
Bio-logical Safeguards

SafeSeq + SafeFind: TEE-Backed Biosecurity Data Rooms and Privacy-Preserving Discovery

Wiktoria Leks
wiki@wikigen.me
Wikigen
Flashbots X

Luc Chartier
lmc7150@rit.edu
Ohio Artificial Intelligence: O(AI)
Flashbots X

With
Apart Research

Abstract

High-resolution biologic and health data are central to early detection of emerging threats and for the operation of resilient health systems. Yet, the same datasets pose privacy and cyber-security risks, when shared in raw form. This tension frequently leads either to over-restriction, where potentially life-saving signals remain locked behind institutional and IP barriers, or to over-exposure, where data circulates without adequate safeguards. We introduce Bio-logical Safeguards, a biosecurity based open sourced protocol system, leveraging trusted execution environments (TEEs) and local large language models (LLMs) to transform high-risk data into actionable lessons, while keeping raw, dual-use and privacy-sensitive material sealed inside confidential compute “safe rooms.” The system comprises two core components. SafeSeq ingests viral, wastewater, and seq-like data into a TEE, computes biosecurity and privacy risk scores, and uses a local LLM to generate policy-filtered narrative reports. Outbound filters and human review prevent leakage of raw sequences, experimental protocols, or re-identifying cohort details. SafeFind is a privacy-preserving discovery layer—a “yellow pages” for datasets, ML pipelines and laboratory capabilities—running in the same TEE, which allows frontline workers and researchers to query for “who might have what we need?”, without logins or exposures of resource-owner identities until both, or all sides, opt in via a double-blind workflow.

Together, SafeSeq and SafeFind instantiate a defensive acceleration approach to biosecurity, leveraging TEEs, defensive AI, and structured human-initiated collaboration to strengthen virus surveillance, genomic privacy and cross-institutional coordination, without IP or data-sharing barriers to undermine pandemic response or personalized discovery. **Keywords:** biosecurity,

1. Introduction

1.1 Background and motivation

Modern biosecurity and healthcare systems face a dual problem:

1. **We need high-resolution surveillance and operational data**
 - Environmental and clinical sequencing (virus genomes, wastewater RNA).
 - RNA-seq and single-cell datasets from patients and tissues.
 - Health-system signals like insulin inventory, ICU capacity, and lab utilization.
2. **The same data is risky to share**
 - Synthetic biology and DNA synthesis have clear **dual-use potential**: long sequences can be misused to design pathogens, toxins, or engineered constructs, motivating calls for robust sequence screening and governance. ([J. Craig Venter Institute](#))
 - Genomic and single-cell expression data can enable **re-identification and private information leakage**, even when nominal identifiers are removed. ([Cell](#))
 - Cyber-biosecurity work has highlighted that genomic datasets and RNA-seq pipelines are attractive targets for attackers, threatening both privacy and data integrity. ([RNA-Seq Blog](#))

The result is a stalemate: valuable data is either **locked away** behind institutional, legal, and IP walls, or **shared too freely**, leaking sensitive details.

At the same time, **discovery is broken** for people **on the front lines**:

- A medical professional cannot easily see whether *anyone* in their region tracks insulin supply or who may use early-warning dashboards for diabetic medications already.
- A wastewater lab may not know which regional center already runs a viral analysis seq pipeline that could help interpret their data, safer.

- A small research group may not know who has a suitable RNA-seq ML pipeline or lab equipment nearby or know who may be able to make use of their data for repurpose and other hypothesis tests.

Everything is gated or recorded behind logins, org charts, and across personal to enterprise networks.

