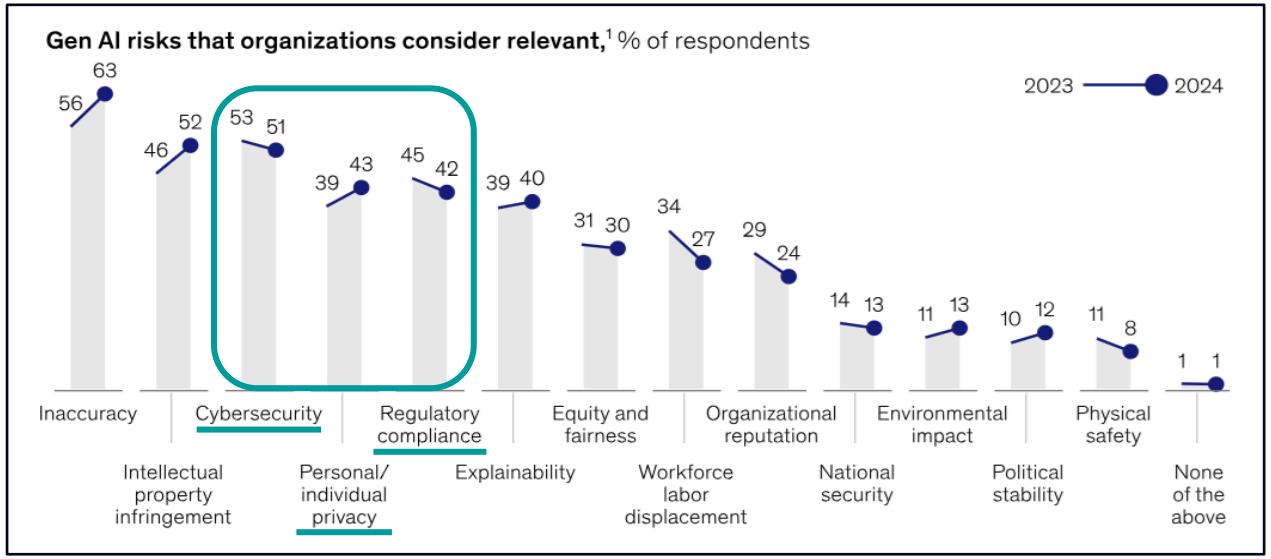
baffle

PGConf NYC 2024

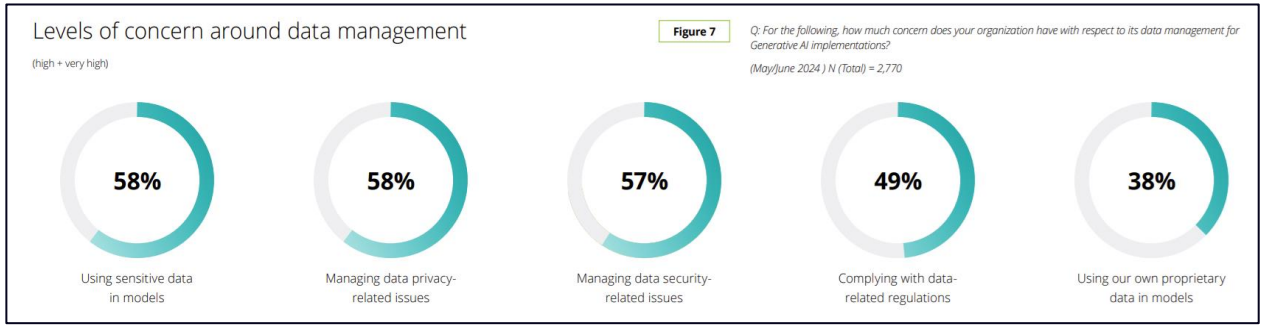
Securing PostgreSQL for use with Generative AI

Data Privacy & Security a Major Issue for GenAI

- Data privacy and security remain a problem despite rapid development of GenAI technical ecosystem
- No <u>provably</u> effective built-in technology for controlling PII and other sensitive data in GenAI



Source: McKinsey State of AI in early 2024



Source: Deloitte's State of Generative AI in the Enterprise Q3 Report, August 2024

Data Points Justifying the Concern

Apr 30, 2024 - Technology

Researchers uncover servers filled with government secrets

Sam Sabin

Databases storing approximately 550 gigabytes of secret data from a government artificial intelligence contractor were exposed on the internet until the end of last month, according to a report released Tuesday.

Why it matters: Plenty of attention has been given to protecting confidential information from entering AI models, but the new research suggests more focus needs to be given to how AI models' training data itself is stored.

Artificial Intelligence

ChatGPT Leaks Sensitive User Data, OpenAI Suspects Hack

The leaks exposed conversations, personal data, and login credentials.

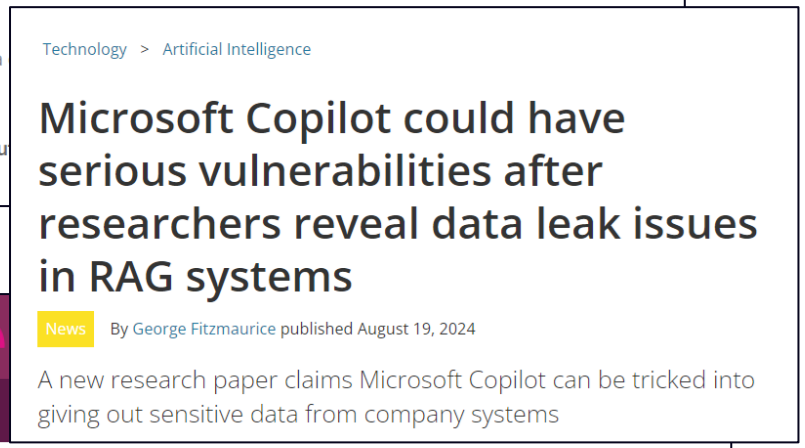
Anuj Mudaliar Assistant Editor - Tech, SWZD February 1, 2024

DARKREADING NEWSLETTER SIGN-UP

Simple Hacking Technique Can Extract ChatGPT Training Data

Apparently all it takes to get a like "poem" forever.

