Libraesva AI usage: technical implementation, governance, privacy and regulatory compliance

Introduction

At Libraesva, we take a decentralized approach to AI, integrating data security into our system's architecture from the outset. This ensures that our business resilience is bolstered by our independence from third-party AI service providers.

We also design and develop our own AI systems in-house, which provides us with full control over our technology and prevents reliance on external cloud-based services or third-party providers. Our AI models operate directly on the device, eliminating any risk of data exfiltration.

By building optimized and specialized AI solutions specifically for distributed architectures, we are able to maintain maximum security, flexibility, and resilience in our systems.

Techical implementation

Our Libraesva AI Engine, integrated into our Email Security solutions, leverages a multilingual tensor-based model that builds upon core generative AI technology. Specifically designed for classification tasks, it provides tailored intelligence for distributed architectures, delivering high accuracy and performance.

The Libraesva AI Engine is optimized to efficiently process the semantics of email messages, without requiring specialized hardware for deployment. While training requires specific hardware, our engine can be deployed on standard hardware, ensuring flexibility and accessibility in real-time threat detection.

By deploying the Libraesva AI Engine directly on our Email Security appliance, we ensure that no customer data is exfiltrated and eliminate reliance on third-party providers. This ensures maximum security, control, and resilience for our customers' email security needs.

Advantages of our approach

No Reliance on Third Parties: Ensures complete independence from external vendors
Ensures Complete Data Privacy: Prevents data exposure by avoiding third-party dependencies
High Resilience Against Geopolitical Instability: Maintains system stability despite international tensions and trade conflicts
Independence from Global Supply Chains: Reduces risk by not relying on components sourced internationally
Predictable Performance Over Time: Avoids potential inconsistencies associated with AI cloud services
Exclusive Use of Distributed, Specialized Models: Employs AI models optimized exclusively for the task at hand

Privacy and compliance

Our Libraesva model is specifically designed to avoid using customer data in its training process, ensuring that no sensitive information is ever utilized. By virtue of its classification-only architecture, the model cannot inadvertently access or transfer training data, providing an added layer of security and compliance.

The model is trained in our EU-based laboratories using a curated dataset of real malicious emails, under the close supervision of our expert analysts. This rigorous quality control process ensures that high-quality data and accurate tagging are always used to develop efficient AI models.

Libraesva takes swift action to keep its appliances secure and up-to-date. We deploy updates to all appliances worldwide within just five minutes of release, ensuring that our customers' email security is always protected against the latest threats.

In addition to frequent releases, we also continuously monitor emerging threat patterns and update our model as needed. Typically, a new version is released once a week or even twice a day, allowing us to stay ahead of evolving malicious activity.

Semantic analysis

Our model is capable of grasping not only the literal meaning of messages but also their underlying intent, taking into account factors such as linguistic nuances, misspellings, and even the language used.

This advanced understanding enables it to abstract meaning from text, allowing it to recognize patterns and detect similar types of attacks across over 100 languages we support.

Large centralized models versus small distributed models

Small models excel in efficiency and accuracy within their specialized contexts compared to larger ones. However, they require careful optimization and expertise to maximize their capabilities with minimal resources.

The Libraesva Adaptive Trust Engine processes email messages by extracting only the relevant information for analysis. This data is then fed into a neural network, which returns a classification (e.g., phishing, business email compromise, transactional, newsletter, etc.) along with a confidence level (ranging from 0% to 100%).

This semantic analysis is integrated with other sources of intelligence, including:

- Local machine learning models.
- Reputation checks.
- Authentication verification.
- Rule-based evaluations.
- Attachment inspection.
- URL analysis.

The result is a holistic and efficient security solution that combines semantic understanding with robust threat detection.

The new AI model doesn’t need any configuration, making its contribution easily observable through two key channels:

1. Message features extracted and displayed as “message indicators” in the ESG User Interface (UI).
2. A detailed technical report accompanying each analyzed message, providing further insight into the model’s performance.

Purpose and proportionality

The primary goal of semantic analysis is to refine the accuracy of message content classification.

The utilization of an AI model for this purpose is entirely justifiable and risk-free, as it leverages cutting-edge technology to detect and classify messages with utmost reliability.

Data protection

Our Email Security appliance ensures that customer data remains isolated and secure, as it processes and analyzes messages entirely on-premises without transmitting any sensitive information outside the appliance.

To maintain the highest level of confidentiality, we do not utilize customer data to train or update our AI model. This approach eliminates any risk of data exfiltration or exploitation.

Our organization adheres to a rigorous AI governance framework, which encompasses risk management and regulatory compliance. The Libraesva AI Governance Committee oversees the deployment and utilization of AI solutions within the company.

Notwithstanding our use of generative AI technology, our models are designed as classifiers rather than extractors. This distinct approach ensures that sensitive information remains inaccessible to the model, thereby preventing potential data breaches or unauthorized disclosure.

By adopting this structured methodology, we remain in alignment with international AI regulatory standards, including those set forth by the EU’s Artificial Intelligence Act.

Our organization’s AI solution is classified as ‘Minimal Risk AI’ under the EU Artificial Intelligence Act, which means it does not fall within the scope of the Act’s specific

requirements and regulations.

As such, our AI approach does not trigger any mandatory obligations or reporting requirements defined by the EU AI Act.