

---

# Portable Raman Spectroscopy with AI detection for Dangerous Synthesis of Novel Pathogens<sup>1</sup>

---

Merika  
UCI

Mohamad  
USC

Dominic  
UCI

Harry  
UCI

Grace  
UCI

With  
Apart Research

## Abstract

*Problem: Deadly pathogens synthesized in at homelab made with the help of AI have been a rising concern among national security.*

*Our approach: With this problem in mind, we aim to create a portable hand-held spectroscopy that is able to detect multiple types of peptides and, with the help of AI, detect whether the chemicals are potential precursors to make deadly diseases.*

*Main takeaway: We want to create a hardware + software system with embedded AI. We aim to promote our product to the government, FDA, places with a lot of traction, airport security etc. This would reduce chances of people smuggling in dangerous chemicals (created using open weight models) into and outside of the country/state for experiment.*

## 1. Introduction

*Recent news of D.I.Y laboratories experimenting with deadly diseases have raised concerns of AI's role in biosecurity. Today, thousands of pathogen genomes are publicly available, and modern*

---

<sup>1</sup> Research conducted at the [AIxBio Hackathon](#), April 2026

*synthetic biology allows organisms to be assembled from small DNA fragments rather than requiring access to a full pathogen. DNA synthesis is becoming globally accessible, and experts project it may soon be easier to synthesize a pathogen than to obtain one naturally. Combined with AI tools that lower expertise barriers, this creates multiple pathways where relatively accessible biological components can be combined into potentially dangerous systems. How can we stop this self-synthesis of bioweapons, or improve detection of it? This is where our product comes in.*

*According to US Customs data, imports of hormone and peptide compounds from China reached \$328 million in the first three quarters of 2025, up from \$164 million in the same period the previous year ([The New York Times](#)). Additionally, the global biodefense market is projected to grow to \$32.87 billion by 2034 ([Towards Healthcare](#), [GlobeNewswire](#)). In the midst of an increasing awareness and possession of biosecurity products, there are multiple ways novel pathogens can be synthesized from precursors that are found in daily life. To prevent underground labs from distributing dangerous weaponized pathogens, we want to create a system that detects biohazard precursors. Similar to how pharmaceutical medications track the drug contents of medications, we want to develop a similar system for detecting peptide precursors in samples that could be potential bioweapons.*

***Our main contributions are:***

- 1. A test kit prototype for assessing the risk level of whether a sample is a pathogen, including a 3D model casing.*
- 2. A web dashboard for biosecurity risk assessment with AI detection.*

## **2. Related Work**

*Prior work has focused on preventing dangerous sequences from being synthesized or on using AI to classify known biological samples, whereas our system combines portable Raman sensing and threat assessment to detect and triage potentially biosynthesized hazards from physical precursors in the field. Our work is most similar to prior work that combines AI-based analysis with biosecurity risk assessment, but the biggest gap it addresses is that prior work mostly helps decide whether a sequence should be synthesized or whether an AI tool is risky in general. Our system is designed to assess a real sample quickly and flag an emerging threat even when the sequence is not already in a curated database. This makes it downstream and operational, closer to a deployable biosurveillance or field forensics tool than an upstream, synthesis-screening policy framework.*

*The most relevant prior works include:*

- [De Haro 2024](#) on AI biosecurity risk assessment in synthetic biology. It is the closest direct precursor on the governance side because it provides a structured method for evaluating AI use in SynBio and explicitly frames dual-use and model maturity as part of the risk process.*

- [CDC biological risk assessment guidance](#), because it gives a general hazard/exposure/control framework that fits our alert-and-mitigation workflow.
- Raman + ML pathogen identification studies, such as [Xu et al. 2023](#) and [Li et al. 2025](#), show that Raman spectra can support rapid, culture-free organism classification, which supports our sensing layer. Related clinical Raman work also demonstrates that Raman can be integrated into fast sample-to-result workflows with strong accuracy.
- Synthetic nucleic acid screening guidance and sequence-of-concern efforts such as the [U.S. 2023 framework](#) and later [SBRC-type efforts](#). They define the current prevention baseline on which our work extends.
- [Function-based screening proposals](#), because they highlight the weakness of pure sequence-matching and the need to detect novel hazardous function, which is conceptually similar to our novelty-detection and threat-scoring layer.