Jai Vijayan, Contribu
December 1, 2023

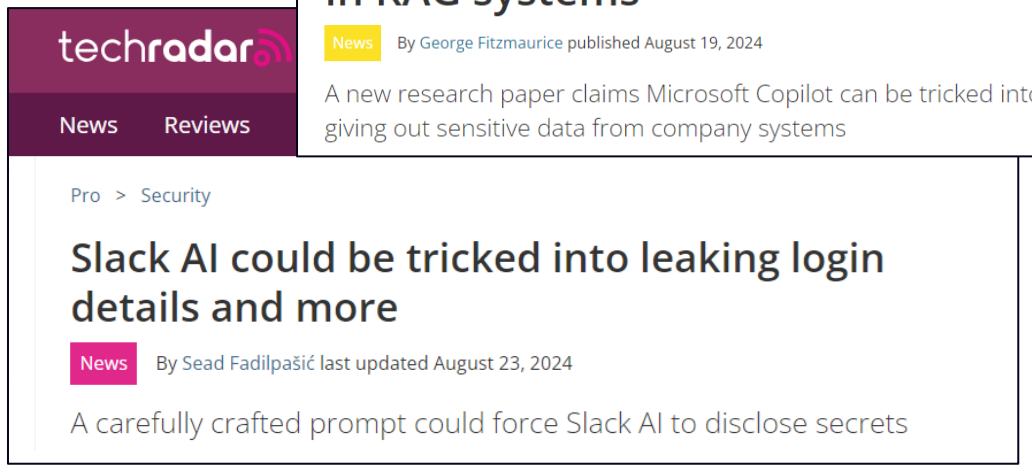


Technology > Artificial Intelligence

Microsoft Copilot could have serious vulnerabilities after researchers reveal data leak issues in RAG systems

News By George Fitzmaurice published August 19, 2024

A new research paper claims Microsoft Copilot can be tricked into giving out sensitive data from company systems



