VIANAI

hila™

Generative AI and Data Security

How we keep even the most sensitive data secure while enabling cutting-edge innovation

Executive Summary

Enterprise adoption of Generative AI will always have unique challenges that do not exist in the consumer world. These include operating costs, speed of response, accuracy, and crucially, data security and privacy.

For much of the nascent history of LLMs, data security and privacy have essentially been a promise, as although a company could send their data to a public LLM - such as those provided by OpenAI, Anthropic or Gemini - and those companies would not train on it and keep the company's information secure, proprietary company data would still have to leave the company's private network .

Many value-added GenAI use cases, however, require deeper access to sensitive company data, the kind (such as data from a financial or HR system) that cannot leave a company's private network. This kind of data is far more important as it has critical information for an enterprise, and, therefore, requires diligent security efforts to prevent any leakage to the outside world. Additionally, while open-source models have almost achieved parity with the public models, the burden is on the IT departments to manage the cost, speed and accuracy challenges when deploying these open-source models.

To address the challenges discussed above, we have engineered the hila Platform to achieve all crucial aspects of an enterprise solution - robust data security, cost, speed and highest accuracy, and at the same time eliminate the need for any critical data to ever leave a company's private network – even when using a public LLM.

The Enterprise Data Security Challenge

While an enterprise GenAI solution needs to solve all key aspects of cost, speed, accuracy and data security, this paper is specific to the topic that has often beguiled companies working in this new generative AI world - data security.

Due to the need to achieve quality outcomes, public models are often the only models that work for an enterprise. However, there are businesses, such as those in financial services or defense or healthcare, which will not allow any proprietary data to leave their private network. But there are also data types that are sensitive for any enterprise to share, such as a company's internal financial and accounting information.

There are open-source models which increasingly have performance on par with public models and can sit inside a company's private network. These models, such as Llama, Qwen or Deepseek, have significant advantages over previous generation models, but often to address the accuracy, cost and speed challenges, a massive amount of energy and expertise needs to be applied to the deployment and management of these models – efforts that can take months or years of specialized understanding and engineering. [NB1] Most enterprises are not prepared for the kind of work that's necessary to implement these open-source models in their landscape. Rather, they'd like to get something that has data security as a fundamental tenant, so they can begin to see immediate value.

To summarize, enterprises adopting Generative AI face several unique and significant challenges compared to consumer applications:

- **Data Security:** Companies in sensitive sectors such as financial services or defense often prohibit data from leaving their internal landscape. General enterprises must protect critical internal financial and accounting data, among other types.

- **Cost:** Implementing high-performance models within an enterprise environment can involve substantial expenses, particularly for open-source models, which require extensive customization and fine-tuning.

- **Speed:** Achieving rapid response times suitable for enterprise use often requires dedicated infrastructure and specialized optimization.

- **Accuracy:** Maintaining high accuracy levels, especially with enterprise-critical data, demands significant effort in fine-tuning and continuous model enhancement.

To obtain value from the models, enterprises must overcome additional challenges:

- Extensive expertise and resources are required for fine-tuning and deployment.
- Significant energy and engineering effort must be applied to achieve cost-effective, accurate, and performant solutions.
- The necessary infrastructure and data integration frameworks must be developed, which can take months to years.

Unique Risks and Vulnerabilities of Generative AI:

While the potential of Generative AI in the enterprise is immense, so are its vulnerabilities. The past two years have revealed just how exposed public LLMs - even those offered by the most advanced AI labs - remain to breaches, leaks, and adversarial misuse. These risks are not theoretical; they are well-documented across a range of high-profile incidents involving the largest model providers in the world.

Several categories of risk have emerged:

- **Data Leakage via Bugs or Misconfigurations:** In March 2023, OpenAI's ChatGPT experienced a memory caching bug that briefly exposed other users' chat titles and billing details. Google Bard similarly exposed shared user conversations in search engine results due to a misconfigured indexing policy. These incidents show how even "read-only" chat environments can inadvertently expose private information.

- **Prompt Injection and Model Misuse:** In December 2024, researchers successfully manipulated Anthropic's Claude into downloading and executing malicious payloads, demonstrating that even constrained AI agents can be tricked into violating their security envelope. The original "Sydney" prompt leaks from Microsoft's Bing AI in 2023 also revealed how system instructions could be extracted and repurposed.