### 3. Methods

*We used CAD and SolidWorks to make an MVP model of the embedded software circuitry and device/kit casing for testing the peptide content of a sample. To classify the peptide content percentages and assess the threat risk, we used AI to help generate the program code for the ARDUINO embedded model and the dashboard UI, and 21st.dev to design the background of the UI.*

*The primary detection method we used is portable Raman spectroscopy because each peptide has unique vibrational signature (molecular fingerprinting), it has been used in existing portable devices with \$10K-50K budget (vs \$500K+ for mass spec), no sample prep is needed (works through glass vials), it produces fast results within 5 minutes, and it works for peptides (can detect amino acid sequences and modifications). The threat level classifications are LOW, MEDIUM, HIGH, and CRITICAL, and because of limited budget, only a limited number of metrics are used, namely water (control), insulin (safe), BSA protein (LOW), ricin (bioweapon, CRITICAL), conotoxin (neurotoxin, HIGH), and unknown peptide (MEDIUM).*

*The UI dashboard includes three tabs, including an Analysis tab and Peptide Database tab. The Database contains 300 pseudorecords of peptides for reference. Before submitting a sample for analysis, the user can select the origin of the sample from a dropdown menu and fill in details of the sample. The home screen also allows sample and device selection, the status of analysis, and a summary of recent analyses.*

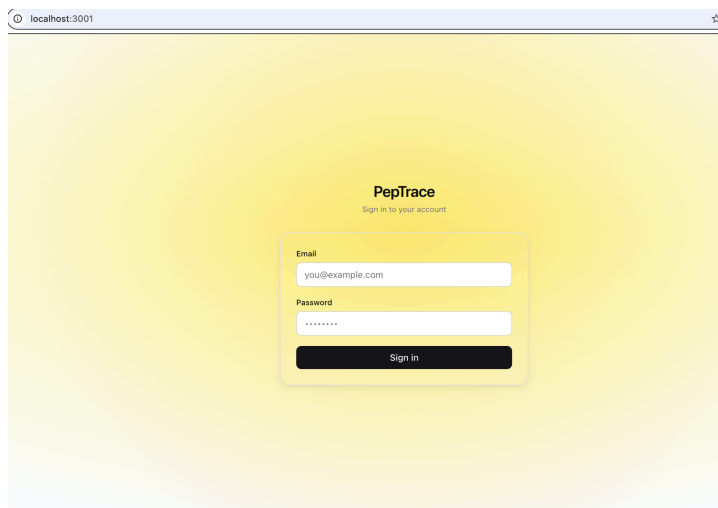
### 4. Results

*In Fig. 4 and 5, the dimensions for Arduino are 2.7x2.1 inches, and for Breadboard are 6.5x2.2 inches.*

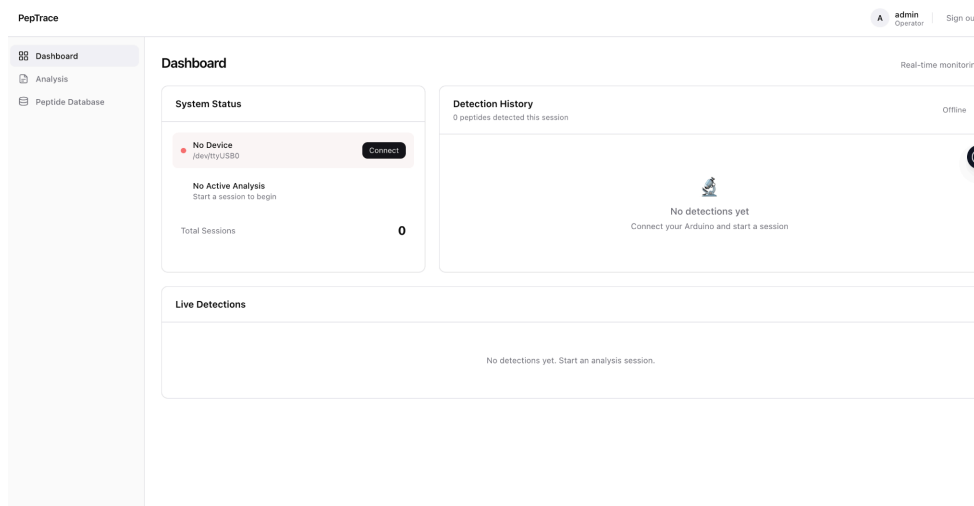
*Our results currently show what a working prototype of our device might produce.*