1.2 Def/acc perspective

The defensive acceleration (def/acc) framing asks us to “**strengthen the shield**”: to build technologies that systematically tilt the offense–defense balance toward safety, rather than only slowing a spear. In this project we focus on strengthening the shield around **biosecurity-relevant data and workflows** by:

- using **confidential computing** to protect data *in use* inside hardware-based, attested TEEs, ([Confidential Computing Consortium](#))
- running **local LLMs inside the enclave** to summarize and red-team sensitive analyses
- building a **discovery and coordination layer** that lets people find help quickly without destroying privacy or IP

1.3 Contribution and research questions

We ask:

1. *How can we design a practical genomic surveillance “safe room” that processes sensitive data and ML or AI frameworks without leaking information, weights, raw sequences or re-identifying individuals?*
2. *How can we let frontline users discover who has relevant datasets, tools, and capabilities (e.g. supply chain signals, laboratory protocol insights, ML pipelines) without pre-existing logins or exposing owner identities?*
3. *Can TEEs and confidential computing, already used for privacy-preserving genomic analysis, be turned into a reusable **biosecurity infrastructure pattern** for AI and data-intensive workflows? ([PubMed](#))*

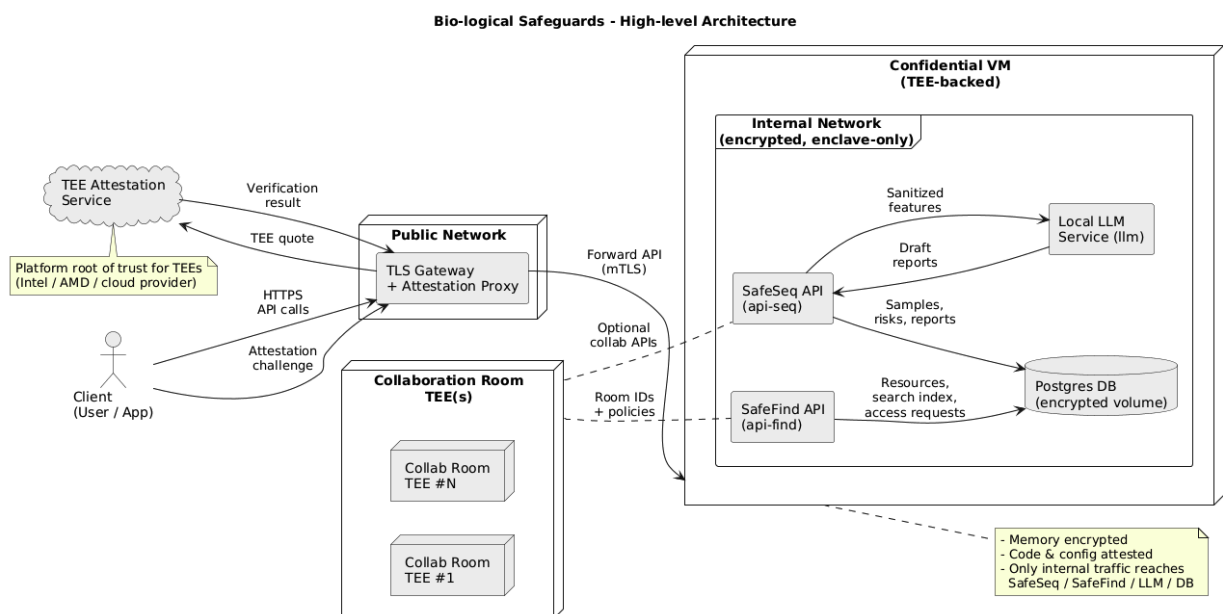
Our hackathon project proposes and partially specifies Bio-logical Safeguards, a system that integrates:

- **SafeSeq** – a TEE-backed genomic and surveillance pipeline.
- **SafeFind** – a TEE-backed discovery and coordination layer.

We describe the architecture, API design, threat-model reasoning, and provide concrete examples that demonstrate how the system could enable safer pandemic risk surveillance and health-data coordination.

