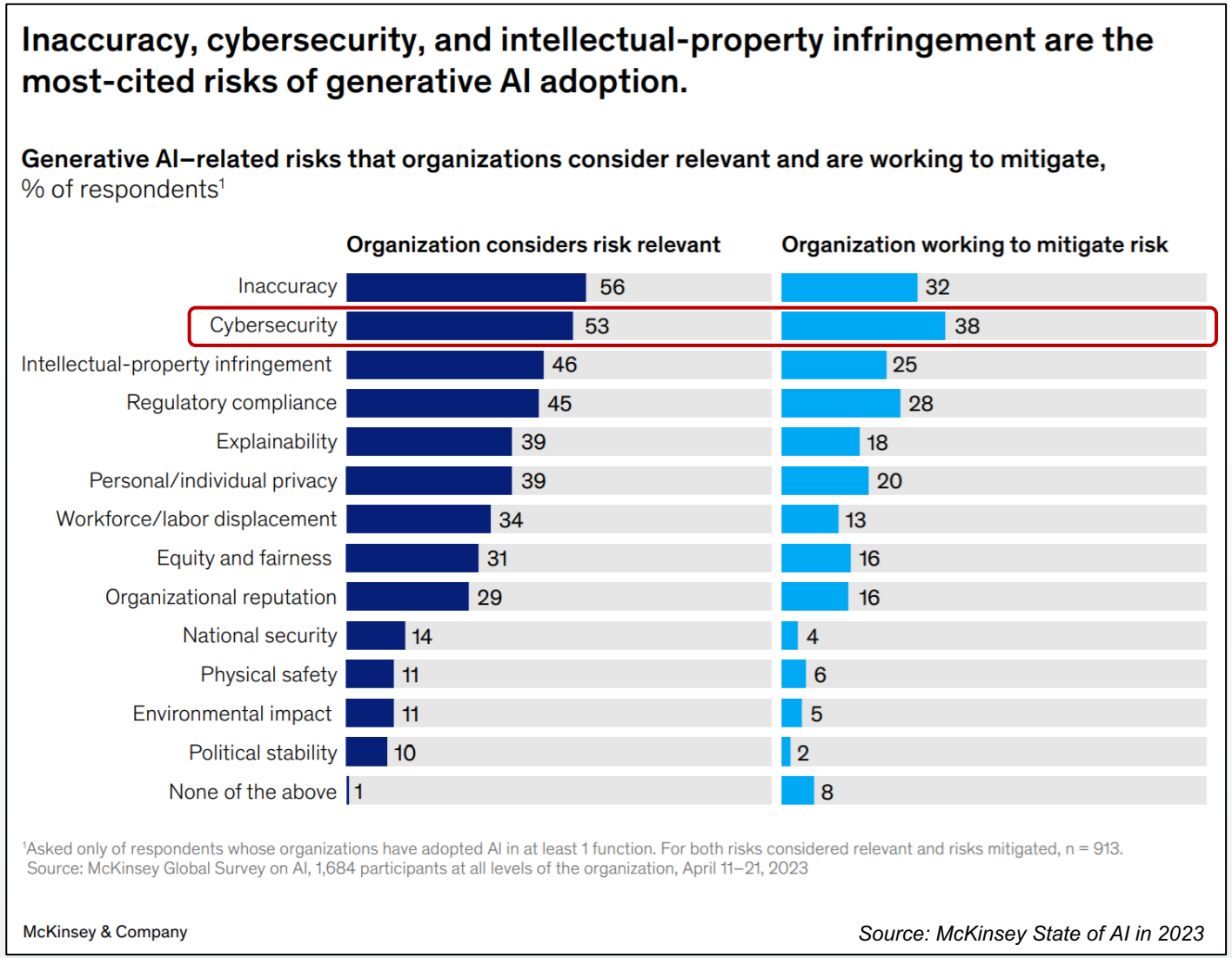
Securing PostgreSQL for use with Generative AI