*We created a layout for a potential device that would be able to distinguish peptides based on their compounds, and a dashboard UI for users to easily see, navigate, and interpret analysis results. Screenshots (Fig. 1, 2, and 3) are included below.*

*After running a test, the dashboard would present device health and connection status, a recent detection summary, and a live recognition chart, and the analysis would present session setup and active controls; a live real-time graph of sensor intensity and temperature; a results panel with peptide names, danger levels, confidence, and match percentages; and active session statistics like start time and elapsed duration.*



*Fig. 1a Login page (test user email admin@example.com and password Password123)*



*Fig. 1b Main UI Dashboard screen page at http://localhost:3001/dashboard*

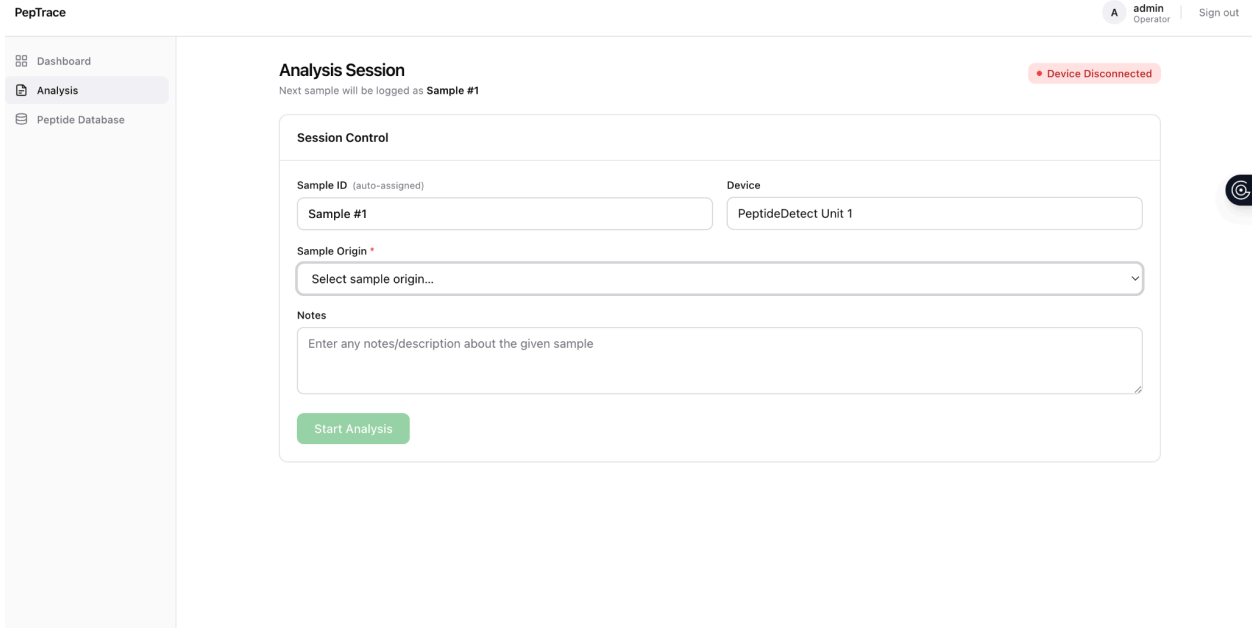


Fig. 2a Analysis screen page at <http://localhost:3001/analysis>

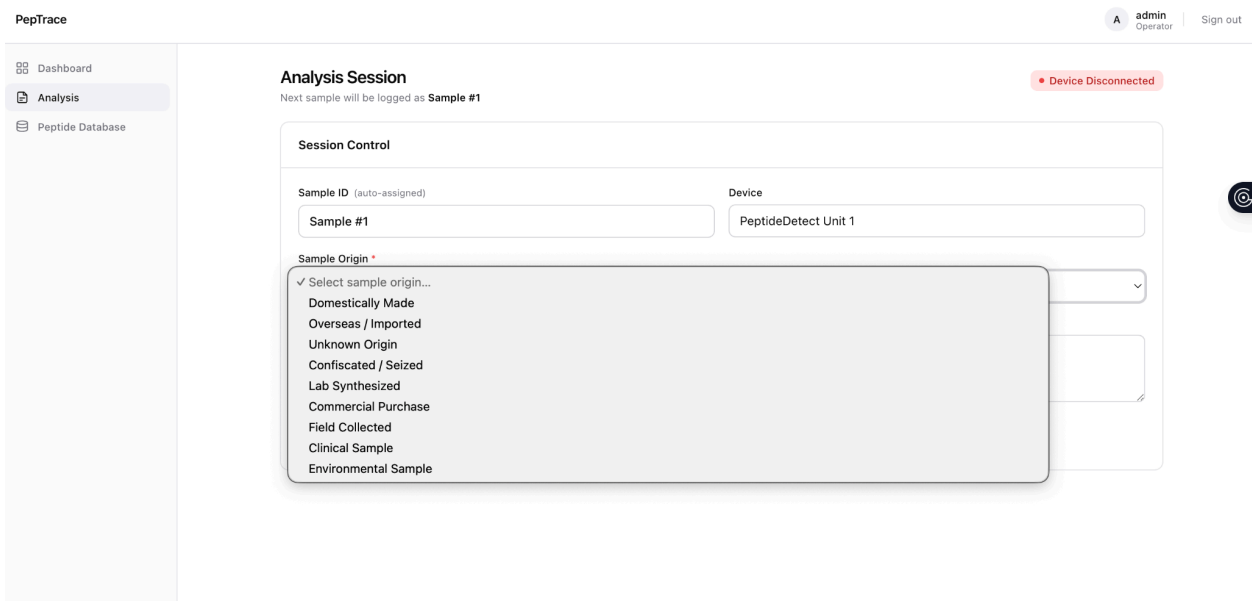


Fig. 2b Analysis screen page with opened dropdown options for Sample Origin field

PepTrace

Peptide Database

Search name, sequence, description...

All danger levels

NAME	SEQUENCE	CLASSIFICATION	MW	DANGER LEVEL
Abeta toxin peptide	EFESGKLVNLLFDK	Toxin	2161.49	Critical
ACTH (1-24)	SYSMEHFRWGKPVGKKRRPKVYYP	Hormone	2933.44	Medium
