privacera | paig

SOLUTIONS

# Data Security Governance for AWS and Foundation Models

Securely Innovate with Generative AI

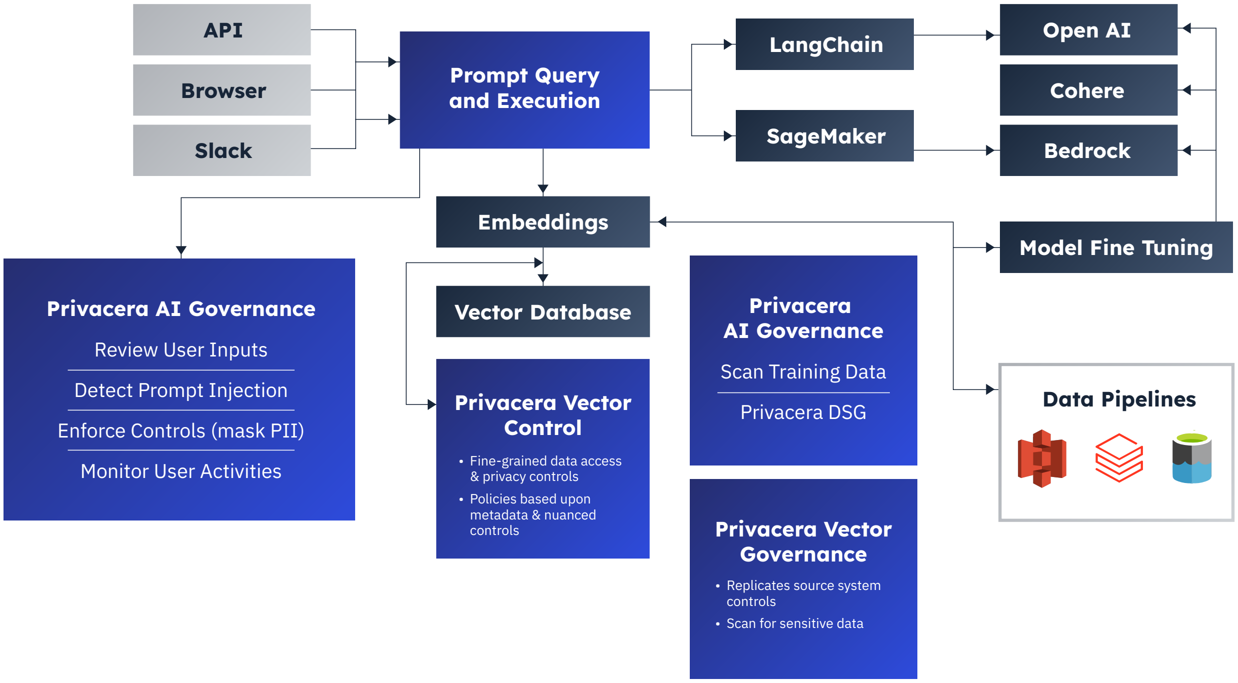
## Generative AI and Large Language Models (LLMs) have captured corporate imaginations.

While the promises are significant, the risks are concrete, including sensitive and unauthorized data exposure, IP leakage, and regulatory compliance failures. Privacera AI Governance (PAIG) added to your AWS ecosystem secures the entire lifecycle of building and deploying Generative AI models and apps, from discovery of sensitive data to protecting and continuously monitoring model usage.

## KEY BENEFITS

- Securely Innovate
- Protect PII
- Prevent IP Leakage
- Track & Audit

## ARCHITECTURE



As shown above, AWS enables enterprises to build and scale generative AI-based applications using foundation models (FMs). AWS Bedrock provides a scalable, reliable, and secure managed service. AWS Generative AI services like Amazon SageMaker Jumpstart, and Amazon Bedrock work with a wide selection of FMs. Using Amazon Bedrock or Amazon SageMaker Jumpstart, organizations can fine-tune their model for a particular task without having to annotate large volumes of data.

