

VaultX

- AI-Driven Middleware for Real-Time PII Detection and Data Security

Diego Sabajo
Affiliation

Egbert Stuger
Affiliation

VaultX Technologies addresses the critical challenge of safeguarding sensitive information in AI systems while ensuring compliance with data privacy laws like GDPR. Our proof-of-concept uses a 3-layered approach to detect and encrypt Personal Identifiable Information (PII) in real time across data sources such as chatbots, email forms, and documents. We also enable seamless decryption for authorized workflows, ensuring sensitive data remains accessible only when needed.

By integrating rules-based detection (regex) Named Entity Recognition (NER) and Language Models, we achieve a balance of accuracy, scalability, and cost-effectiveness. Unlike traditional solutions, VaultX is designed as middleware that seamlessly integrates into existing workflows, significantly reducing the complexity of achieving compliance and security for businesses in data-sensitive sectors such as fintech, healthcare, and customer service.

Commercial viability is further enhanced by VaultX's real-time processing capabilities, which enable businesses to proactively secure sensitive data without disrupting operations. By lowering the risk of data breaches and hefty non-compliance fines, estimated to cost businesses billions annually, VaultX delivers clear ROI. Its lightweight and scalable design make it accessible even to small- and medium-sized enterprises, opening a vast and underserved market segment.

1. Problem overview

The adoption of AI-powered tools has transformed how businesses operate, but it has also dramatically increased the amount of Personally Identifiable Information (PII) being collected, processed, and stored. This has created a dual challenge: complying with complex and evolving data privacy laws (e.g., GDPR, CCPA) while safeguarding sensitive information against a rising tide of cyberattacks.

Data breaches and non-compliance fines have reached staggering levels, costing businesses billions annually. For instance:

- The 2013 data breach at Target Corporation, involving the loss Of 40 million credit and debit card numbers and 70 million other customer details illustrates the high stakes that are involved

<https://www.aiat.report/report/AIAssuranceLandscape/AI-Resilient-IT-Security>

VaultX

- National Public Data (NPD) will probably be 2024's biggest data breach. The mammoth breach potentially impacts 2.9 billion records, close to the most significant data leak ever <https://nordlayer.com/blog/data-breaches-in-2024/>

Protecting PII is challenging due to the limitations of current solutions, which are often reactive, manual, and tailored to specific use cases like email systems. These methods struggle to scale across unstructured data sources, such as chatbot conversations or web forms, and their fragmented nature creates blind spots, leaving businesses exposed to vulnerabilities.

The urgency of this problem has never been greater. With the proliferation of AI-powered tools, companies must adopt proactive solutions to safeguard sensitive information while meeting regulatory demands. The stakes are high—not just in terms of financial penalties, but also in maintaining customer trust and protecting organizational reputation.

2. Your solution

By 2027, **VaultX Technologies** will evolve into a robust middleware platform enabling real-time detection, encryption, and secure management of PII at scale. This solution will seamlessly integrate into a variety of workflows, including chatbots, web forms, email systems, and document processing pipelines, ensuring businesses remain compliant with global data privacy laws while safeguarding sensitive information from misuse or breaches.

a. Technical Foundation:

VaultX combines three powerful technologies:

1. **Rules-Based Detection (Regex):** For structured and well-defined patterns like phone numbers, SSNs, and email addresses, enabling deterministic accuracy for specific PII types.
2. **Named Entity Recognition (NER):** Leveraging SpaCy or similar frameworks to identify broader and more context-dependent PII categories like names, locations, and dates.
3. **Language Models (LLMs):** Fine-tuned models for identifying nuanced and domain-specific PII that regex and NER may miss. This layer also ensures adaptability to industry-specific terminology and emerging patterns.

b. Addressing the AI Safety Challenge:

VaultX tackles the dual challenge of privacy compliance and data security in AI systems. As AI models increasingly interact with sensitive user data (e.g., personalized healthcare assistants, legal advisors), these systems risk exposing private information during training, inference, or post-deployment use. VaultX ensures that AI systems interact with encrypted or anonymized data, mitigating risks of data breaches, model inversion attacks, and non-compliance with data protection regulations. This safeguards users while maintaining AI utility.

c. Competitive Advantages and IP Potential:

VaultX

- **Middleware Integration:** Unlike traditional monolithic tools, VaultX seamlessly integrates into existing infrastructures, reducing adoption friction.
- **Multi-Layered Detection:** Combining regex, NER, and LLMs offers unparalleled detection accuracy and adaptability.
- **Real-Time Processing:** VaultX is designed for real-time workflows, enabling on-the-fly data masking and encryption without latency issues.
- **Scalable Architecture:** Modular design ensures easy scaling to handle large datasets or high-throughput environments.
- **IP Opportunities:** Innovations in encryption-decryption workflows, domain-adapted NER models, and middleware design create opportunities for patents.

d. Technical Diagrams and Proof-of-Concept Results:

We envision diagrams illustrating:

1. Workflow Pipeline:

- a. Data sources (e.g., chatbot, web form, email) → Detection layers (regex, NER, LLMs) → Encryption → Storage.

2. Integration Model:

- a. Middleware interacting with downstream AI/ML systems or existing business applications.

e. Innovation and Scalability:

VaultX's **modular architecture** allows businesses to customize detection models, encryption standards, and integration methods based on their unique needs. Its scalability is supported by parallel processing (e.g., `ThreadPoolExecutor`) and robust key management for multi-tenant environments.

By 2027, VaultX aims to expand beyond PII, addressing sensitive data categories like financial transactions, health records, and intellectual property, making it indispensable across industries.

3. Pilot experiment or demo

The primary goal of the hackathon was to develop a **Minimum Viable Product (MVP)** for VaultX Technologies, showcasing the detection, encryption, and secure management of Personally Identifiable Information (PII).

1. PII Detection Pipeline:

- Developed a multi-layered PII detection pipeline. The pipeline detects and masks PII in various text inputs, such as chatbot messages, email forms, and documents.

2. Encryption Module:

- a. Implemented a secure encryption mechanism using **AES-based Fernet encryption**.
- b. Integrated **key management features**, including key rotation and secure storage, to meet industry standards.

3. Decryption and Re-encryption:

VaultX

- a. Added functionality to decrypt masked PII for authorized use and re-encrypt data during key rotations, ensuring security and long-term manageability.
4. **Integration and Simulation:**
 - a. Built a middleware simulation that processes incoming data streams in real-time, demonstrating compatibility with chatbots and web forms.

Limitations and Future Developments

Balancing detection accuracy and processing speed was a significant challenge, addressed by combining regex for simple patterns and NER for contextual detections. Managing false positives, especially for names and ambiguous entities, required implementing a placeholder mechanism to avoid overlaps between regex and NER results. However, the prototype is not yet production-ready, lacks scalability testing, and struggles with detecting some complex PII patterns.