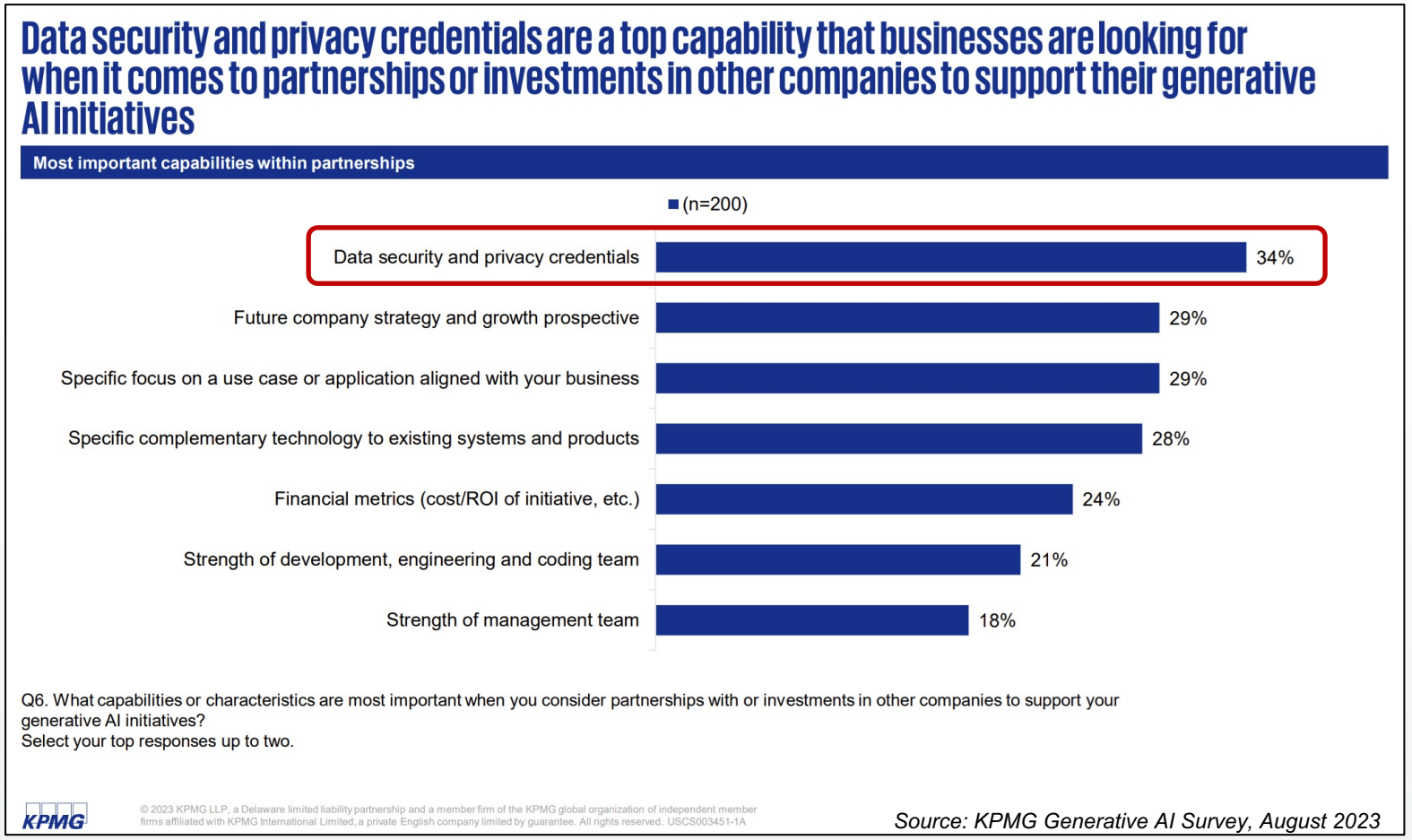
Security is a Blocker for GenAI

- Cybersecurity risk is a well recognized problem for generative AI
- Security is the top risk that organization are looking to address
- Lack of effective technical controls to meet privacy and security compliance requirements inhibits GenAI adoption especially when private data is used



Source: McKinsey State of AI in 2023

Security is an Enabler for GenAI



Source: KPMG Generative AI Survey, August 2023

- Businesses looking for evidence of strong data security and privacy capabilities in GenAI partners
- Successful GenAI applications and providers must go to market with strong data security and privacy assurances

Compliance Requires Provable Access Control

- Core to all data privacy and security regulations is requirement that access to sensitive data **must be** controlled and auditors check for proof of this control
- Traditional information systems rely on some variation of an access control list (ACL) or capability token tied to an **object** to define, enforce, and prove access controls
- Current generative AI systems cannot implement the required access control because there are no such objects

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."