Ac-YVAD-CMK	YVAD	Inflammatory	628.68	Medium
Adiponectin (globular)	QEQYMKVQLPLTEYIEAIQKCLKLSNELREIQIQRLEFQEK...	Metabolic	5448.17	Safe
Adrenomedullin (1-52)	YRQSPNMFVGLRSGCRFGTCTVOKLAHQIYQFTDKDKDNVAP...	Hormone	5734.47	Safe
Alamethicin	Ac-A1b-Pro-A1b-Ala-A1b-Gln-A1b-Val-A1b-	Antimicrobial	1964.33	High
Alarin	GSPLASLGDPEEDHTFNWGLQREFEQMLMLQLSGG	Neuropeptide	3969.34	Safe
Alpha-MSH	SYSMEHFRWGKPV	Hormone	1664.87	Safe
Alpha-Neuroendorphin	YGGFLRKYPK	Opioid	1318.56	Medium
Anaphylatoxin C3a	SVQLTEKRMVKYKPKELRKCCEQGMRENPMRFSCQRRTRFT...	Inflammatory	5492.28	Medium
Angiotensin (1-7)	RFVTDP	Vasodilator	809.03	Safe
Angiotensin I	RFVTDPNL	Vasodilator	1295.48	Safe
Angiotensin II	RFVTDPF	Vasodilator	1046.18	Medium
Angiotensin III	RFVTDPF	Vasodilator	831.09	Medium
Angiotensin IV	YVDFP	Vasodilator	774.92	Low
Anaxina	SAFA-61a-PRVLA	Metabolic	303.32	Safe
Antimicrobial M9-78	CSPLAKAKYQKATKLLKPK	Antimicrobial	2748.90	Medium
Apelin	ENKQVPTALCARRQDK	Toxin	1964.37	High
Apelin-13	QPFLLSRKPPF	Vasodilator	1000.86	Safe
Arenicin-3	INCCVYKTRVSGVLRVYRQW	Antimicrobial	2707.38	Medium

300 total, page 1 of 10

Fig. 3a Peptide Database screen page that shows the backend database seeded with 300 peptides at <http://localhost:3001/database>

PepTrace

Peptide Database

Search name, sequence, description...