- **Credential Theft at Scale:** Hackers have increasingly targeted public LLM accounts. In both 2023 and 2025, over 100,000 ChatGPT accounts were compromised via infostealer malware and resold on dark web forums. While not breaching the models themselves, these attacks expose vulnerabilities in the broader AI ecosystem - namely, the lack of endpoint protections and centralized credential reuse.

- **Training Data Memorization and Exposure:** In late 2023, researchers found that ChatGPT could be coaxed into revealing sensitive snippets from its training data, including emails, adult content, and proprietary text. This kind of unintended memorization highlights a major risk in model development, particularly when training data includes sensitive or copyrighted content.

- **Accidental Internal Data Exposure by Users:** Samsung engineers, like many others, used ChatGPT to streamline tasks, unknowingly pasting confidential code and internal documents into a public AI model. These types of errors can lead to irreversible exposure of trade secrets and underline the necessity of strong internal policy and technical guardrails when using external models.

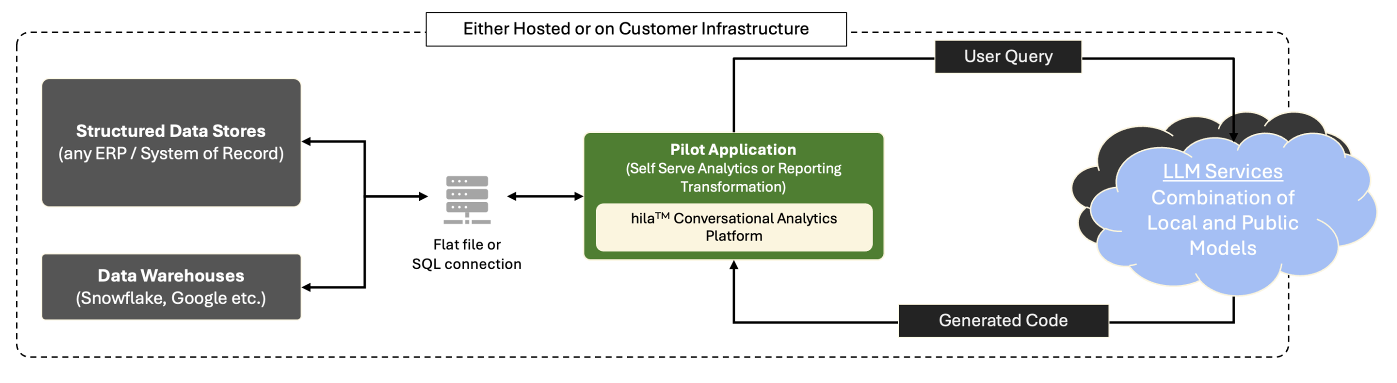
- **Legal and Regulatory Violations:** In both 2023 and 2024, Italy's data protection authority temporarily banned ChatGPT and later fined OpenAI €15 million for non-compliance with GDPR, including insufficient notice of past breaches and unlawful processing of personal data. These events signal the increasing scrutiny enterprises may face when adopting public AI tools that lack transparent data governance.

These incidents - spanning bugs, social engineering, malware, and legal oversight - illustrate that **generative AI cannot be deployed safely in an enterprise context without a secure, controllable, and transparent architecture.**

Enterprises cannot afford to compromise data security, and as these examples make clear, relying on public models to safeguard your data introduces unacceptable risks - both to privacy and to compliance. As a result, Vianai's hila Platform architecture ensures that **your sensitive data never leaves your environment**, and our platform is hardened to withstand the kinds of vulnerabilities now well-documented across the AI industry.

Security Framework and Architecture:

The hila Platform sits on a containerized solution and is not a SaaS offering. As such, it can run all the components in the customer's private network.


Either Hosted or on Customer Infrastructure
User Query
Structured Data Stores
(any ERP / System of Record)
Data Warehouses
(Snowflake, Google etc.)
Flat file or
SQL connection
Pilot Application
(Self Serve Analytics or Reporting Transformation)
hila™ Conversational Analytics Platform
LLM Services
Combination of Local and Public Models
Generated Code

An example of a standard installation on a private cloud

Moreover, as the architecture diagram shows, we never send any sensitive data to any public LLM. We rely on our various fine-tuned models to interpret the user query and our knowledge models to understand the question and how it can be answered. The LLMs then generate the appropriate code that is executed locally inside the customer network to analyze the data, build the charts, and provide explanations. In particular, the public LLM never sees the data, and the code is run, and the data analyzed separately from the models.