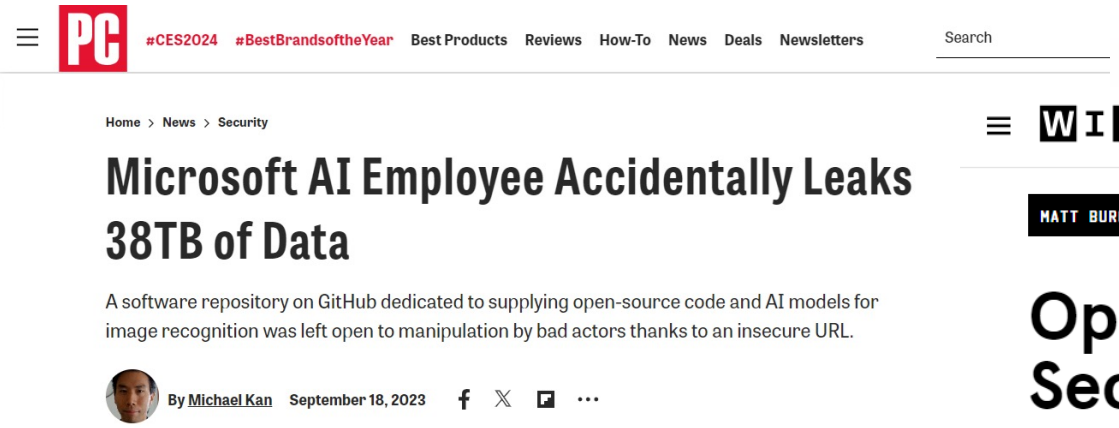
HIPAA

"(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

"The controller and processor shall take steps to ensure that any natural person acting under the authority of the controller or the processor who has access to personal data does not process them except on instructions from the controller..."

Consequences of Forgoing Access Control

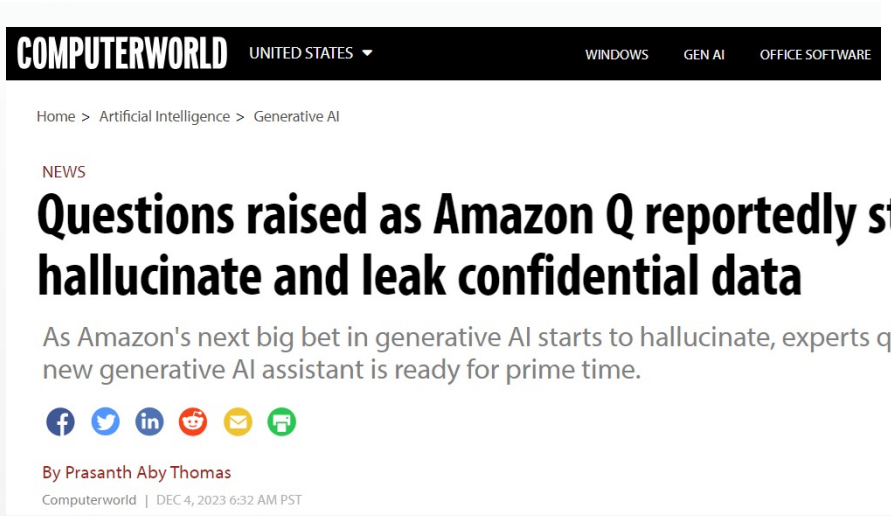

PC #CES2024 #BestBrandsoftheYear Best Products Reviews How-To News Deals Newsletters Search

Home > News > Security

Microsoft AI Employee Accidentally Leaks 38TB of Data

A software repository on GitHub dedicated to supplying open-source code and AI models for image recognition was left open to manipulation by bad actors thanks to an insecure URL.