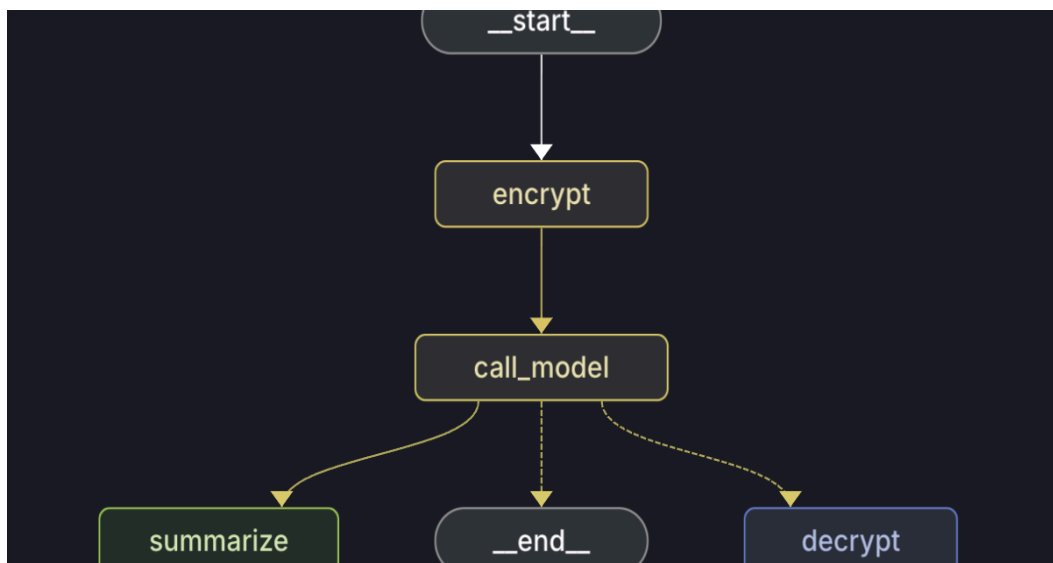


Figure 1 – Flow of Vault X using a chatbot

VaultX

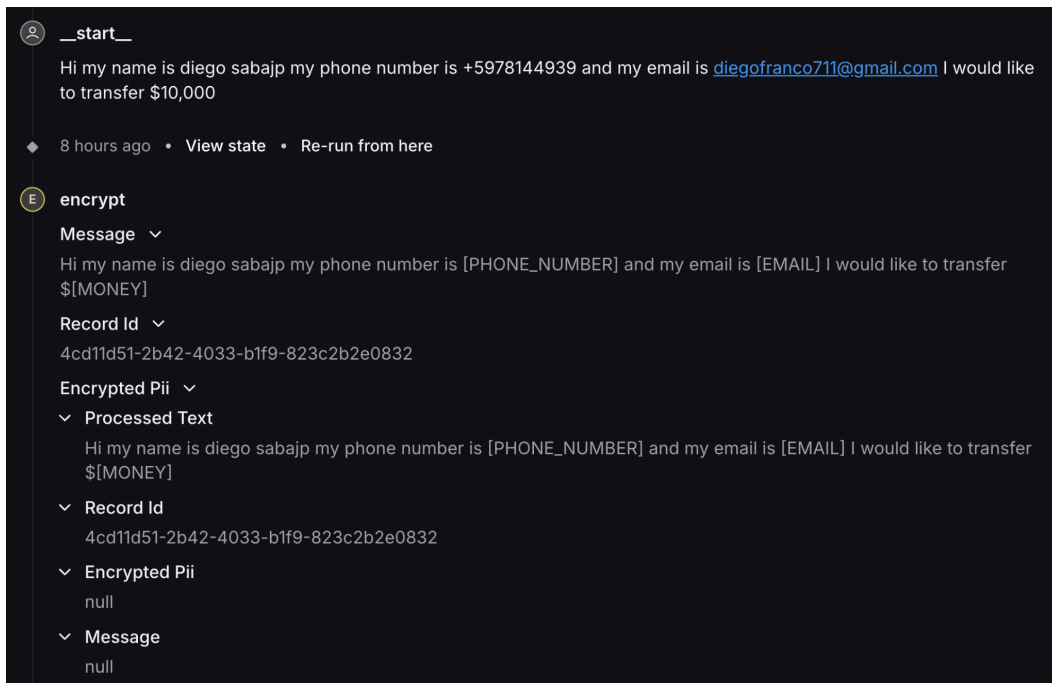


Figure 2 – Real Time Encryption, PII doesn't reach LLM

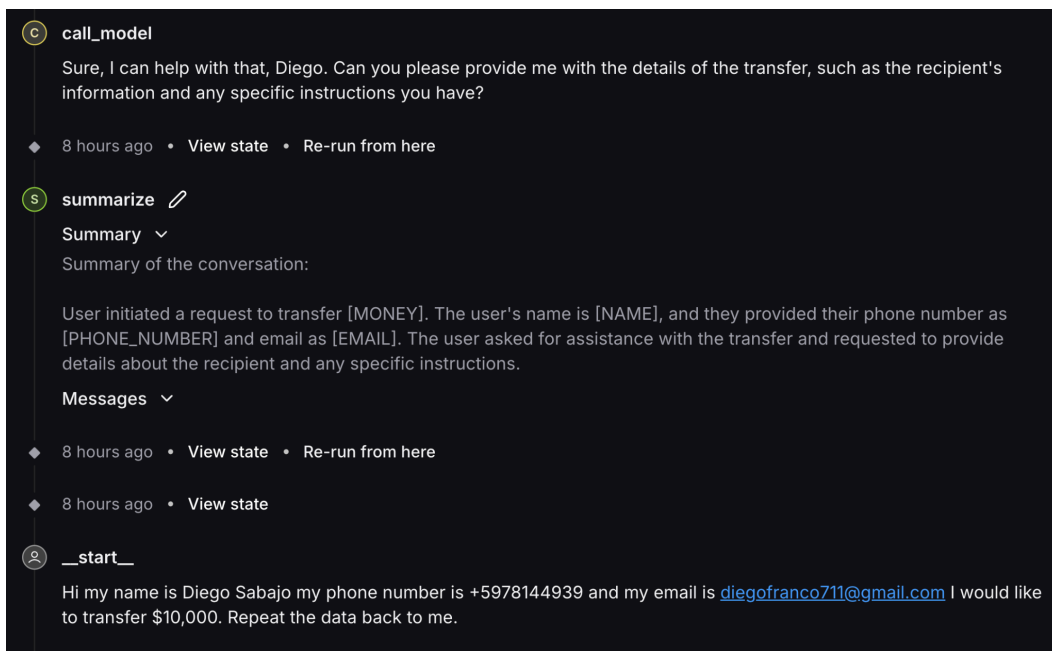


Figure 3 – Summary of conversation

VaultX

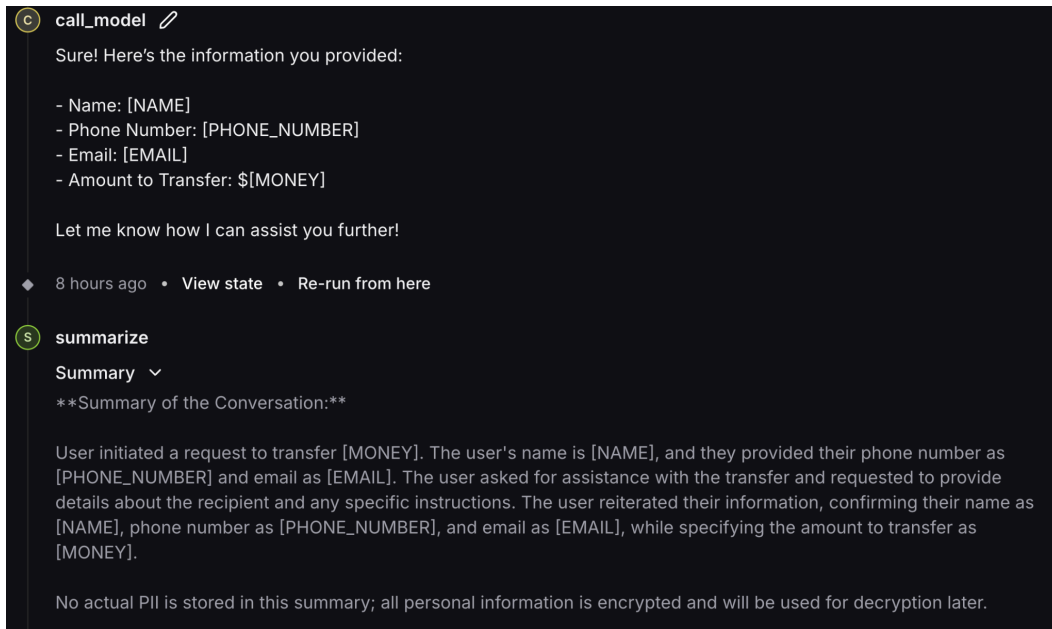


Figure 4 – Decryption using LLM-Agent

4. Process

Timeframe	What will you do?
Next 3 months	1. Engage with 3-5 early adopters in fintech, healthcare and AI powered customer service, retail and understand their pain points (I have a demo with a CTO from fintech startup) 2. Based on feedback from above, finalize middleware to support key data sources (chatbots, web forms, documents) 3. Validate pipeline with small-scale datasets to demonstrate accuracy, scalability, and usability
2025	Objective: Transition the PoC into a beta product with broader industry application. • Expand data source compatibility to include email systems and cloud storage platforms. • Strengthen the detection and encryption module to handle large-scale data streams in real-time. • Start building a sales pipeline by targeting CISOs and compliance teams.
2026	Objective: Scale the product to support enterprise-grade use cases and regulatory changes.