techradar

News Reviews

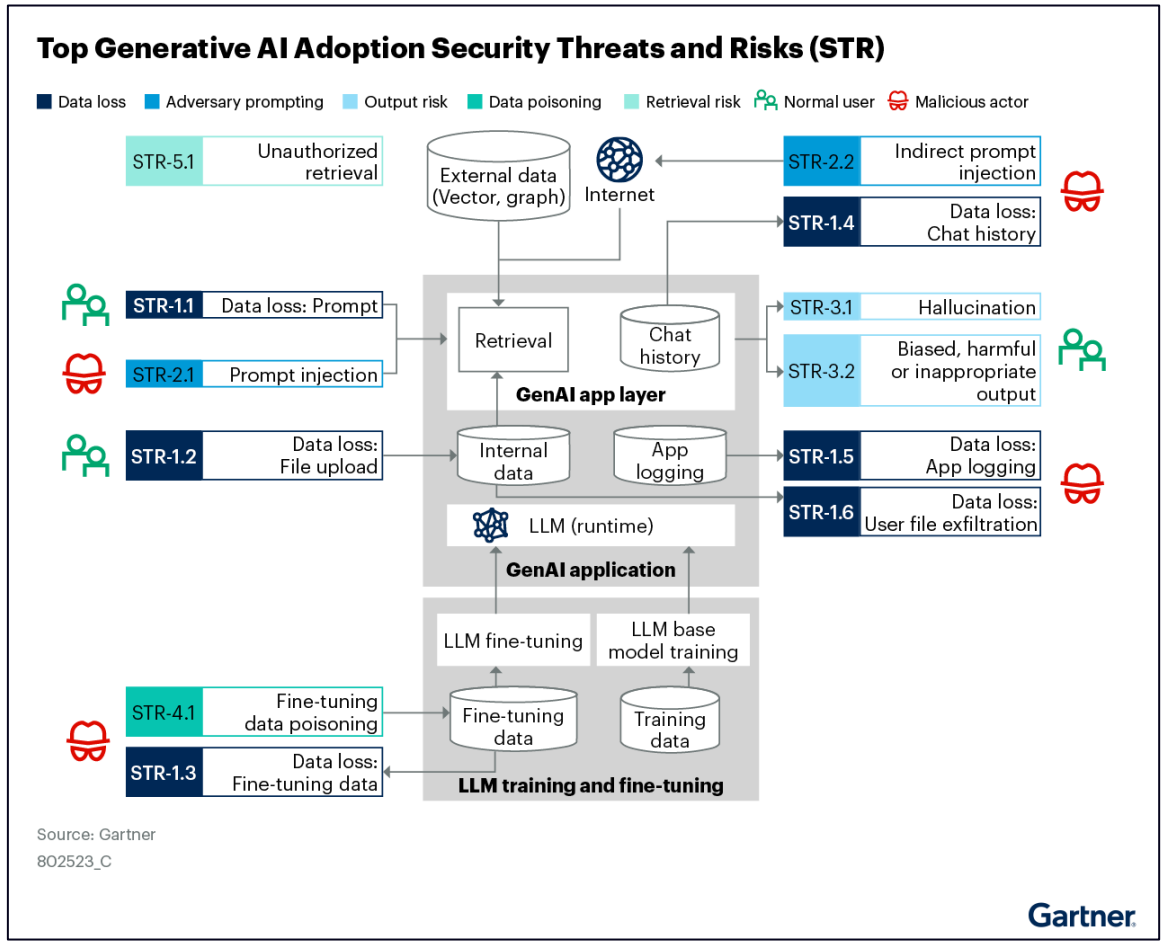
Pro > Security

Slack AI could be tricked into leaking login details and more

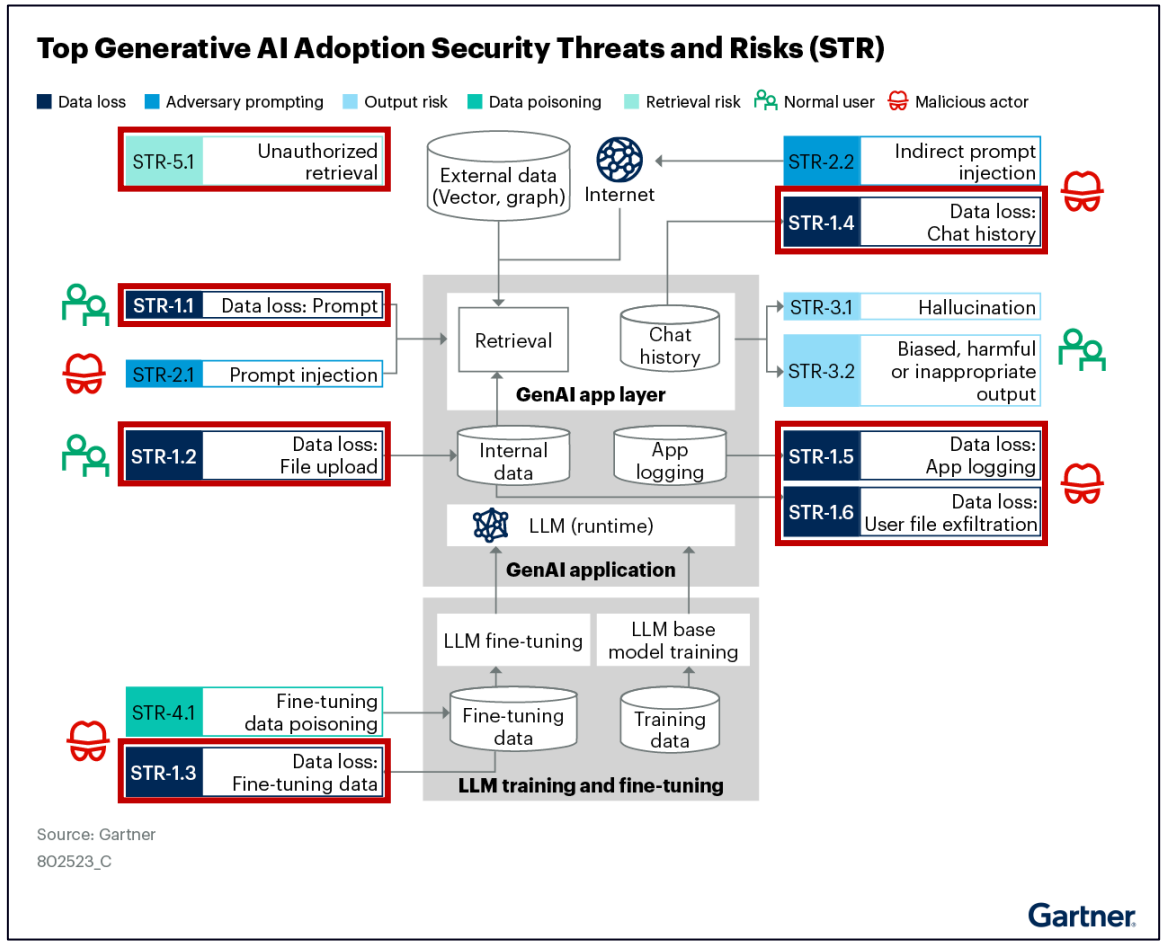
News By Sead Fadilpašić last updated August 23, 2024

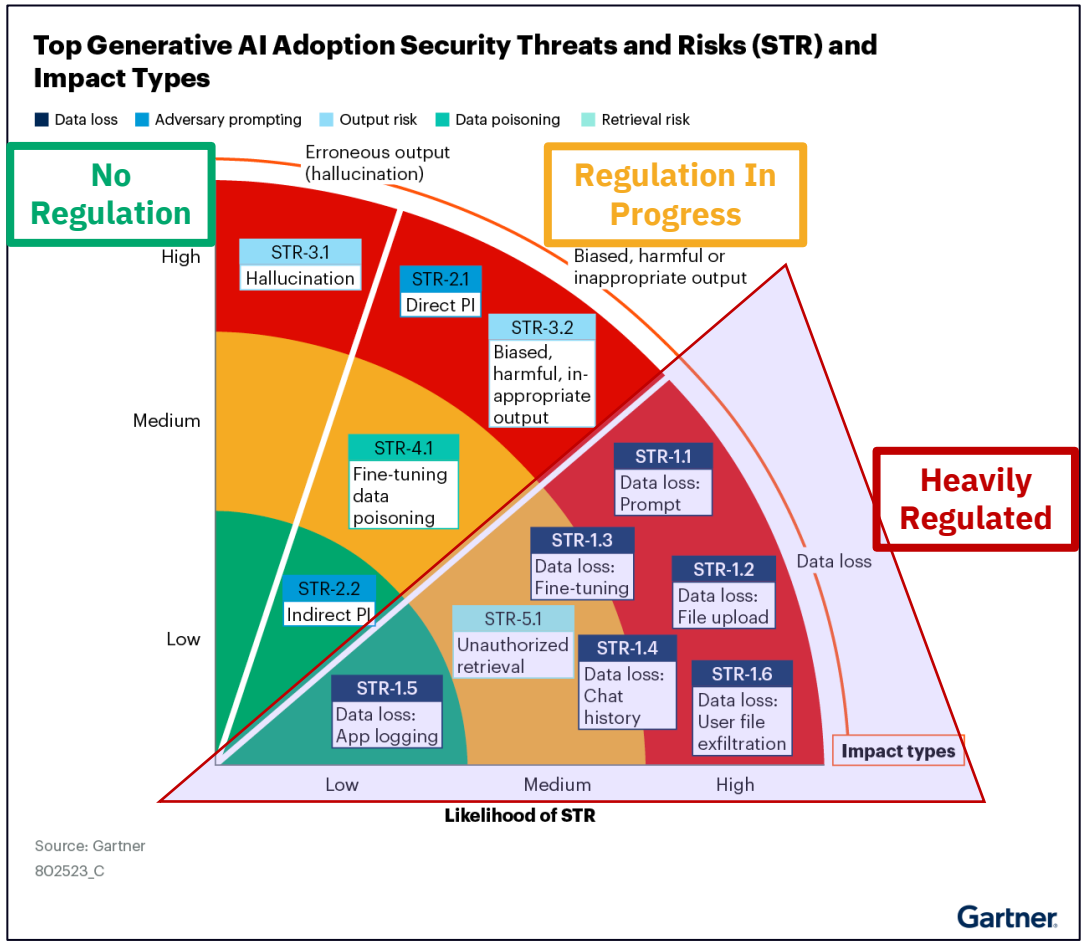
A carefully crafted prompt could force Slack AI to disclose secrets

A More Systematic Look at the Risks



Why the Risks Matter





Compliance Requires Provable Access Control

- Core to all data privacy and security regulations is ensuring the control of sensitive data
- Provability is required for demonstrate effectiveness of the controls in audits

PCI DSS 4.0

"An access control system(s) is in place that **restricts access based on a user's need to know** and covers all system components."