By Michael Kan September 18, 2023


COMPUTERWORLD UNITED STATES WINDOWS GEN AI OFFICE SOFTWARE

Home > Artificial Intelligence > Generative AI

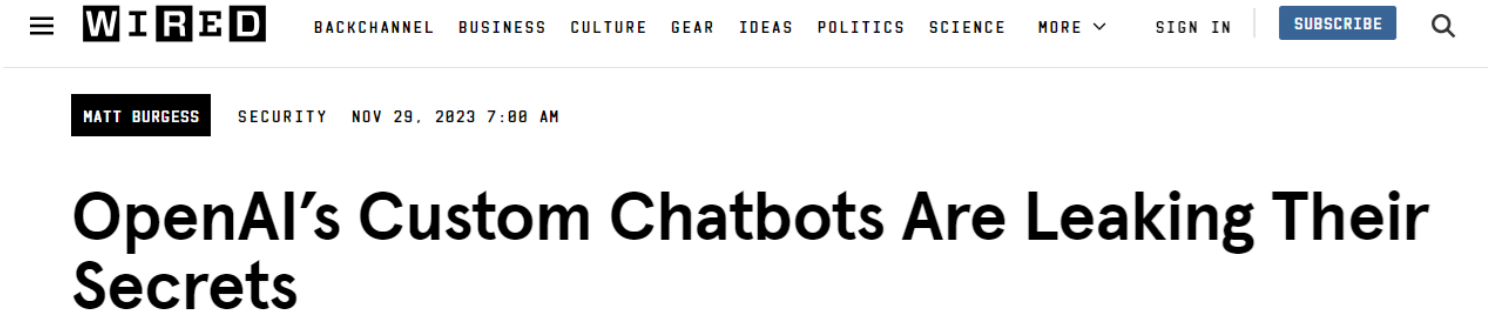
NEWS

Questions raised as Amazon Q reportedly starts to hallucinate and leak confidential data

As Amazon's next big bet in generative AI starts to hallucinate, experts question if the new generative AI assistant is ready for prime time.

By Prasanth Aby Thomas

Computerworld | DEC 4, 2023 6:32 AM PST


WIRED BACKCHANNEL BUSINESS CULTURE GEAR IDEAS POLITICS SCIENCE MORE SIGN IN SUBSCRIBE

MATT BURGESS SECURITY NOV 29, 2023 7:00 AM

OpenAI's Custom Chatbots Are Leaking Their Secrets

Released earlier this month, OpenAI's GPTs let anyone create custom chatbots. But some of the data they're built on is easily exposed.


DARKREADING NEWSLETTER SIGN-UP

Simple Hacking Technique Can Extract ChatGPT Training Data

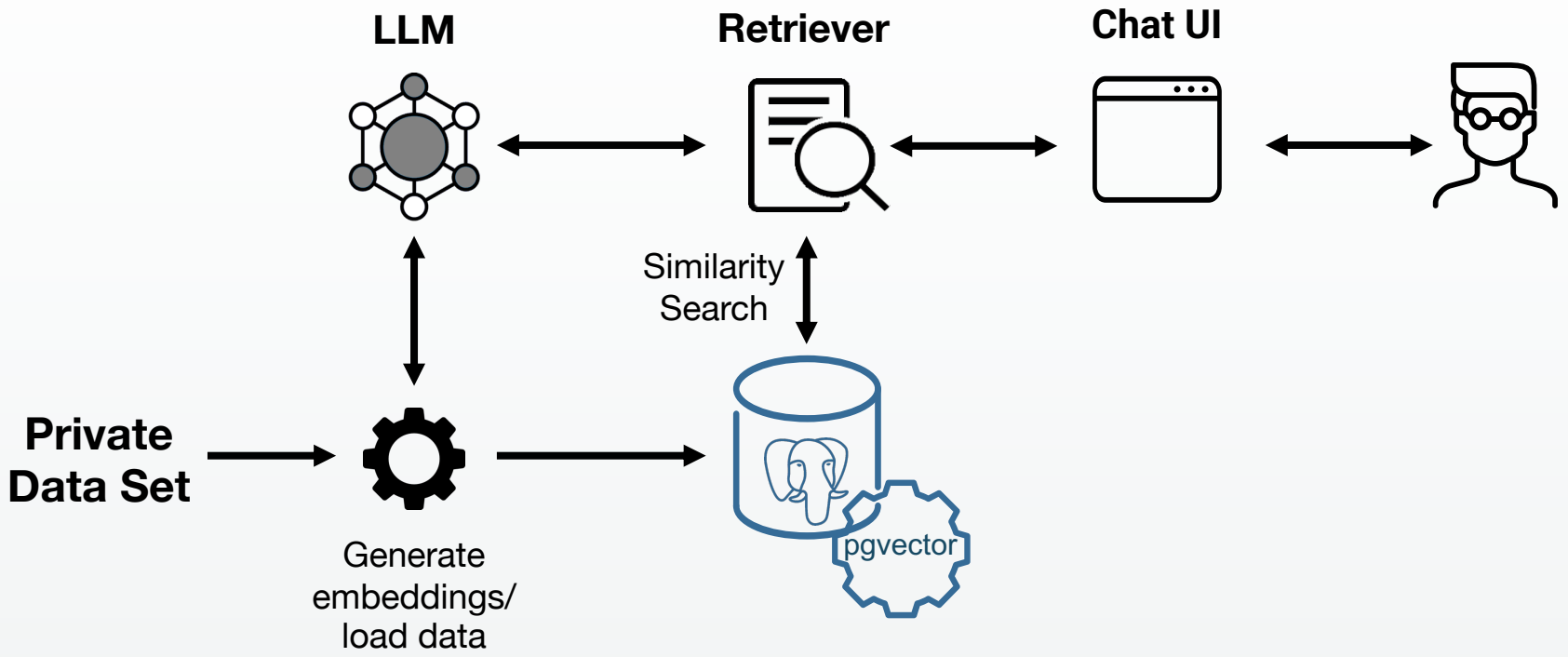
Apparently all it takes to get a chatbot to start spilling its secrets is prompting it to repeat certain words like "poem" forever.