All danger levels

- safe
- low
- medium
- high
- critical
- unknown

NAME	SEQUENCE	CLASSIFICATION	MW	DANGER LEVEL
ACTH (1-24)	SYSMEHFRWGKPVGKKRRPKVYYP	Hormone	2933.44	Medium
Ac-YVAD-CMK	YVAD	Inflammatory	628.68	Medium
Adiponectin (globular)	QEQYMKVQLPLTEYIEAIQKCLKLSNELREIQIQRLEFQEK...	Metabolic	5448.17	Safe
Adrenomedullin (1-52)	YRQSPNMFVGLRSGCRFGTCTVOKLAHQIYQFTDKDKDNVAP...	Hormone	5734.47	Safe
Alamethicin	Ac-A1b-Pro-A1b-Ala-A1b-Gln-A1b-Val-A1b-	Antimicrobial	1964.33	High
Alarin	GSPLASLGDPEEDHTFNWGLQREFEQMLMLQLSGG	Neuropeptide	3969.34	Safe
Alpha-MSH	SYSMEHFRWGKPV	Hormone	1664.87	Safe
Alpha-Neuroendorphin	YGGFLRKYPK	Opioid	1318.56	Medium
Anaphylatoxin C3a	SVQLTEKRMVKYKPKELRKCCEQGMRENPMRFSCQRRTRFT...	Inflammatory	5492.28	Medium

Fig. 3b Peptide Database screen page with opened dropdown menu for filtering peptides based on threat/danger level

## 5. Discussion and Limitations

Our project demonstrates the promise of AI-enabled biosafety screening as an early-warning layer, while also underscoring a core AI safety reality: capability advances must be matched by equally robust governance, oversight, and adaptive safeguards. Broader AI safety implications include that:

- *AI lowers barriers to dual-use knowledge: As open models and biological datasets become more accessible, safety layers that detect and mitigate risky requests or outputs may become increasingly important.*
- *Guardrails are necessary but incomplete: Screening systems can reduce misuse risk and triage suspicious activity, but no static safeguard will fully address novel threats, adversarial adaptation, or evolving scientific capability.*
- *Defensive AI needs continuous updating: Effective safety requires iterative retraining, red-teaming, and integration of new threat intelligence.*
- *Security is sociotechnical: Technical safeguards must be paired with governance, access controls, auditing, policy frameworks, and responsible disclosure norms.*
- *Transparency and explainability matter: For high-stakes domains like biosecurity, interpretable safety systems are more trustworthy and operationally useful than opaque black-box filters.*
- *Safety-performance tradeoff: Overly restrictive models may hinder beneficial science, while permissive systems may increase misuse risk, highlighting the importance of calibrated, context-aware controls.*

*Our work reflects a broader shift from building only more capable models to building capable and governable systems, where deployment safety, monitoring, and misuse prevention are first-class design goals.*

## **Limitations**

*Limitations of our work include the limited metrics for our device and the pseudo-database for our dashboard UI. Our system estimates risk from sequence/spectral similarity, anomaly signals, and known references, but does not determine actual pathogenicity, toxicity, or synthesis feasibility. Its performance is constrained by the coverage and quality of curated datasets (e.g., known toxins, precursor libraries, reference spectra), so novel, engineered, or out-of-distribution threats may evade detection. Aggressive screening can also overflag benign content and increase review burden while missed detections remain possible, so there is a risk of false positives and negatives, especially for short, modified, or adversarially designed sequences. Additionally, Raman and sequence-only approaches each have blind spots, biological risk often requires broader contextual interpretation, and human oversight is still required for the system. Current results are based on limited datasets, synthetic augmentation, and hackathon-scale evaluation rather than operational deployment, real-world field testing, or regulatory benchmarking.*

## **Future Work**

*Next steps for our work include upgrading our system and increasing the number of metrics used to evaluate the threat of a sample, to hopefully be able to identify more peptides and improve threat*

*classification/assessment accuracy; expanding the peptide database on our dashboard; and running/evaluating our model on live data for more reliable results.*