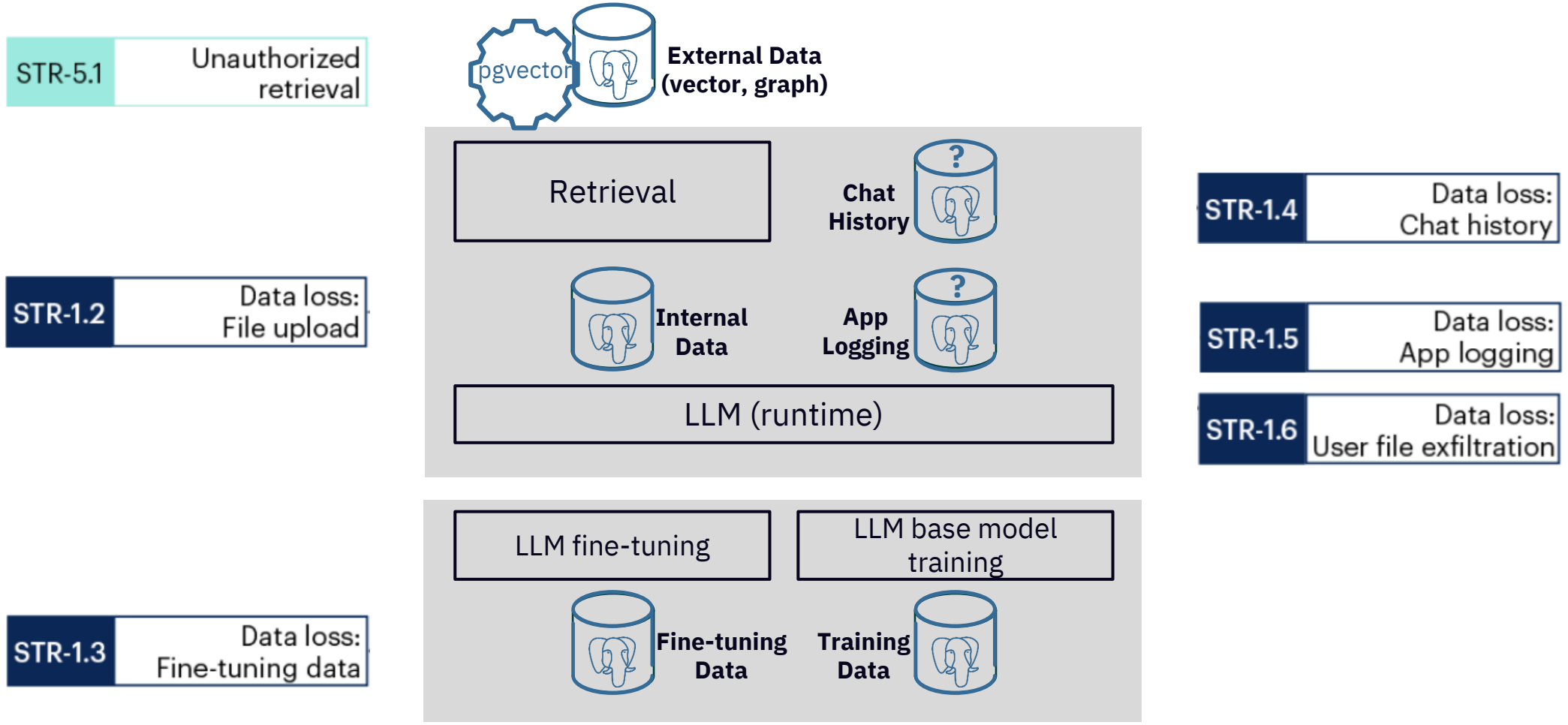
HIPAA

"(a)(1) Standard: Access control. Implement technical policies and procedures for electronic information systems that maintain electronic protected health information to **allow access only to those persons** or software programs that have been granted access rights as specified in §164.308(a)(4)."

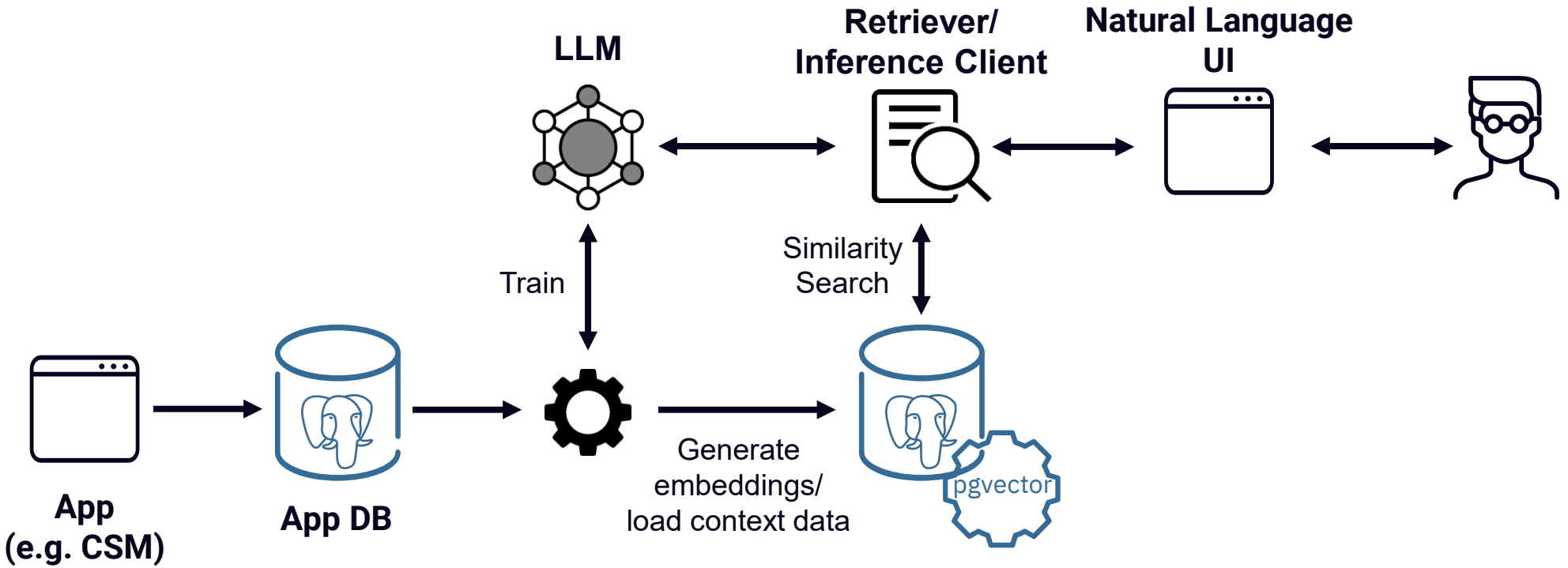
GDPR

"Personal data should be processed in a manner that ensures appropriate security and confidentiality of the personal data, including for **preventing unauthorised access to or use of personal data** and the equipment used for the processing."