Jai Vijayan, Contributing Writer

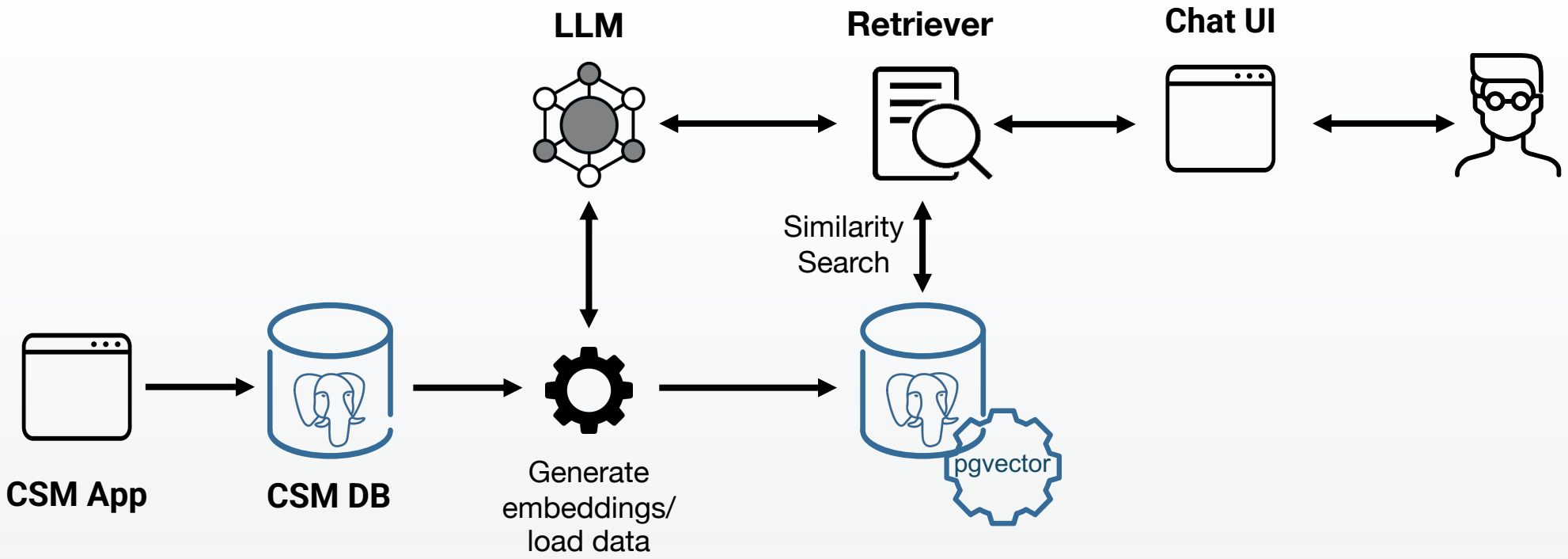
December 1, 2023

5 Min Read

What Does This Mean for Postgres?