Our architecture can work securely on both public models and with our own fine-tuned, open-source models. Our own models can also enable a deep analysis mode, in which hila acts as the business analyst. In this case, we keep the data entirely in the customer landscape, and use our fine-tuned, open-source models to perform robust analysis of the data.

Fine-Grained Access Controls (RBAC + Row/Object-Level Privileges):

hila applies enterprise-grade role-based access control (RBAC) across every surface of the platform—applications, datasets, prompts, connectors, and exports. Roles (e.g., Viewer, Analyst, Data Steward, Admin) encapsulate least-privilege permission sets for actions such as connecting data sources, generating and running code, viewing results, creating or sharing assets, and administering policies. RBAC evaluates the user's identity and group memberships at request time and issues a tightly scoped authorization context that follows the query from intent understanding through code generation and execution. Crucially, the public LLM never receives data; it only receives the minimum metadata needed to produce code, and any subsequent execution runs inside the customer environment under the user's effective permissions. All access decisions and sensitive actions are fully audited to support compliance and forensics.

Beyond roles, hila enforces fine-grained data entitlements via row-, column-, and object-level controls. Users only see the data they are authorized to see: hila computes an "effective permission" by inheriting existing policies from the underlying sources (e.g., warehouse/database/ERP authorization, native row-level filters, masking policies) and/or by applying overlay policies defined in hila. At query time, hila's policy engine injects mandatory predicates (row filters), applies column masking/redaction, and blocks access to unauthorized objects—guaranteeing that LLM-generated SQL or code cannot bypass governance. These controls extend to derived outputs (charts, tables, explanations), to cached artifacts (caches are segmented by user and policy context), and to data exports and shares. In short: whether permissions are managed in hila or in the underlying systems it queries, the platform preserves and enforces them end-to-end so that the data returned is always appropriate for the requesting user.

Ensuring Compliance and Governance:

In today's stringent regulatory landscape, enterprise adoption of Generative AI necessitates rigorous compliance with globally recognized security standards. At Vianai, we recognize that compliance and governance form the backbone of secure enterprise solutions. Our hila Platform adheres strictly to internationally recognized information security standards, including ISO27001 and SOC2 Type 2.

ISO27001 Certification: The ISO27001 certification validates that our information security management system (ISMS) meets globally accepted standards for managing sensitive enterprise data. By adhering to ISO27001, we ensure that:

- Comprehensive risk assessments and data protection measures are consistently applied.

- Continuous improvement processes are implemented for ongoing security enhancements.

- All data handling procedures follow strict governance policies, maintaining confidentiality, integrity, and availability.

SOC2 Compliance: Our adherence to SOC2 standards demonstrates our commitment to safeguarding enterprise data with strict operational controls and robust governance practices. SOC2 compliance ensures:

- Rigorous security controls for preventing unauthorized access and ensuring data privacy.
- Clear operational transparency and accountability, with regular independent audits validating our security practices.
- Ongoing monitoring and evaluation of our security posture to proactively address and mitigate emerging threats.

Through our certified compliance frameworks, we empower enterprises to deploy advanced Generative AI solutions securely and confidently, while effectively navigating the complex regulatory environment. Our governance model provides comprehensive visibility, control, and assurance, enabling organizations to focus on innovation without compromising data security or compliance obligations.

Actionable Recommendations for Enterprise Leaders:

- Prioritize solutions that ensure sensitive data never leaves the internal environment.
- Invest in technologies that offer strong compliance frameworks, such as ISO27001 and SOC2.
- Evaluate the infrastructure required for deploying Generative AI to ensure optimal speed, cost efficiency, and accuracy.
- Train teams on the proper handling of AI systems to prevent accidental exposure of sensitive information.
- Regularly audit and update your AI security and compliance practices to keep pace with evolving risks and regulations.

Conclusion:

Generative AI offers unprecedented potential for innovation and efficiency, yet it comes with significant responsibilities regarding data security and compliance. By leveraging secure, compliant, and high-performance solutions such as the hila Platform, enterprises can safely harness the transformative power of Generative AI, maintaining complete control over their sensitive data while meeting stringent regulatory requirements.