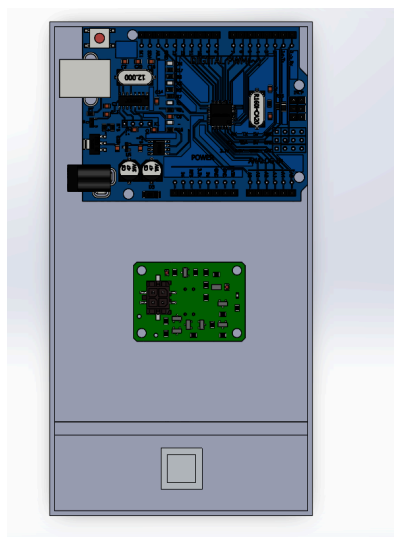
## 6. Conclusion

*With increased levels of globalization and more imports of bio goods, earlier identification and prevention of bioweapons is increasingly critical for biosecurity. Our project will complement current security checks and tests to catch and rank the threat of live substances containing peptides, to assess biosecurity and potential bioweapons, using an embedded system device that runs a model to evaluate peptide percentages' correlation with danger/threat levels. Our project is an MVP prototype of our vision to improve biosecurity by preventing bioweapon's distribution and accessibility.*

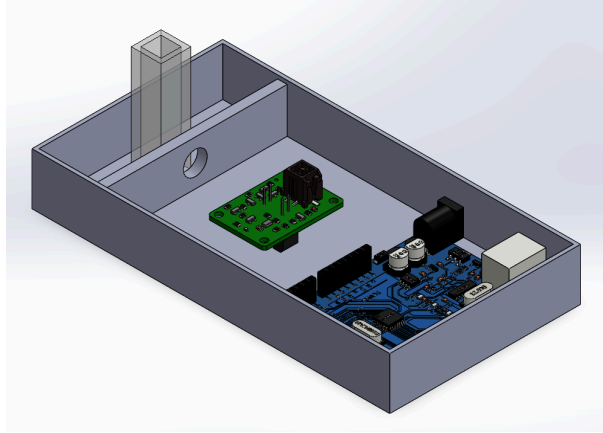
## Code and Data

### **Code repository:**

- *Microcontroller Code*  
[\[https://github.com/MohamadAbakar/Biosecurity-Hackathon/blob/main/Software/Microcontroller Firmware for Electrochemical Peptide Assay\]](https://github.com/MohamadAbakar/Biosecurity-Hackathon/blob/main/Software/Microcontroller%20Firmware%20for%20Electrochemical%20Peptide%20Assay)
- *Breadboard Design*  
[\[https://github.com/MohamadAbakar/Biosecurity-Hackathon/blob/main/Hardware/Breadboard%20Design.png\]](https://github.com/MohamadAbakar/Biosecurity-Hackathon/blob/main/Hardware/Breadboard%20Design.png)
- *Website UI*  
[\[https://github.com/MohamadAbakar/Biosecurity-Hackathon/tree/main/Software/User-Interface/peptide-detection-app\]](https://github.com/MohamadAbakar/Biosecurity-Hackathon/tree/main/Software/User-Interface/peptide-detection-app)
- *CAD Design*