2. Methods

2.1 System architecture



We design Bio-logical Safeguards to run inside **TEE-backed confidential VMs** (e.g. Intel TDX / Dstack (<https://github.com/Dstack-TEE/dstack>), where:

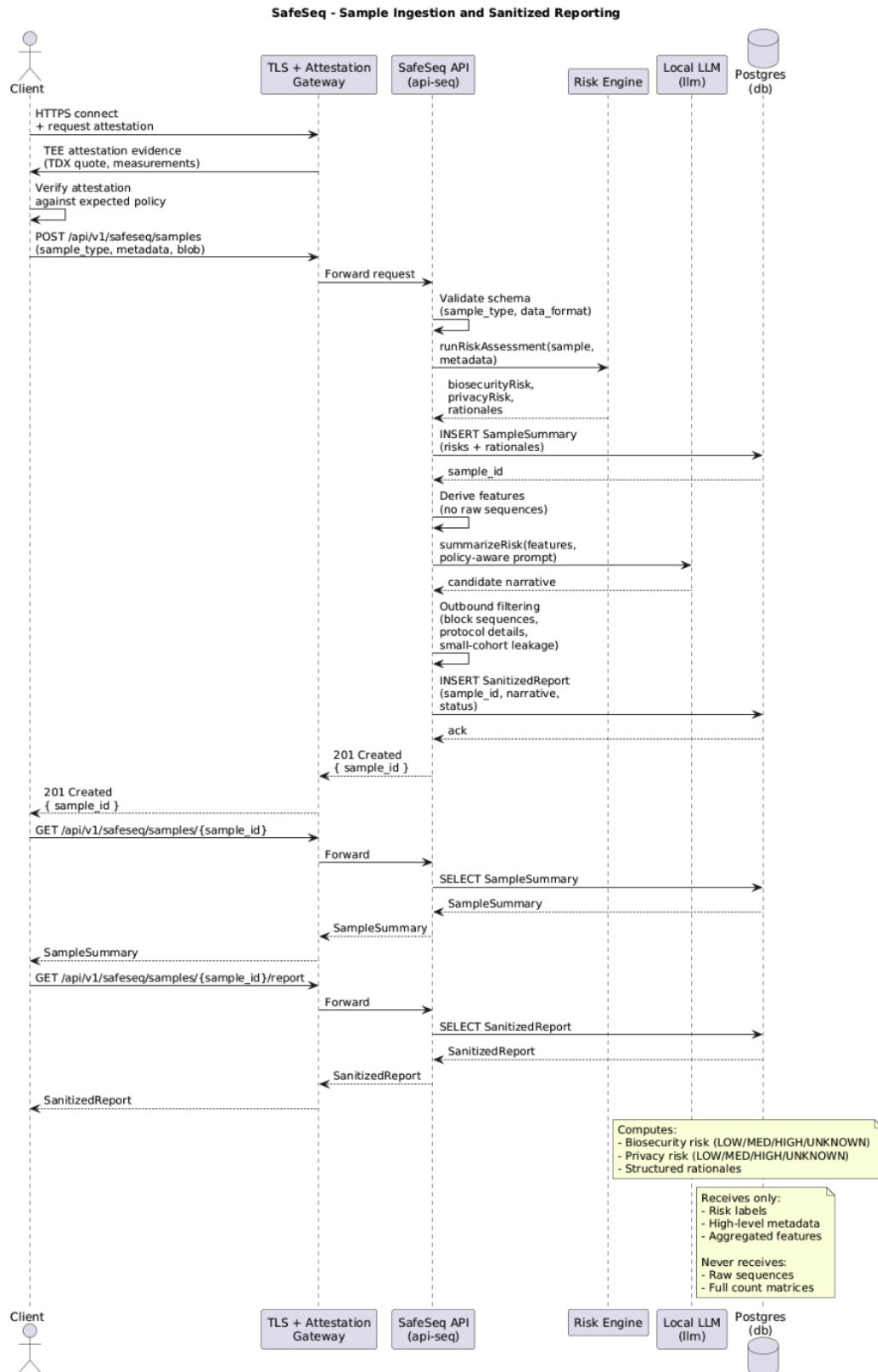
- application code and data are encrypted in memory
- only attested code runs inside the enclave
- remote attestation allows clients to verify that they are talking to the correct deployment with the expected configuration ([Confidential Computing Consortium](#))

Inside each TEE, we run the following Dockerized components:

- **api-seq** – a Python service implementing **SafeSeq**, the genomic / surveillance API.
- **api-find** – a Python service implementing **SafeFind**, the resource discovery API.
- **llm** – a containerized **local LLM** (oss-120b) exposing an OpenAI-style HTTP endpoint *only on the internal network*.
- **db** – a Postgres database for **SampleSummary**, **SanitizedReport**, **ResourceSummary**, and **AccessRequest** tables.