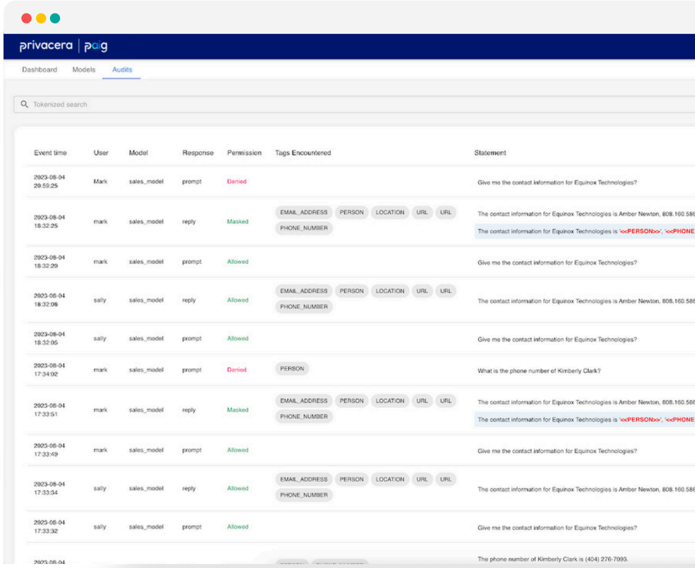
By adding PAIG to the AWS ecosystem and architecture, organizations can secure and govern Generative AI applications running in AWS powered by Amazon SageMaker and Amazon Bedrock. PAIG provides the ability to responsibly govern sensitive or regulated data within FMs. PAIG does this by seamlessly integrating with Bedrock-supported, open-source, and proprietary LLM models and workflows. PAIG enables a comprehensive suite of capabilities to address privacy, security, and compliance concerns associated with the use of LLMs.

How does this architecture work for data consumer interactions? The data consumer asks a question (prompt) to the Generative AI application. The application, in turn, sends this prompt through Bedrock to the LLM framework. A Privacera plugin transparently intercepts the prompt and checks for prompt injection, user authorization, and prompt details. The plugin overrides the initialization methods and registers itself within the execution path, ensuring it intercepts all further interactions. It then sends the data to the LLM hosted as Amazon SageMaker or Amazon Bedrock endpoint. The Privacera plugin verifies the response from LLMs for unauthorized data and applies any required response redaction. All details related to the user, context, prompt, response, and action are audited and securely stored for future reference. The sanitized and approved response is then returned back to the user through the client application.

## KEY CAPABILITIES

## Secure Model Inputs and Outputs

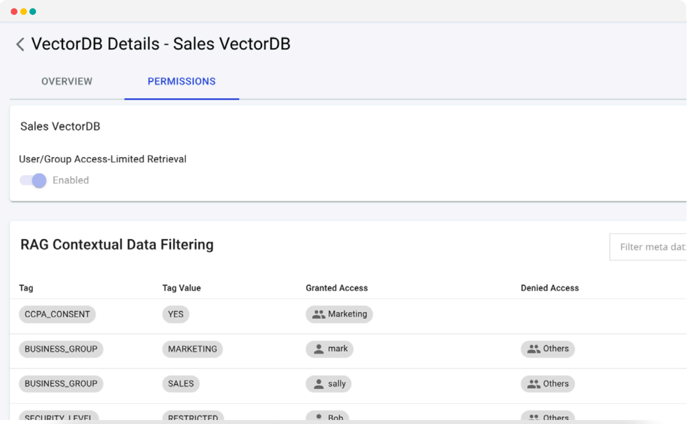
PAIG protects data exposure using context-aware data protection, inspecting user-prompted queries and masking or redacting data requests containing improper or sensitive data. Additionally, Attribute-Based Access Control (ABAC) and Tag-Based Access Controls (TBAC) can be applied to mask sensitive data model output, ensuring users can only see data they are authorized to see. User prompts into the application and model are also inspected. Unauthorized questions that could expose sensitive data or are deemed toxic can be denied. Not only does this capability add an additional layer of security, but it also eliminates the massively expensive and wasteful LLM compute costs associated with processing unauthorized requests.