What Does GenAI Mean for PostgreSQL



Sensitive Data Enters Early in Gen AI Pipeline



Hypothetical Source Data and Embeddings

Support Tickets Table – support_tickets

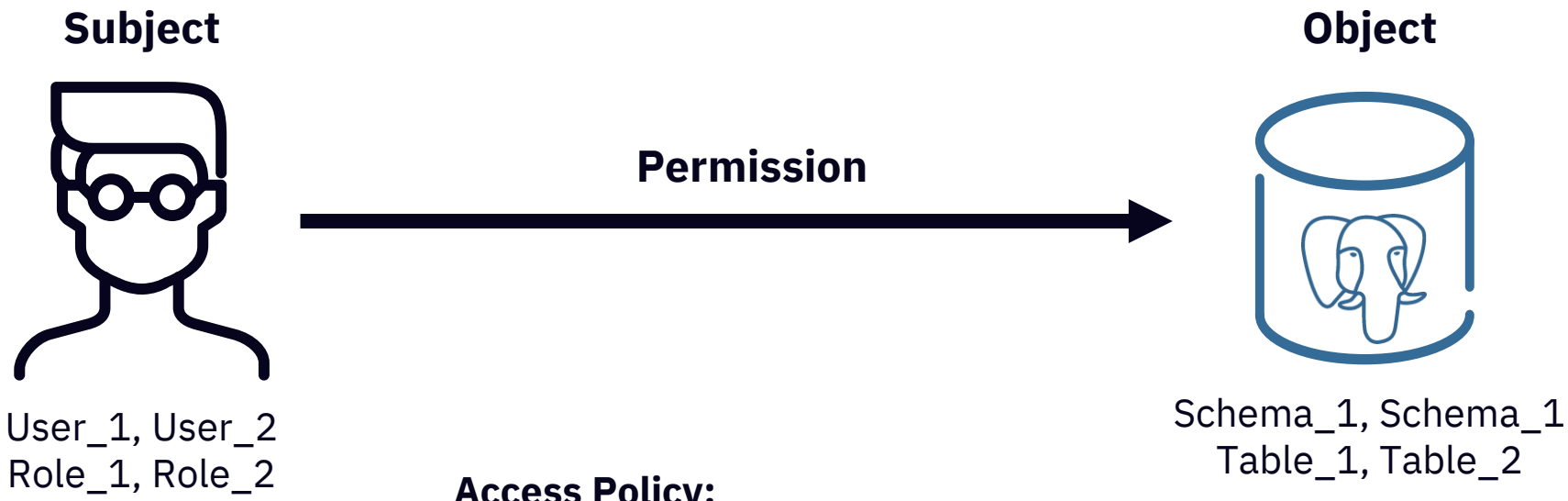
Ticket_ID	Email	Subject	Details
1001	jdoe@acme.com	Unable to Access Online Account	... I am writing to report an issue I encountered while attempting to ..
1002	jane.smith@example.com	Transaction Discrepancy	... I'm writing to bring to your attention a discrepancy ..
1003	robert.jones@example.com	How to close my account	... I am writing to inquire about the process for closing my credit card account ..
1004	sarah.davis@sample.com	Problem with bill payment	...I encountered an issue while trying to use my card for an online bill payment ..
1005	mwilson@gmail.com	Error accessing your info update site	...I received an email from you that my account would be closed if I don't ..

Embeddings Table – support_tickets_embeddings

Content	Tokens	Embeddings
Filer email is doe@acme.com, Subject:Unable to Access Online Account Description: ...I am writing to report an issue I encountered while attempting to ..	539	[0.021440856158733368, 0.022003607824444772, -0..
Filer email is jane.smith@example.com Subject is Transaction Discrepancy Description is ...I'm writing to bring to your attention a discrepancy ..	753	[0.0245039766559492878, -0.000169642977416515, 0....
Filer email is robert.jones@example.com Subject is How to close my account Description is ... I am writing to inquire about the process for closing ..	320	[0.03550934555492730, 0.047169963686414836, 0....
Filer email is sarah.davis@sample.com Subject is Problem with bill payment Description is...I encountered an issue while trying to use my card for an online ..	289	[0.011440856158733368, 0.00847360782495234, -0.0..
Filer email is mwilson@gmail.com Subject is Error accessing your.. Description is ...I received an email from you that my account would be closed ..	134	[0.022517921403050423, -0.00191582809210237303, ...

What Happened to the Access Control Policy?

Data Access Control in Postgres Database



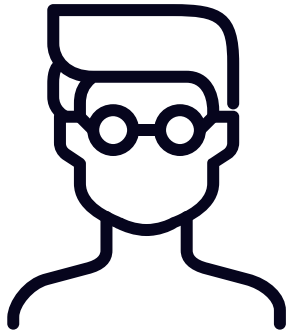
Access Policy:

```
SELECT grantee, privilege_type
FROM information_schema.role_table_grants
WHERE table_name = 'Table_1';

 grantee | privilege_type
---------+---------------
   User_1|         INSERT
   User_1|         SELECT
   User_1|         UPDATE
      …  |             …
```

Data Access Control for GenAI Systems

Subject

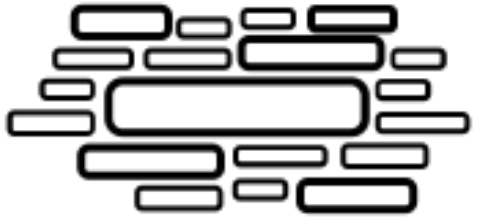


User_1, User_2
Role_1, Role_2

Permission

Access Policy: ???

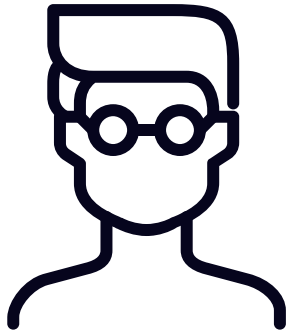
Object?