*Fig. 4 Birds-eye view of 3D case*



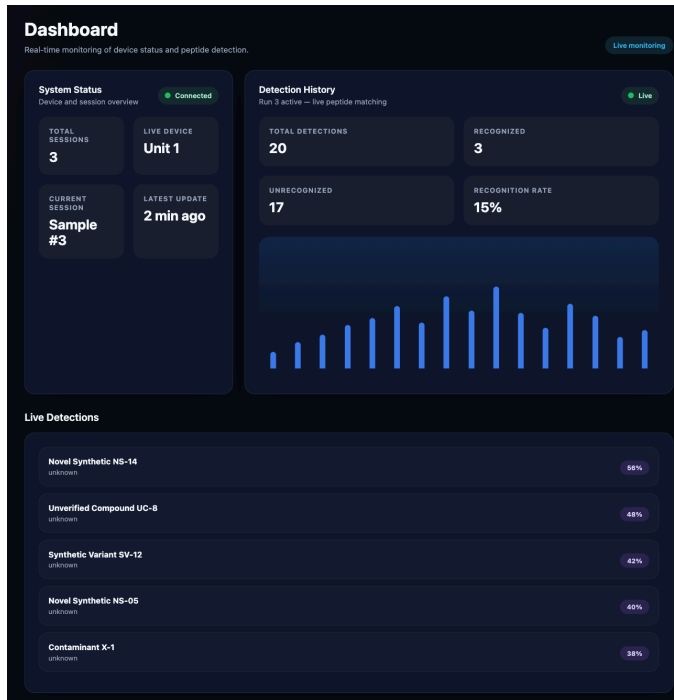
*Fig. 5 Lateral view of 3D case*

- 3 pseudoruns of peptide detection analysis (sample output)
  - Current UI flow from:
    - ``frontend/src/components/dashboard/Dashboard.jsx``
    - ``frontend/src/components/dashboard/LiveAnalysis.jsx``
    - ``frontend/src/components/analysis/AnalysisSession.jsx``
    - ``frontend/src/components/analysis/ResultsPanel.jsx``
    - ``frontend/src/components/analysis/RealTimeGraph.jsx``
  - Each run shows what the dashboard and analysis pages would look like after a live session.
  - ### Run 1 — Domestic sample, mostly safe detections
    - **\*\*Session\*\***
      - Sample: ``Sample #1``
      - Origin: ``Domestically Made``
      - Device: ``PeptideDetect Unit 1``
      - Status: ``Running``
      - Elapsed: ``03:14``
    - **\*\*Dashboard\*\***
      - Top row:
        - Device status: ``Connected``
        - Live detection card: ``15 peptides detected this session``
      - Recognition pie:
        - ``Recognized: 12``
        - ``Unrecognized: 3``
        - ``Recognition rate: 80%``
      - RecentResults list:
        - ``Oxytocin`` — ``safe`` — ``98%``
        - ``Somatostatin-14`` — ``safe`` — ``91%``
        - ``Neuropeptide Y`` — ``safe`` — ``87%``
    - **\*\*Analysis page\*\***

- *Session control card:*
  - *'Next sample will be logged as Sample #1' before start*
  - *Once started: active summary with elapsed timer and Stop button*
- *Real-time chart:*
  - *'intensity' waveform rising with repeated peaks*
  - *'temperature' stable around '36.9-37.3 °C'*
- *ResultsPanel:*
  - *Top detections:*
    - *'Oxytocin' — 'safe' — '98.0% confidence'*
    - *'Somatostatin-14' — 'safe' — '91.3%'*
    - *'Neuropeptide Y' — 'safe' — '87.6%'*
  - *Bars colored green for safe matches*
  - *Stats card preview:*
    - *'15 total scans'*
    - *'12 recognized'*
    - *'3 unrecognized'*
- *### Run 2 — Confiscated / seized sample, mixed danger levels*
  - *\*\*Session\*\**
    - *Sample: 'Sample #2'*
    - *Origin: 'Confiscated / Seized'*
    - *Device: 'PeptideDetect Unit 1'*
    - *Status: 'Running'*
    - *Elapsed: '05:02'*
  - *\*\*Dashboard\*\**
    - *Alerts:*
      - *'Critical peptide detected in current session'*
    - *Live detection card:*
      - *'15 peptides detected this session'*
      - *'Recognized: 7'*
      - *'Unrecognized: 8'*
      - *'Recognition rate: 47%'*
    - *RecentResults list:*
      - *'Sarafotoxin S6b' — 'critical' — '93%'*
      - *'Melittin' — 'high' — '88%'*
      - *'Crotamine' — 'critical' — '84%'*
  - *\*\*Analysis page\*\**
    - *Active session summary:*
      - *'Sample #2'*
      - *'Device Connected'*
      - *'Elapsed: 05:02'*

- *Real-time chart:*
  - *noisy intensity trace with high spikes*
  - *temperature rising slightly from `36.8°C` to `37.4°C`*
- *ResultsPanel:*
  - *`Sarafotoxin S6b` — `critical` — `93.2%`*
  - *`Melittin` — `high` — `88.7%`*
  - *`Crotamine` — `critical` — `84.1%`*
  - *`Unknown Sample UNK-23` — `unknown` — `52.8%`*
- *Status badges:*
  - *critical results show red badge*
  - *high-risk hits show orange/red bar*
- **### Run 3 — Lab-synthesized / unknown origin, novel detections**
  - **\*\*Session\*\***
    - *Sample: `Sample #3`*
    - *Origin: `Lab Synthesized`*
    - *Device: `PeptideDetect Unit 1`*
    - *Status: `Running`*
    - *Elapsed: `02:47`*
  - **\*\*Dashboard\*\***
    - *No high-priority alert banner*
    - *Live detection card:*
      - *`20 peptides detected this session`*
      - *`Recognized: 3`*
      - *`Unrecognized: 17`*
      - *`Recognition rate: 15%`*
    - *RecentResults list:*
      - *`Novel Synthetic NS-14` — `unknown` — `56%`*
      - *`Unverified Compound UC-8` — `unknown` — `48%`*
      - *`Synthetic Variant SV-12` — `unknown` — `42%`*
  - **\*\*Analysis page\*\***
    - *Session control:*
      - *active run with Stop button*
    - *Real-time graph:*
      - *jagged intensity trace suggesting complex sample*
      - *temperature around `37.0°C`*
    - *ResultsPanel:*
      - *`Novel Synthetic NS-14` — `unknown` — `56.0%`*
      - *`Unverified Compound UC-8` — `unknown` — `48.4%`*
      - *`Synthetic Variant SV-12` — `unknown` — `42.1%`*
      - *additional unrecognized results below*
    - *Interpretation:*