| Event time | User | Model | Response | Permission | Tags Encountered | Statement |
|---|---|---|---|---|---|---|
| 2023-08-04 20:59:25 | Mark | sales_model | prompt | Denied | | Give me the contact information for Equinox Technologies? |
| 2023-08-04 18:32:25 | mark | sales_model | reply | Masked | EMAIL_ADDRESS PERSON LOCATION URL URL PHONE_NUMBER | The contact information for Equinox Technologies is Amber Newton, 808.160.586<br>The contact information for Equinox Technologies is '<<PERSON>>', '<<PHONE |
| 2023-08-04 18:32:20 | mark | sales_model | prompt | Allowed | | Give me the contact information for Equinox Technologies? |
| 2023-08-04 18:32:08 | sally | sales_model | reply | Allowed | EMAIL_ADDRESS PERSON LOCATION URL URL PHONE_NUMBER | The contact information for Equinox Technologies is Amber Newton, 808.160.586 |
| 2023-08-04 18:32:05 | sally | sales_model | prompt | Allowed | | Give me the contact information for Equinox Technologies? |
| 2023-08-04 17:34:02 | mark | sales_model | prompt | Denied | PERSON | What is the phone number of Kimberly Clark? |
| 2023-08-04 17:33:51 | mark | sales_model | reply | Masked | EMAIL_ADDRESS PERSON LOCATION URL URL PHONE_NUMBER | The contact information for Equinox Technologies is Amber Newton, 808.160.586<br>The contact information for Equinox Technologies is '<<PERSON>>', '<<PHONE |
| 2023-08-04 17:33:49 | mark | sales_model | prompt | Allowed | | Give me the contact information for Equinox Technologies? |
| 2023-08-04 17:33:34 | sally | sales_model | reply | Allowed | EMAIL_ADDRESS PERSON LOCATION URL URL PHONE_NUMBER | The contact information for Equinox Technologies is Amber Newton, 808.160.586 |
| 2023-08-04 17:33:32 | sally | sales_model | prompt | Allowed | | Give me the contact information for Equinox Technologies? |
| 2023-08-04 | | | | | PERSON PHONE_NUMBER | The phone number of Kimberly Clark is (404) 276-7093. |

## Enforce and Maintain Privacy Controls on Vector Databases and Enrichments

PAIG enables fine-grained data access and privacy controls for all vector databases and embeddings, transforming your GenAI application security, privacy, and compliance from brick to a titanium firewall for your GenAI applications. PAIG accomplishes this for you via the following mechanisms:



Sales VectorDB

User/Group Access-Limited Retrieval: Enabled

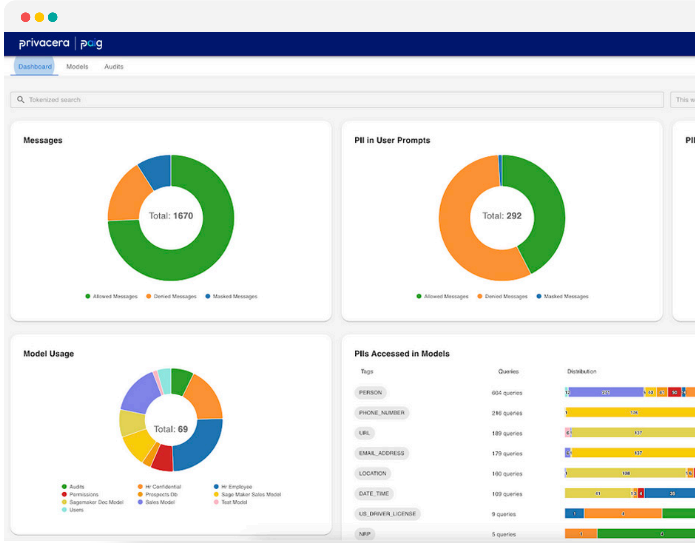
RAG Contextual Data Filtering

| Tag | Tag Value | Granted Access | Denied Access |
|---|---|---|---|
| CCPA_CONSENT | YES | Marketing | |
| BUSINESS_GROUP | MARKETING | mark | Others |
| BUSINESS_GROUP | SALES | sally | Others |
| SECURITY_LEVEL | RESTRICTED | Bob | Others |