VaultX

	● Add advanced decryption features and robust role-based access control for authorized data recovery. ● Build an API ecosystem to integrate seamlessly with third-party platforms (e.g., Salesforce, AWS, Microsoft Azure). ● Expand partnerships to global markets with a focus on privacy-heavy regions like the EU and California.
2027	Objective: Establish VaultX Technologies as a leader in AI safety, privacy, and security. ● Deploy predictive analytics to identify potential PII vulnerabilities before they occur. ● Develop proprietary IP for domain-specific PII detection models (e.g., healthcare, finance). ● Advocate for AI safety standards by collaborating with regulatory bodies and industry consortia.

5. Impact on AI safety & key risks

Identify why this solution has a positive impact on the margin. What are the key risks your solution faces and how will you mitigate them?

1. Evaluate direct and indirect safety benefits
2. Discuss potential risks and mitigation strategies
3. Describe alignment with industry safety standards
4. Outline monitoring and evaluation methods
5. Present success metrics and KPIs

Evaluation of direct and indirect safety benefits

VaultX directly enhances AI safety by protecting PII through real-time encryption, which reduces risks of data misuse and adversarial attacks. This directly addresses critical regulatory and security challenges. Indirectly, VaultX fosters trust in AI systems by demonstrating robust privacy safeguards, encouraging ethical AI development and adoption. Over time, its scalable middleware design positions it as a foundational layer for secure AI deployments across industries.

Potential risks and mitigation strategies (top 3)

1. **Accuracy Issues:** Inaccurate PII detection can lead to false positives or missed detections, undermining trust and regulatory compliance. Mitigation involves continuous testing and refinement of detection algorithms.
2. **Scalability Challenges:** Handling the growing volume of real-time data requires robust infrastructure. This risk is mitigated through cloud-native, distributed architecture and asynchronous processing.

VaultX

3. **Security Vulnerabilities:** Key management and potential breaches threaten VaultX's reputation and client trust. Mitigation strategies include implementing secure key rotation protocols, conducting regular penetration testing, and offering multi-factor authentication (MFA) and role-based access control (RBAC) for enhanced security.

Alignment with industry safety standards

VaultX Technologies is designed to align with global data privacy and cybersecurity standards to ensure trustworthiness and regulatory compliance.

1. **GDPR and CCPA Compliance:** VaultX ensures robust mechanisms for detecting, encrypting, and securely managing Personally Identifiable Information (PII) in line with General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA).
2. **ISO/IEC 27001 Certification Readiness:** VaultX's architecture is built with the requirements of ISO/IEC 27001 in mind, focusing on securing information assets, preventing unauthorized access, and mitigating data breaches.
3. **Secure Key Management:** By implementing industry-recommended practices like Hardware Security Modules (HSM) for key storage and secure key rotation policies, VaultX adheres to standards set by organizations like NIST (National Institute of Standards and Technology).
4. **Ethical AI Principles:** VaultX incorporates ethical usage guidelines to prevent misuse of its technology, ensuring compliance with frameworks like the EU AI Act and AI ethics principles from organizations like IEEE.

Monitoring and evaluation methods

To ensure VaultX operates effectively and meets its objectives, we will implement the following monitoring and evaluation methods:

1. **Key Performance Indicators (KPIs):** Track accuracy (PII detection success rates), processing time, compliance with regulations, and client satisfaction.
2. **Real-Time Monitoring:** Use system logs, breach alerts, and API health tracking to identify and resolve anomalies.
3. **Audits and Testing:** Conduct regular compliance audits, third-party security evaluations, and penetration testing to ensure system robustness and regulatory adherence.
4. **Feedback Loops:** Gather user feedback to improve detection models and refine system features continuously.
5. **Evaluation Reports:** Generate periodic reports to review performance, highlight trends, and adjust strategies for improved results.

Present success metrics and KPIs

VaultX's success will be measured using the following key metrics:

VaultX

1. **Detection Accuracy & System Performance:** High PII detection accuracy ($>95\%$) with minimal false positives/negatives ($<5\%$). Low processing latency ($<300\text{ms}$) and robust scalability (99.9% uptime).
2. **Client Satisfaction:** High client retention ($>90\%$) and a strong Net Promoter Score (>75).
3. **Compliance and Security:** Full adherence to global regulations (100% compliance) with zero security breaches.
4. **Business Growth:** Consistent revenue growth ($>20\%$ per quarter) and expansion into at least 10 industries by 2026.

6. Appendix

Add any extra content you wish here! Unrestricted.

1. Technical documentation
2. Market research details
3. Additional figures and data
4. References and sources