“Filer email address is jdoe@acme.com”

The Solution is to Control Individual Data Values

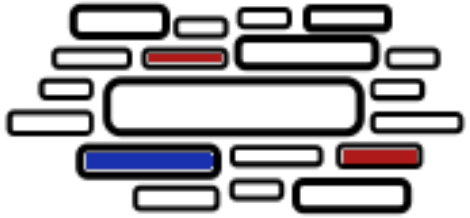
Subject



User_1, User_2
Role_1, Role_2

Permission

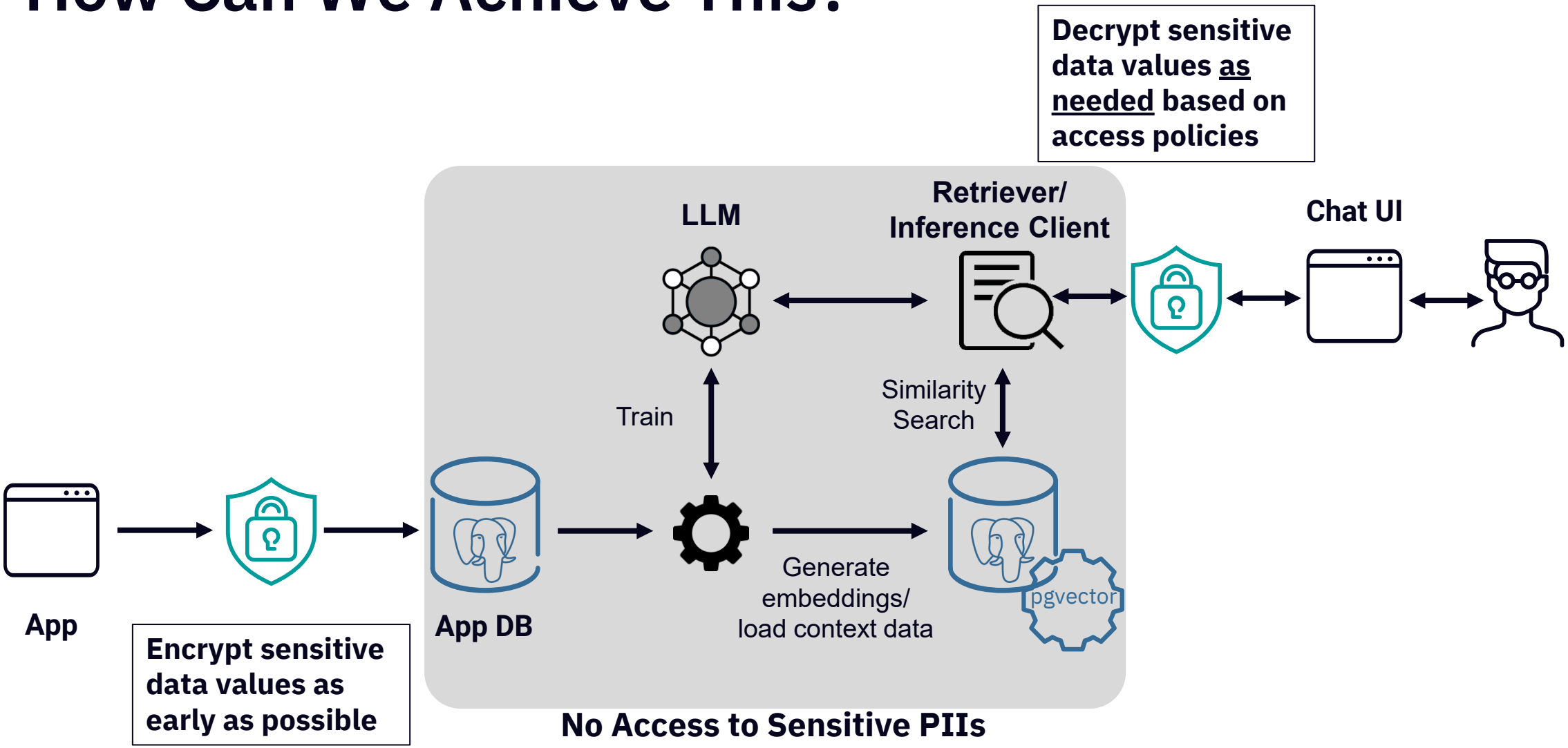
Object



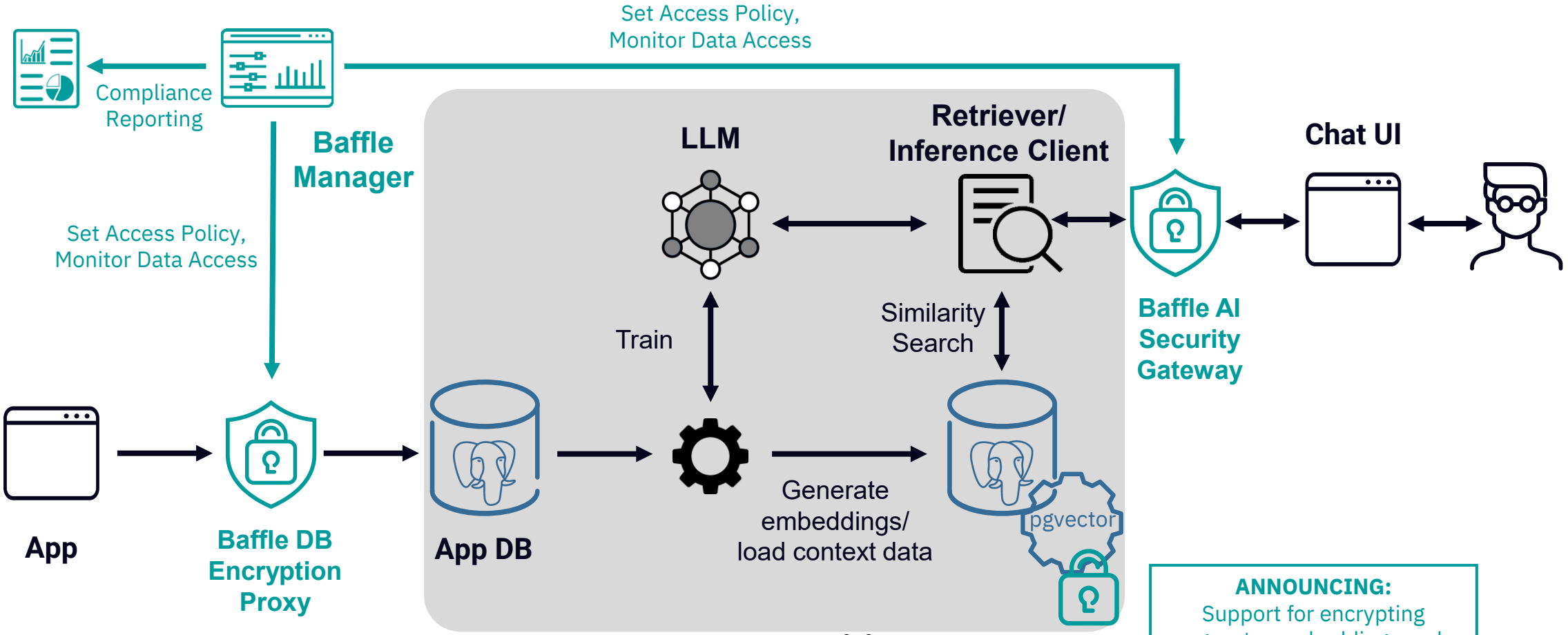
“Filer email address is [MASKED].