- All access and policies applicable on the source systems is retained and replicated into the vector database. To do this, PAIG applies user/group permissions where vector databases (VectorDBs) including OpenSearch contain sensitive information only users with clearance can access.
- Additionally, fine-grained authorization can be implemented to comply with regulatory requirements. PAIG enables enforcement of policies based upon real-time content scanning.

- During retrieval, PAIG filters results, returning only data chunks the user or group has permissions and entitlements to access. This ensures users receive only appropriate information, but also limits the amount of data processed and returned. To work, administrators create user and group-level policies for VectorDB collections. As data chunks are imported, they are tagged with appropriate classifications. PAIG then manages the tags accessible for individual users, either directly or through group memberships. This determines which data chunks should be accessible when querying through a GenAI application.
- Fine-grained metadata filtering takes this a step further by adding a layer of specificity based on the values within each tag. This enables policies based on metadata values. It also allows for nuanced access control that can be tailored to organizational needs and compliance requirements.

## Comprehensive Compliance Monitoring



| Tags | Queries | Distribution |
|---|---|---|
| PERSON | 604 queries | 12 221 5 40 41 50 |
| PHONE_NUMBER | 216 queries | 1 176 |
| URL | 189 queries | 6 137 |
| EMAIL_ADDRESS | 179 queries | 6 137 |
| LOCATION | 160 queries | 1 138 |
| DATE_TIME | 109 queries | 61 13 4 35 |
| US_DRIVER_LICENSE | 9 queries | 1 4 |
| NRP | 5 queries | 1 4 |

PAIG provides comprehensive dashboards and audit logs of all application and model activity, detailing what sensitive data is leveraged in each model, how it is protected, and who is accessing it. PAIG audit logs show:

- Who is accessing what models
- What sensitive data they are accessing
- When they accessed the model
- Flagged, inappropriate conversations
- What protections were applied

Additionally, PAIG provides a security and compliance dashboard that provides a view of your entire model landscape, including an overview of approved requests, denied requests, and requests that require masking to be applied. The dashboard also provides an overview of all sensitive data across models. PAIG's audit log and dashboard simplifies model monitoring and compliance.

## Support for Open Standards

PAIG leverages the open standards and approaches already used in the Privacera Data Security Platform. In addition, PAIG is designed to embed easily into existing Generative AI application libraries and frameworks to make it frictionless to add to an existing Generative AI application. PAIG supports seamless integration into Open AI, Bedrock, Sagemaker, Langchain, and key prompt and query execution systems.

Fortune 500 enterprises trust Privacera for their universal data security, access control, and governance. Discover how to streamline data security governance with Privacera.

**Take a unified approach to data access, privacy, and security with Privacera.**

REQUEST A DEMO →    CONTACT US →

Privacera, based in Fremont, CA, was founded in 2016 by the creators of Apache Ranger™ and Apache Atlas. Delivering trusted and timely access to data consumers, Privacera provides data privacy, security, and governance through its SaaS-based unified data security platform. Privacera's latest innovation, Privacera AI Governance (PAIG), is the industry's first AI data security governance solution. Privacera serves Fortune 500 clients across finance, insurance, life sciences, retail, media, consumer, and government entities. The company achieved AWS Data and Analytics Competency Status, and partners with and supports leading data sources, including AWS, Snowflake, Databricks, Azure and Google. Privacera is recognized as a leader in the 2023 GigaOm Radar for Data Governance; was named a 2022 CISO Choice Awards Finalist; and received the 2022 Digital Innovator Award. The company is also named a "Sample Vendor" for data security platforms in the Gartner® Hype Cycle™ for Data Security, 2023. Learn more at Privacera.com.

privacera