Sensitive Data Enters Early in Data Pipeline



Hypothetical Source Data and Embeddings

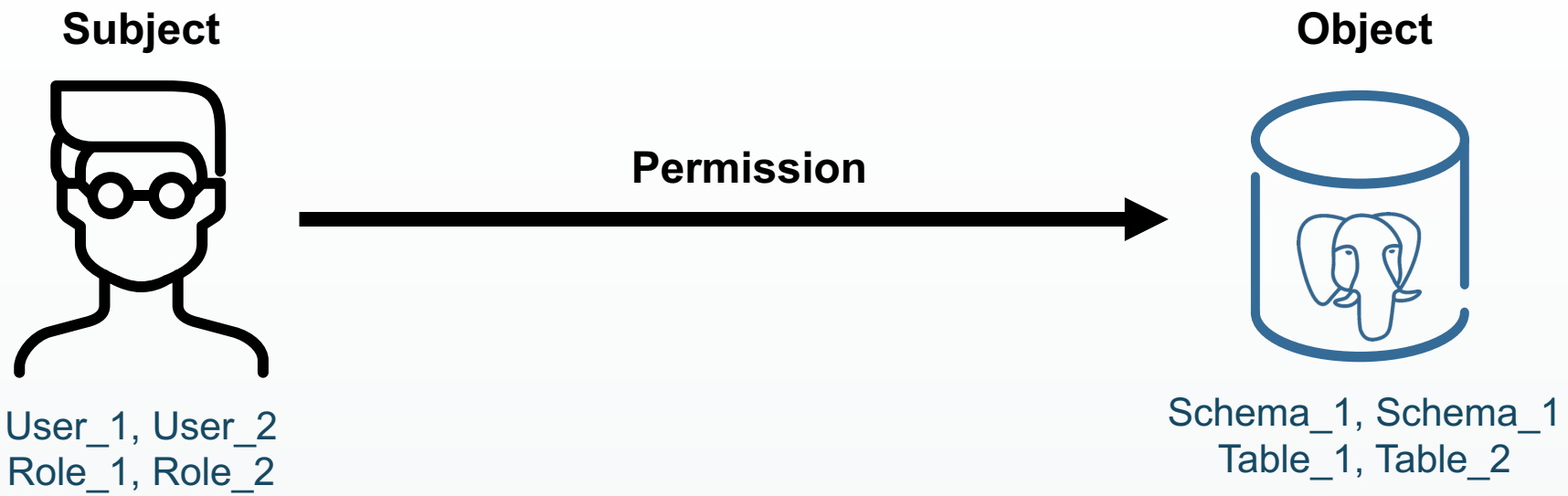
Support tickets Table

Ticket_ID	Email	Subject	Details
1006	jdoe@acme.com	Password reset	Hello, I can't seem to remember my..
1007	awier@acme.com	Can't access internal portal	To Support, The internal portal is..
1008	sdfh2@gmail.com	Urgent request from the CEO	Hi, This is the CEO, please wire $1M to..
1009	ssmit@acme.com	Changing my email address	Can you change my email address to…
1010	nira@acme.com	Forgot my password	Hi, I'm locked out of my account after 5..
1011	pfael@acme.com	Support website down	The support website looks like it's down..

Embeddings Table

Content	Tokens	Embedings
Email:jdoe@acme.com, Subject:Password reset Description:Hello, I can't seem to remember my password for..	539	[0.021440856158733368, 0.02200360782444477, -0..
Emal:awier@acme.com Subject:Can't access internal portal Description:To Support, The internal portal is not available. I'm getting..	753	[0.0245039766559492878, -0.000169642977416515, 0....
Email:sdfh2@gmail.com Subject:Urgent request from the CEO Description:Hi, This is the CEO, please wire $1M to the following account..	320	[0.03550934555492730, 0.047169963686414836, 0....
Email:ssmit@acme.com Subject:Changing my email address Description:Can you change my email address to rsmit@acme.com?	289	[0.011440856158733368, 0.00847360782495234, -0.0..
Email:nira@acme.com Subject:Forgot my password Description:Hi, I'm locked out of my account after 5 attempted logins..	134	[0.022517921403050423, -0.0019158280920237303,...
Email:pfael@acme.com Subject:Support website down Description:The support website looks like it's down, and I can't file a ticket..	431	[0.02050386555492878, 0.010169642977416515, 0....