Access Policy:
User_1 can decrypt and see email addresses
Role_2 can see masked email addresses

How Can We Achieve This?



Baffle Provides a Complete Solution

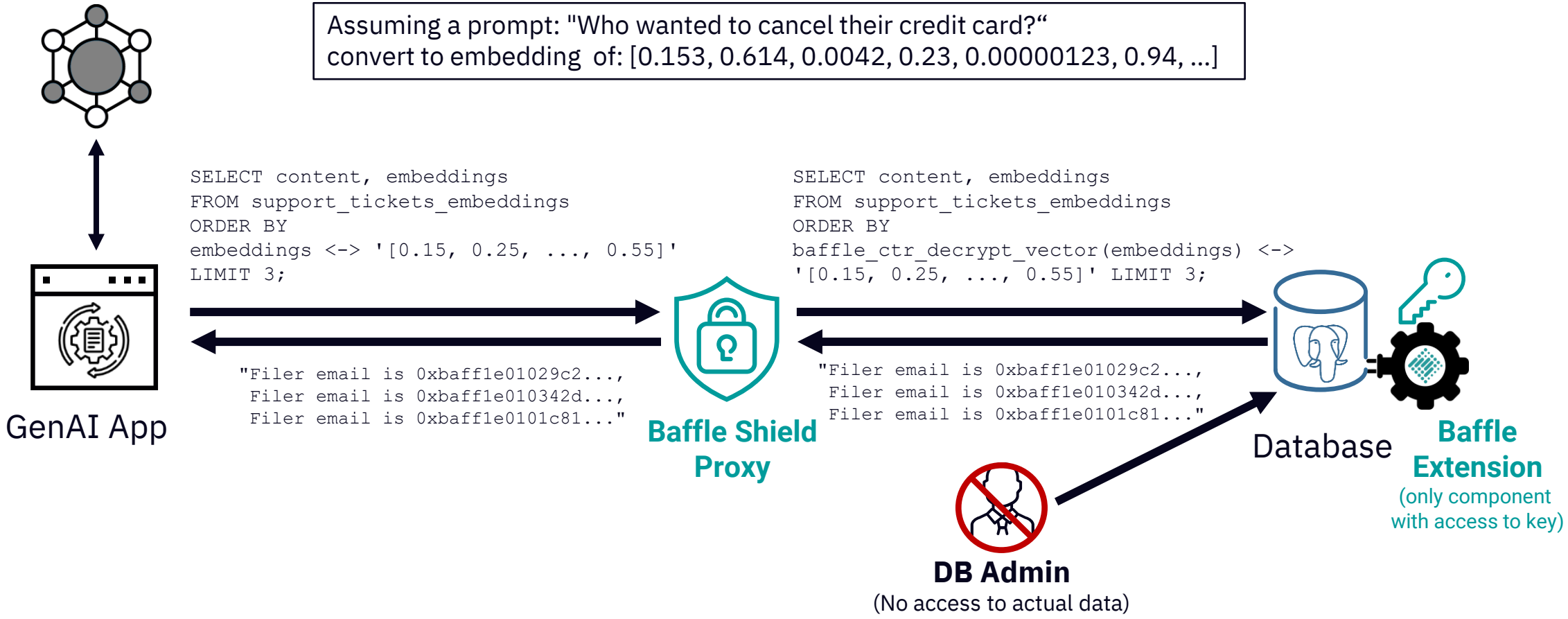


ANNOUNCING:
Support for encrypting pgvector embeddings and operating (e.g. similarity search) on encrypted vectors

