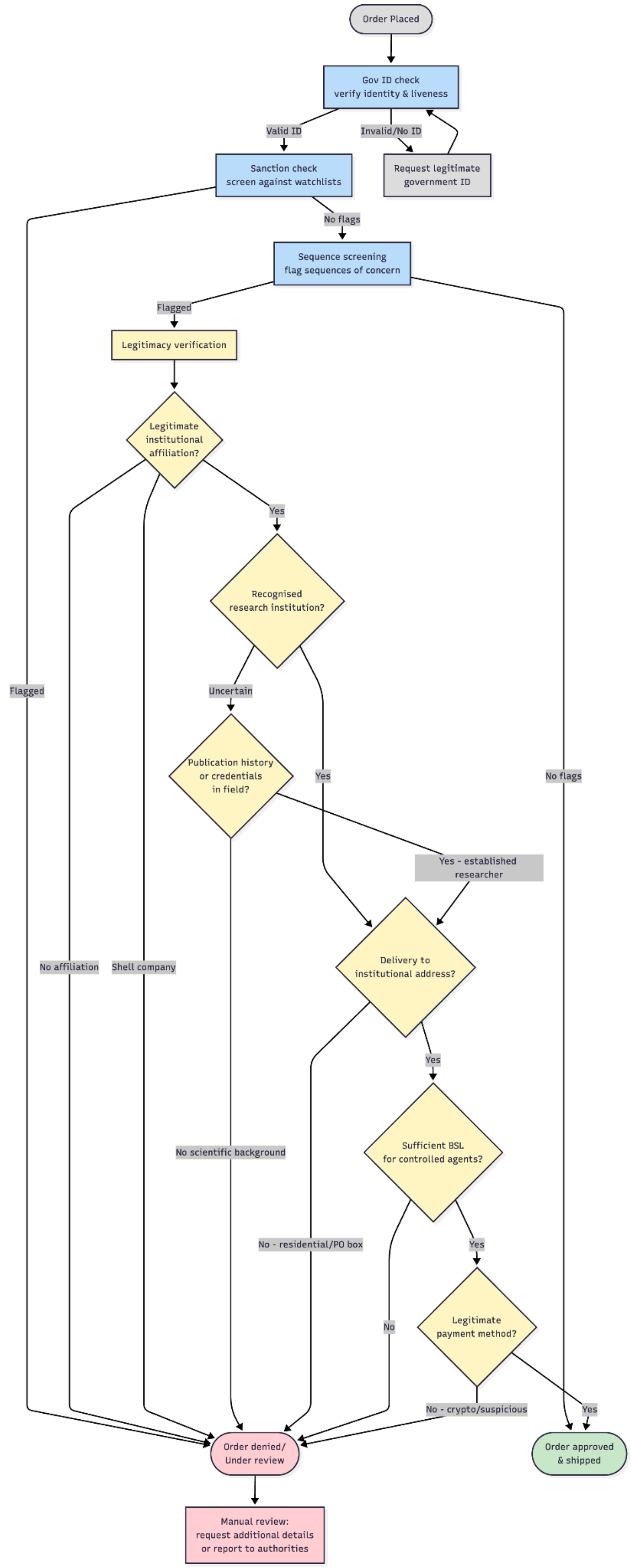

Figure 3 – Flowchart detailing screening decision points and checks

The system processes DNA sequences in 2-3 seconds on standard CPU hardware. Identity verification through Veriff typically completes within 2-5 minutes. Institutional verification completes in under 1 second for most queries.

4. Discussion and Conclusion

BioSecure demonstrates that comprehensive automated customer screening for DNA synthesis becomes feasible with modern identity verification services and AI-enabled tools, addressing the gap between improving sequence screening and rudimentary customer verification.

The compressed development timeline imposed several constraints:

- **User experience:** The interface prioritises functionality over polished user experience. More fundamentally, we did not interview DNA synthesis providers or customers to identify key pain points with existing screening processes
- **Reliability testing:** We could not conduct comprehensive stress testing or validate system behaviour across diverse edge cases
- **Database coverage:** Sequence screening uses only a tiny example dataset from the Common Mechanism as initial proof of concept rather than comprehensive threat databases
- **Provider comparison:** We selected verification providers based on rapid deployment feasibility rather than exhaustively comparing available options for database comprehensiveness and cost

Future work

The critical next step involves interviewing DNA synthesis providers and customers to understand actual pain points in current screening workflows. Which verification steps consume most staff time? Where do customers experience friction? These insights would guide development toward genuinely improving screening rather than simply automating existing processes. The framework applies broadly beyond DNA synthesis to reagent suppliers, cloud laboratories, cell line repositories, live pathogen collections, and AI labs providing access to biological language models or protein design capabilities. Each context requires different verification thresholds, but the core architecture adapts readily across these dual-use biotechnology access points.

Conclusion

AI-enabled threats require AI-enabled defences. As AI lowers barriers to sophisticated biological design, customer screening becomes essential defence-in-depth against sequences designed to evade traditional detection. BioSecure demonstrates that comprehensive automated screening has become technically feasible within remarkably short timeframes. The question is not whether such systems become necessary, but how quickly we deploy them. This work represents a shift from reactive governance to proactive defence, where technical capability development runs ahead of policy mandates. Prevention remains vastly preferable to response.

5. References

Alexanian, T., & Carter, S. R. (2024). *Verifying Legitmacy: Findings from the Customer Screening Working Group, 2020-2023*. https://ibbis.bio/wp-content/uploads/2024/02/IBBIS_Whitepaper_2024_Verifying_Customer_Legitimacy-1.pdf

Lentzos, F., Koblentz, G. D., Ameneiros, M., Earnhardt, B., Houser, R., Rodgers, J., & Wingo, H. (2023). *Global Biolabs Report 2023*. https://www.kcl.ac.uk/warstudies/assets/global-biolabs-report-2023.pdf

Muluneh, A. G., Moa, A., Lim, S., & MacIntyre, C. R. (2025). Mapping biosafety level 3 (BSL-3) and BSL-4 laboratories for public health threats reduction. *Journal of Public Health*. https://doi.org/10.1007/s10389-025-02492-3

Wittmann, B. J., Alexanian, T., Bartling, C., Beal, J., Clore, A., Diggans, J., Flyangolts, K., Gemler, B. T., Mitchell, T., Murphy, S. T., Wheeler, N. E., & Horvitz, E. (2025). Strengthening nucleic acid biosecurity screening against generative protein design tools. *Science*, *390*(6768), 82–87. https://doi.org/10.1126/science.adu8578

6. Appendix

Security considerations

Limitations and future work are listed in the discussion.