Data Access Control In Postgres Database



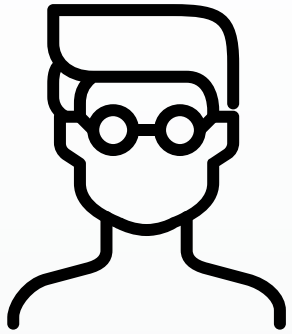
Access Policy:

```
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 in GenAI Application

Subject



User_1, User_2
Role_1, Role_2

Permission

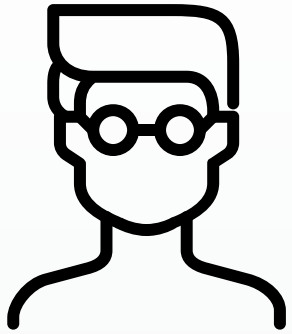
Object?

“The email address is jdoe@acme.com”

Access Policy:
???

The Solution is to Control Individual Data Values

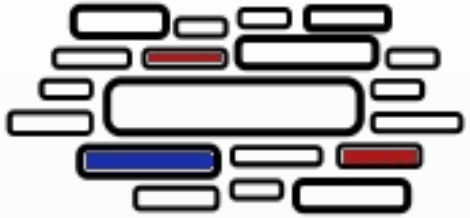
Subject



User_1, User_2
Role_1, Role_2

Permission

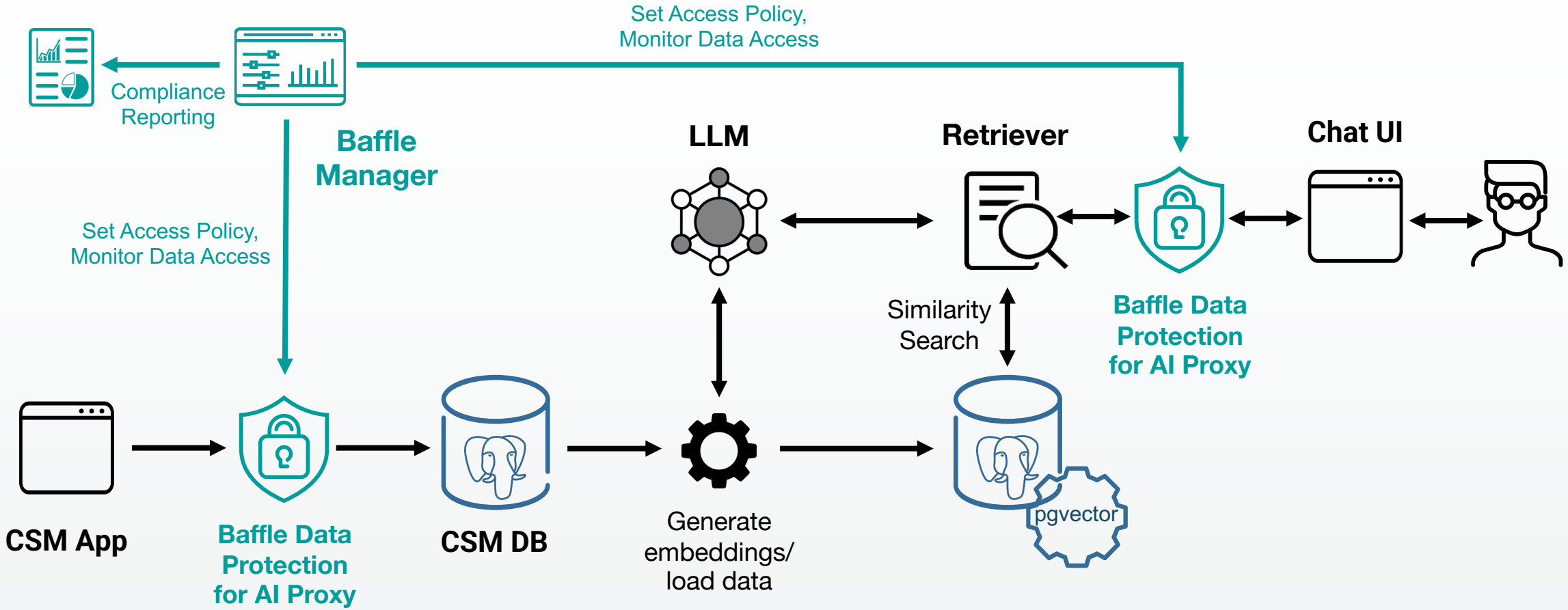
Object



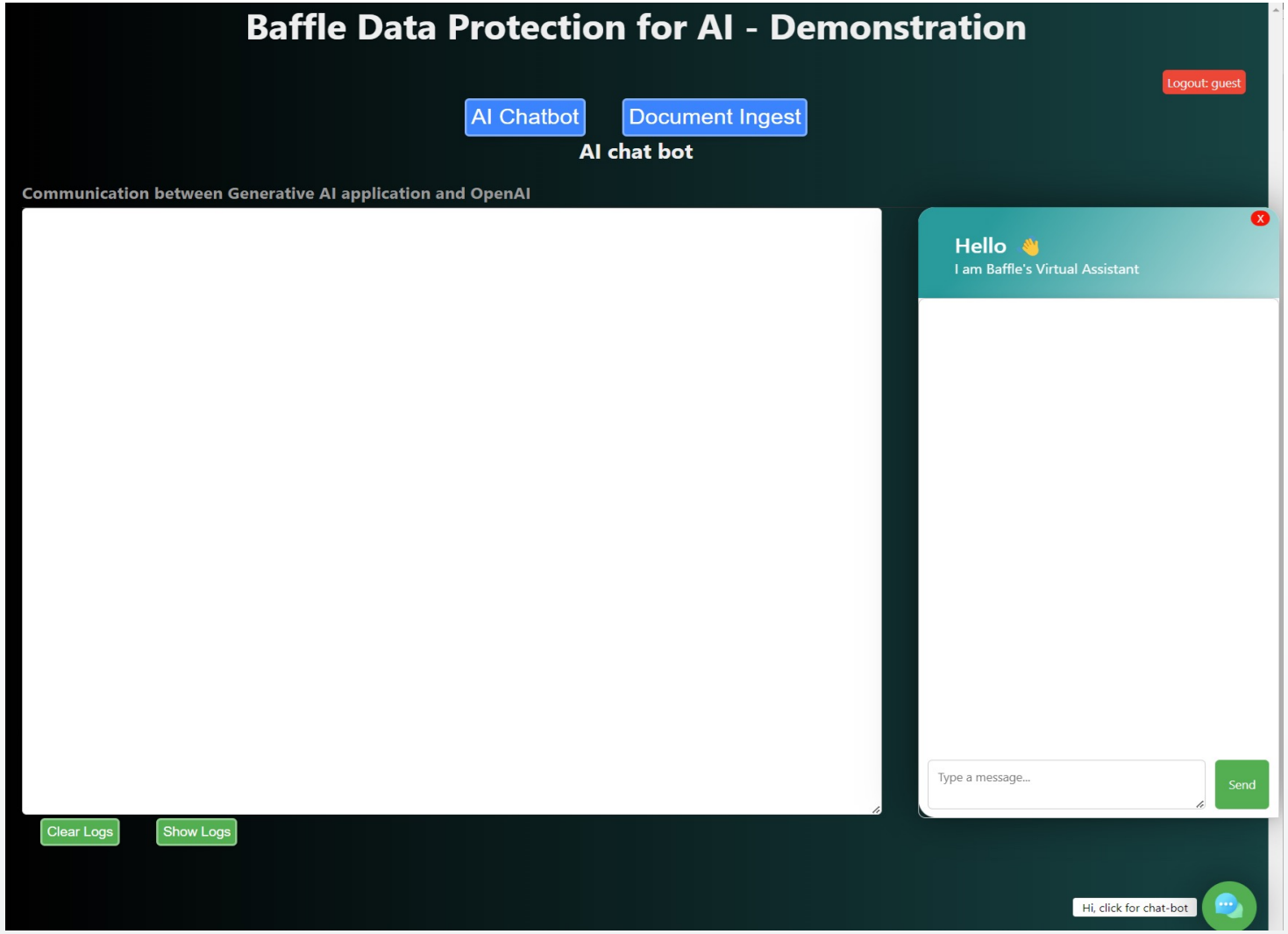
“The email address is **********”

Access Policy:
User_1 can decrypt and unmask “jdoe@acme.com”
Role_2 can decrypt and see partially masked “jdoe@acme.com”

How Do We Achieve This?



Demonstration



Key Takeaways

- Security is huge problem for broader GenAI usage, especially in context of data privacy compliance
- As a source data store Postgres plays a greater role in GenAI application pipelines than pgvector
- To ensure security and compliance data must be protected at the field level (i.e. per PII values)
- 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