- system would show mostly unknown matches
- results panel emphasizes exploratory sample state



*Fig. 6 Mockup of Run 3 results*

## Author Contributions (optional)

*Mohamad made the circuitry diagram and UI dashboard.*

*Merika made the 3D casing in SolidWorks.*

## References

1. Sun, Jasmine. “Chinese Peptides’ Are the Latest Biohacking Trend in the Tech World.” *The New York Times*, 3 Jan. 2026, [www.nytimes.com/2026/01/03/business/chinese-peptides-silicon-valley.html](http://www.nytimes.com/2026/01/03/business/chinese-peptides-silicon-valley.html).
2. Towards Healthcare Research & Consulting. “Biodefense Market Size to Capture USD 35.2 Billion by 2035.” *Towardshealthcare.com*, Towards Healthcare Research & Consulting, 23 Apr. 2026, [www.towardshealthcare.com/insights/biodefense-market-sizing](http://www.towardshealthcare.com/insights/biodefense-market-sizing). Accessed 25 Apr. 2026.
3. Precedence Research. “Biodefense Market to Attain USD 32.87 Billion by 2034, Driven by Biotech Advancements and Government Funding.” *GlobeNewswire News Room*, *GlobeNewswire*, 31 Oct. 2025, [www.globenewswire.com/news-release/2025/10/31/3178499/0/en/Biodefense-Market-to](http://www.globenewswire.com/news-release/2025/10/31/3178499/0/en/Biodefense-Market-to)

- Attain-USD-32-87-Billion-by-2034-Driven-by-Biotech-Advancements-and-Government-Funding.html*. Accessed 25 Apr. 2026.
4. *The Counterfactual*. “Uncensorable, Unmonitorable, Uncontrollable: U.S. Policy Options for Open-Weight Biosecurity Risk.” *Uncensorable.ai*, 2026, *uncensorable.ai/*. Accessed 25 Apr. 2026.
  5. Kim, Sophie. “The Open-Weight Problem.” *Substack.com*, *The Counterfactual*, 18 Mar. 2026, *thecounterfactual.substack.com/p/the-open-weight-problem*. Accessed 25 Apr. 2026.
  6. Alexander, Scott. “Next-Token Predictor Is an AI’s Job, Not Its Species.” *Astralcodexten.com*, *Astral Codex Ten*, 26 Feb. 2026, *www.astralcodexten.com/p/next-token-predictor-is-an-ais-job*. Accessed 25 Apr. 2026.
  7. Alexander, Scott. “What Is Man, That Thou Are Mindful of Him?” *Astralcodexten.com*, *Astral Codex Ten*, 2 Sept. 2025, *www.astralcodexten.com/p/what-is-man-that-thou-art-mindful*. Accessed 25 Apr. 2026.
  8. Casper, Stephen, et al. “Open Technical Problems in Open-Weight AI Model Risk Management.” SSRN, October 26, 2025, <http://dx.doi.org/10.2139/ssrn.5705186>. Accessed 25 Apr. 2026.
  9. Li, Jinjin, et al. “ProjFusNet: Deep Neural Network for Peptide Precursor Prediction Using Projection-Fused Protein Language Model and Structural Features.” *Journal of Cheminformatics*, vol. 17, no. 1, Springer Science and Business Media LLC, Dec. 2025, <https://doi.org/10.1186/s13321-025-01117-x>. Accessed 25 Apr. 2026.
  10. Ollama. “Ollama.” *Ollama.com*, *ollama.com/*.

## LLM Usage Statement

*We used Anthropic Claude, OpenAI ChatGPT, Google Gemini, Perplexity, Copilot, and Cursor AI to brainstorm approaches, find relevant prior work and resources, template our prototype code and program, and help draft sections of this report. All results and claims were independently verified.*