All internal traffic stays inside the TEE’s private network. A gateway outside the TEE exposes a single HTTPS endpoint and attaches **attestation evidence** to the TLS handshake.

2.2 SafeSeq design



SafeSeq defines typed endpoints such as:

- **POST /api/v1/safeseq/samples** – upload a sample with:
 - **sample_type** (virus sequence, bulk RNA-seq, single-cell RNA-seq, etc.),
 - **data_format** (e.g. FASTA, counts matrix, JSON summary),
 - metadata (region, collection time, tissue),
 - opaque payload (kept inside the TEE).
- **GET /api/v1/safeseq/samples/{sample_id}** – retrieve a

SampleSummary:

- **biosecurity_risk** (LOW/MEDIUM/HIGH/UNKNOWN) and structured reasons,
- **privacy_risk** (LOW/MEDIUM/HIGH/UNKNOWN) and reasons,
- current status (**pending_review**, **approved_for_internal**, etc.).
- **GET /api/v1/safeseq/samples/{sample_id}/report** – retrieve a **SanitizedReport**:
 - TEE-generated, policy-filtered narrative suitable for public-health users.
- **POST /api/v1/safeseq/samples/{sample_id}/review** – reviewers can approve or block reports for specific audiences

Risk engine. Inside **api-seq** we implement a risk engine that:

- tags viral and environmental sequences as **dual-use-sensitive**, informed by DNA synthesis and synthetic biology biosecurity guidance. ([J. Craig Venter Institute](#))
- tags human-linked single-cell and small-cohort datasets as **high privacy risk**, reflecting empirical work showing re-identification and private information leakage from single-cell expression matrices. ([ScienceDirect](#))

LLM summarization. For each sample, SafeSeq:

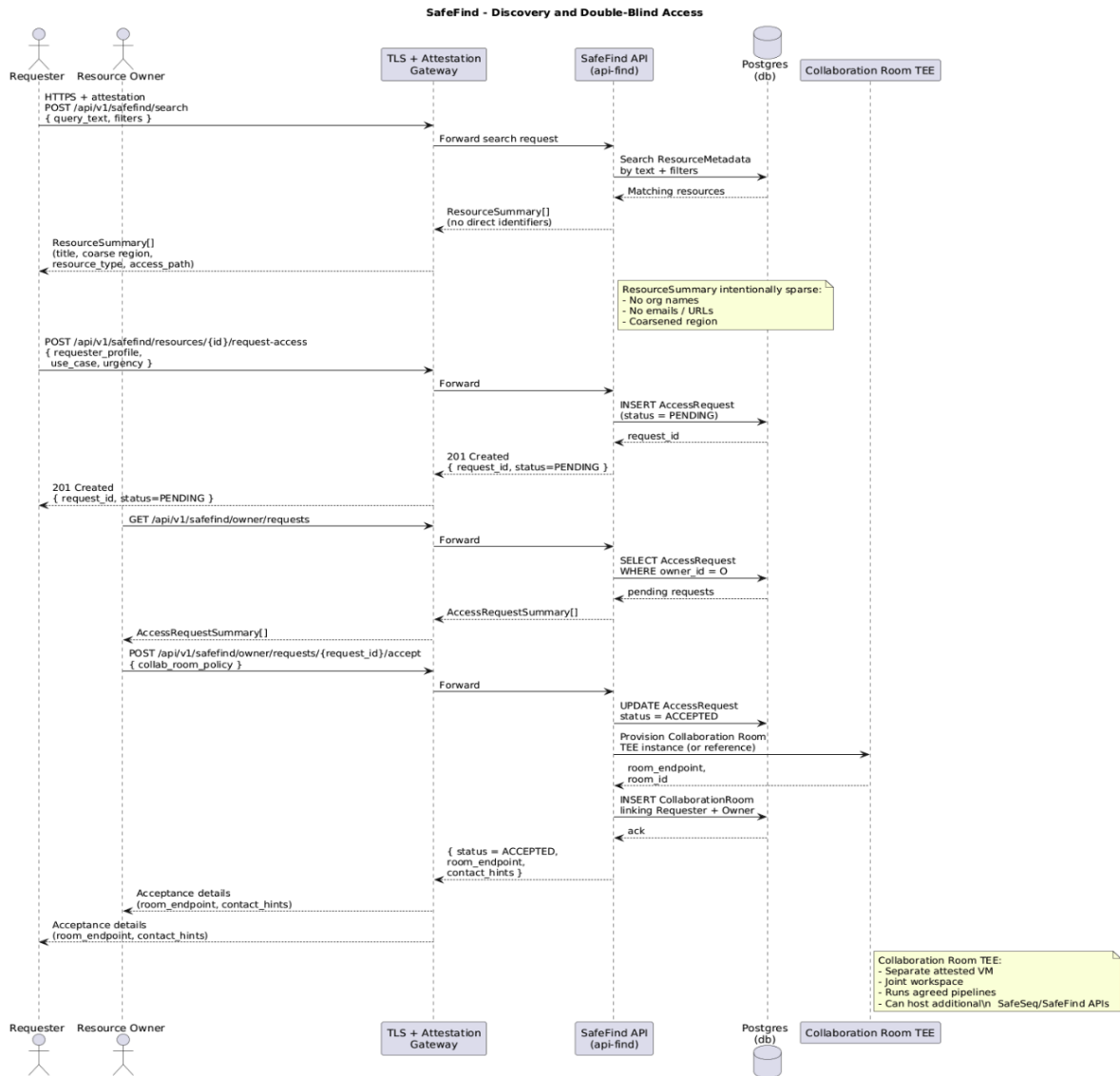
1. Derives sanitized feature vectors (risk labels, high-level metadata, qualitative signals).
2. Calls **llm** internally (e.g. **/v1/chat/completions**) with a prompt that:
 - instructs the model to avoid sequences, protocols, or identities,
 - requests explanations of risk and recommended surveillance/response steps.

Outbound filters. SafeSeq then applies:

- deterministic filters (e.g. regex blocking of `[ACGTU]{N}` strings above a threshold),
- phrase and pattern blocks for lab protocols (“step 1: incubate...”, etc.),
- simple cohort-size sanity checks.

Reports that pass are stored as **SanitizedReports**; others are flagged for manual review.

2.3 SafeFind design



SafeFind maintains a **Resource Registry** inside the TEE, with entries like:

- *Dataset* – “Insulin supply & dispensation trends, ZIP3 441, 2023-2025.”
- *ML pipeline* – “Wastewater viral RNA load analysis pipeline, state-level.”
- *Lab capability* – “BSL-2 lab with short-read sequencer, region X.”
- *Service* – “Omics analysis as a service for small hospitals.”

Each entry stores:

- **ResourceMetadata**: title, description, **resource_type**, topics, **coarsened region** (ZIP3/county/state), and timestamps.
- **ResourceOwnerPublicView**: neutral owner label (“regional health network”, “public-health lab”) and org type.
- **access_path** (open, request-only, or collaboration-room-only) and **contact_model** (mediated/directed).

Search.

POST /api/v1/safefind/search supports low-friction queries with:

- free-text **query_text** (“insulin supply ZIP 44138”),
- optional filters (**region**, topics, resource types).

The response returns a list of **ResourceSummary** objects, each of which **does not include any direct owner identity or raw data details or references**.

Double-blind access requests.