Pgvector Encryption Support

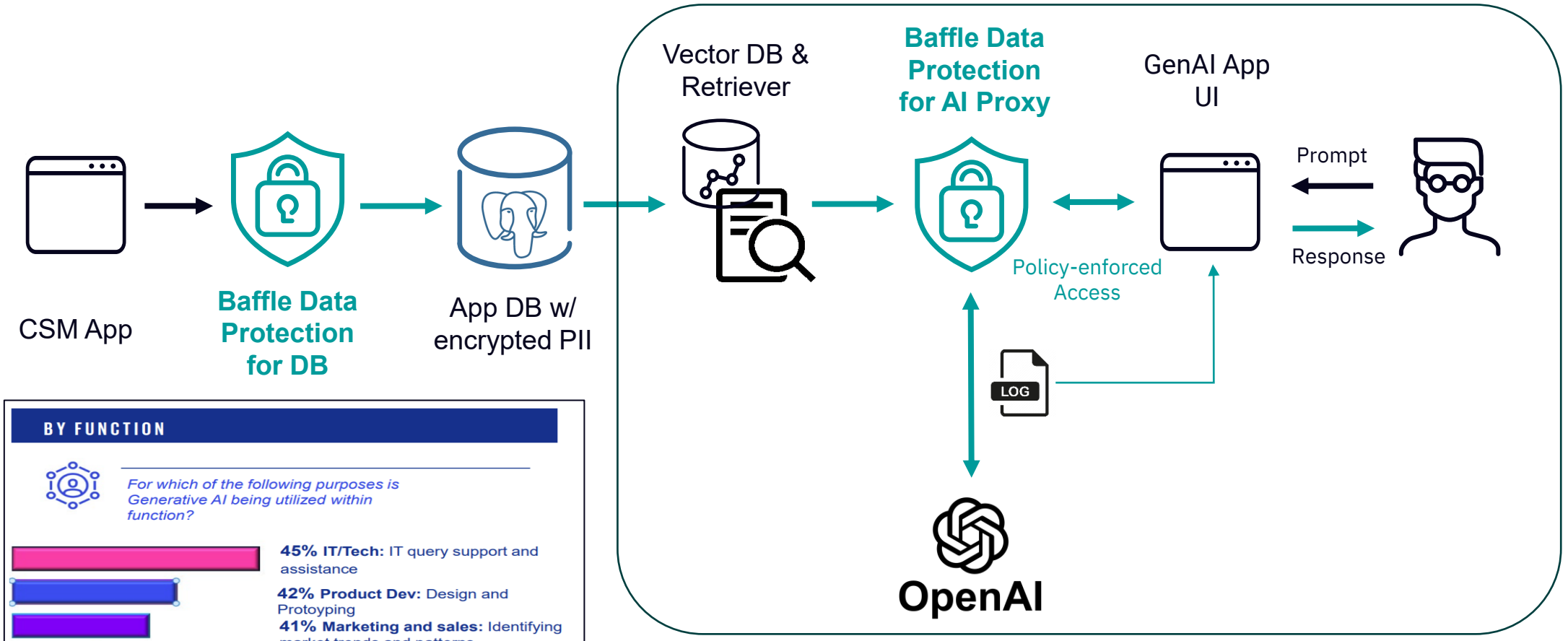
- Allows embeddings to be generated on original values rather than encrypted values
- Produces same response as unencrypted data while eliminating data leakage risks
- Best used for Retrieval Augmented Generation over private data set using public or non-sensitive models

Querying Encrypted Vectors

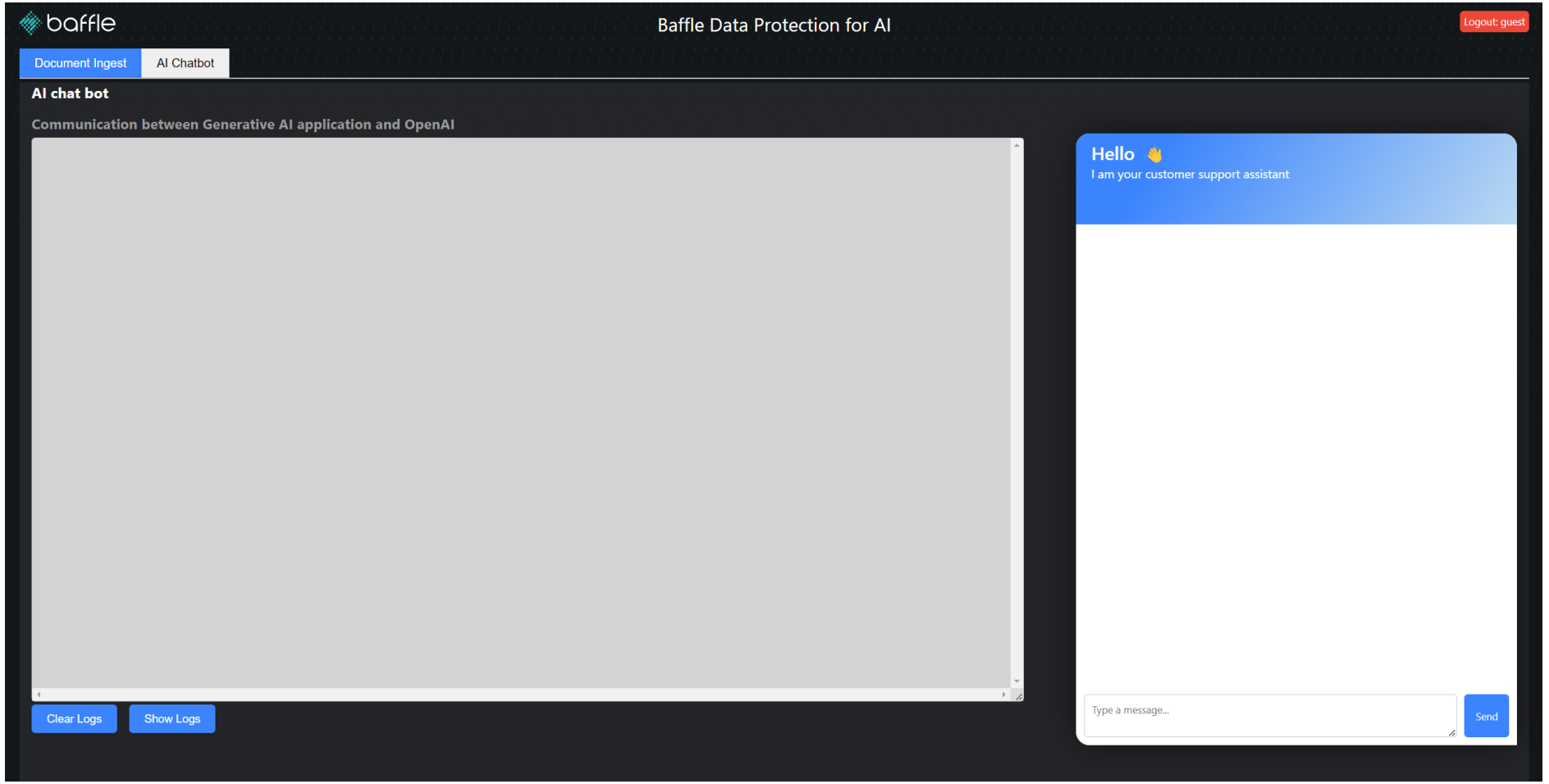


Demonstration

Demonstration Setup (IT Support Chat)



Demonstration



Key Takeaways

- Data privacy and security is huge problem for generative AI systems with no viable solutions available in the GenAI toolbox
- As a popular structured data store, PostgreSQL plays a large role in GenAI application pipelines and in the security of GenAI applications
- The best and most effective approach is to encrypt sensitive data values at column level as early as possible in PostgreSQL and decrypt at end of GenAI pipeline on as-needed basis
- Baffle provides an easy way to enable field-level encryption and access control for Postgres giving GenAI applications that use Postgres a path towards compliant usage

Thank You!