POST /api/v1/safefind/resources/{resource_id}/request-access accepts an **AccessRequestBody** with:

- **requester** profile (role, org type, coarse region, optional affiliation hint),
- use case (“ICU nurse seeing insulin shortages”)
- urgency

SafeFind creates an **AccessRequestSummary** and exposes it only to the resource owner’s admin view. Owners can then:

- accept and assign a collaboration room ID (e.g. a D-stack room inside the TEE), or
- decline, optionally with a high-level reason

2.4 Implementation details

For the hackathon we:

- specify the FastAPI models and endpoint signatures for SafeSeq and SafeFind;
 - sketch a Docker Compose configuration for deploying **api-seq**, **api-find**, **llm**, and **db** into a confidential VM;
 - outline how attestation and key management fit into the workflow;
 - and provide example flows and prompts that could be used for demonstration and future implementation.
-

3. Results

Our “results” at hackathon time are primarily **architectural** and **workflow design** outcomes,

We host a demo of Dstack at vmm.agioh.io

plus example scenarios showing how the system could be used in practice. [Wikigen.me](https://wikigen.me) has accumulated over millions of data sets through its connections, commercial epigenetic biomarkers, metabolomic seq profiles, ancient tibet scrolls, and even Alcor’s dead bodies..

3.1 System design outcome

We produce:

- a unified architecture that combines TEEs, local LLMs, SafeSeq, and SafeFind into a single def/acc-aligned platform;
- a typed API surface for:
 - SafeSeq sample ingestion, risk scoring, reporting, and reviewer actions,
 - SafeFind resource registration, search, and access requests;
- a set of **policy primitives** (biosecurity and privacy filters) and their connection to risk labels and LLM prompt scopes;
- an explicit mapping from these components to hackathon tracks, open sourced to private databases, and propel knowledge graphs (biosecurity, cyber/infra, cross-cutting coordination)

3.2 Example workflows (qualitative “case studies”)

3.2.1 Insulin shortage triage (SafeFind)

- A nurse in ZIP 44138 notices repeated **insulin stockouts**.
- Through a simple web UI backed by SafeFind, they issue a query for insulin supply datasets in their region.
- SafeFind returns two pseudonymous resources:
 - “Dataset · Insulin inventory & dispensation · ZIP3 441xx · Regional health network (request-only).”
 - “ML pipeline · Insulin stockout risk model · State-level public-health team (collaboration-room-only).”
- The nurse sends a double-blind request explaining the situation.
- A regional analytics team accepts, and a safe room is opened; both sides then coordinate redistribution and share aggregate risk forecasts, without any public doxxing of who owns which dataset.

3.2.2 Wastewater viral escalation (SafeSeq + SafeFind)

- A city lab uploads new wastewater viral measurements into SafeSeq.
- SafeSeq flags certain samples as medium-high biosecurity risk, and generates sanitized reports recommending intensified sampling and targeted PCR testing with up to date real world risk factors, taking into account considerations from diverse fields of research.
- A public-health analyst uses SafeFind to discover that another lab in the state has a compatible wastewater pipeline and historical data.
- The analyst submits a request; the second lab accepts; a collaborative safe room is created and a TEE can be implemented.
- Teams compare SafeSeq reports and align on a joint early-warning response, with **no raw sequences ever leaving either enclave**.

3.2.3 Small research hospital with single-cell data (SafeSeq)

- A small hospital runs a single-cell RNA-seq study for oncology patients.
 - They process data entirely inside SafeSeq, which:
 - tags it as **high privacy risk**,
 - prevents any full expression matrices from leaving the enclave,
 - and generates risk explanations citing relevant privacy concerns. ([ScienceDirect](#))
 - This documentation (plus TEE attestation) can be used to support Institutional Review Boards and regulatory review.
-

4. Discussion and Conclusion

4.1 Relation to prior work

Our design builds on three strands of existing work:

1. **Confidential computing and TEEs** – widely explored as a way to protect data-in-use via hardware-based, attested environments. ([Confidential Computing Consortium](#))
2. **Privacy-preserving genomic analysis** – several systems use TEEs or other PETs to run genotype imputation, phasing, and genomic analytics without exposing raw genomes to the service operator. ([PubMed](#))
3. **Genomic privacy and data sanitization** – emerging methods quantify leakage from biologic sequenced material and single-cell data and propose sanitization techniques for reducing identifiability. ([ScienceDirect](#))

SafeSeq adapts these ideas to **biosecurity-oriented surveillance**, focusing on early-warning, dual-use mitigation, and privacy-aware reporting. SafeFind pushes further into **coordination and discovery**, which is under-served: we are not only asking “how do we protect data?” but also “how do we make it *findable* by those who need it without necessarily having to collapse privacy or IP?”

4.2 Limitations

- **Prototype vs. production.**
Our hackathon work specifies architecture and APIs but does not yet provide a fully hardened implementation, comprehensive side-channel mitigations, or large-scale evaluation.
- **Policy complexity.**
The safety of SafeSeq reports depends on the quality of **biosecurity and privacy policies**. Designing and validating robust policies (e.g. around what constitutes actionable sequence information or unacceptable re-identification risk) is non-trivial and will require domain-expert input.
- **Owner incentives.**
SafeFind assumes that organizations will be willing to register at least some metadata into the registry. Incentive design, liability questions, and governance remain open challenges.

4.3 Conclusion and future work

Bio-logical Safeguards demonstrates how **TEE-backed safe rooms, local LLMs, and structured collaboration** can be combined into a **defensive acceleration** system for biosecurity and health data:

- SafeSeq protects and operationalizes virus, wastewater, and omics data.
- SafeFind lets frontline workers and small teams discover who might help them, without premature identity or data exposure.

Future and ongoing work includes:

- implementing and benchmarking the full stack on real confidential-compute hardware
 - adding richer risk models and DNA synthesis-screening integrations, ([NIST](#))
 - formalizing policies and governance around SafeFind participation
 - evaluating SafeSeq+SafeFind in realistic drills with health systems and public-health agencies in collectives
-

5. References

Confidential computing & TEEs

Confidential Computing Consortium. *About – Confidential Computing Consortium*. Retrieved 2025 from confidentialcomputing.io. [Confidential Computing Consortium+1](#)

Google Cloud. *Confidential Computing – Product Overview*. Retrieved 2025 from cloud.google.com/security/products/confidential-computing. [Google Cloud+2Google Cloud Documentation+2](#)

Microsoft Azure. *Azure Confidential Computing – Overview*. Retrieved 2025 from learn.microsoft.com/azure/confidential-computing/overview. [Microsoft Learn](#)

Confidential Computing Consortium. *Defining and Enabling Confidential Computing* (CCC Overview Whitepaper). Linux Foundation, 2023. [Confidential Computing Consortium](#)

Google Cloud. *Confidential VM Overview*. Retrieved 2025 from docs.cloud.google.com/confidential-computing/confidential-vm/docs/confidential-vm-overview. [Google Cloud Documentation](#)

Genomic privacy, sanitization & TEEs for genomics

Walker, C. R., Li, X., Chakravarthy, M., et al. “Private information leakage from single-cell count matrices.” *Cell*, 187(23), 6537-6549.e10, 2024. [Cell+2PubMed+2](#)

Gürsoy, G., Chen, Y., Ramachandran, S., & Gerstein, M. “Data sanitization to reduce private information leakage from functional genomics reads.” *Cell*, 183(4), 905-917.e16, 2020. [PubMed+2ScienceDirect+2](#)

Dokmai, N., Tople, S., Kamm, J., & Shokri, R. “Privacy-preserving genotype imputation in a trusted execution environment.” *Cell Systems*, 12(11), 1004-1017.e6, 2021. [ScienceDirect+2PubMed+2](#)

Dokmai, N., et al. “TX-Phase: Secure phasing of private genomes in a trusted execution environment (TEE).” In *Lecture Notes in Computer Science* (Proc. 2024), 2024. [PubMed+2bioRxiv+2](#)

Bernier, A., Liu, H., & Knoppers, B. M. “Computational tools for genomic data de-identification: Facilitating data protection law compliance.” *Nature Communications*, 12, 6942, 2021. [Nature+2ResearchGate+2](#)

Walker, C. R., et al. “Private information leakage from single-cell count matrices.” Coverage summary in *RNA-Seq Blog*, 2024. [RNA Seq Blog](#)

Gürsoy, G., et al. “Private information leakage from functional genomics data.” *bioRxiv* preprint, 2020. [bioRxiv](#)

Biosecurity, DNA/RNA synthesis & screening frameworks

National Institute of Standards and Technology (NIST). *Biosecurity for Synthetic Nucleic Acid Sequences*. Program overview, 2024. [NIST+1](#)

U.S. Office of Science and Technology Policy (OSTP). *Framework to Encourage Providers of Synthetic Nucleic Acid Sequences to Implement Comprehensive, Scalable Screening* (Nucleic Acid Synthesis Screening Framework), September 2024. [ASPR+1](#)

Center for Health Security, Genetic Engineering and Society Center. *Resources on Nucleic Acid Synthesis Screening*. Johns Hopkins University & partners, 2024. [Gene Synthesis Screening Information Hub](#)

General genomic privacy & de-identification

Bernier, A., Liu, H., & Knoppers, B. M. “Computational tools for genomic data de-identification: Facilitating data protection law compliance.” *Nature Communications*, 12, 6942, 2021. (and 2022 author correction).

Kane, A., et al. “Screening State of Play: The Biosecurity Practices of Synthetic DNA Providers.” 2024.([PMC](#))

JCVI Policy Team. “DNA Synthesis and Biosecurity: Lessons Learned and Options for the Future.”([J. Craig Venter Institute](#))

NIST. “Biosecurity for Synthetic Nucleic Acid Sequences.” 2024.([NIST](#))

NTI | bio. “Preventing the Misuse of DNA Synthesis Technology.”([The Nuclear Threat Initiative](#))

6. Appendix – Security Considerations

1. TEE side-channels and implementation risks.

TEEs reduce many threats but are not a silver bullet; side-channel attacks and implementation bugs remain an active research area. Production deployments should incorporate state-of-the-art side-channel mitigations, patch management, and, where necessary, complementary PETs (e.g. MPC/FHE for particularly sensitive computations).([PubMed](#), tee.fail)

2. Policy bypass and prompt-based leakage.

LLM-based systems can be prompt-hacked into revealing more than intended. Our design partially mitigates this by:

- feeding only sanitized features to the LLM,
- applying deterministic outbound filters,
- and adding human review gates.

However, a full security analysis of LLM-prompt channels and system-prompt hardening is needed.

3. Registry abuse and inference attacks.

Even coarsened metadata in SafeFind could, in principle, be used to infer who owns what (e.g. by cross-referencing with public information).

Defensive measures include:

- rate-limiting and monitoring queries,
- avoiding overly specific topics/regions combinations,
- and providing “bundled” responses at higher levels of aggregation.

4. Governance and misuse scenarios.

A powerful coordination layer could, in a badly governed environment, be abused (e.g. to locate vulnerable facilities). Mitigations include:

- restricting what kinds of resources can be listed,
- vetting participants,
- and including abuse-detection heuristics in query and request logs (stored inside the TEE, with careful access controls).

5. **Humanity factors.**

No technical system can fully compensate for misconfigured policies, social engineering or poor training. Bio-logical Safeguards should be paired with:

- clear operational playbooks
- training for staff on what *not* to upload or approve
- feedback loops with regulators and biosecurity experts

Overall, Bio-logical Safeguards is designed as a **defensive acceleration building block**, not a final answer. We do improve flexible revokability to data access and lock down sensitive data from AI or humans with mal-intent. Our hackathon version focuses on architecting a strong, extensible foundation on which more rigorous security analysis, policy design, and real-world pilots can be built.