CASE STUDY

CS-2026-11

ioSENTRIX

Your Trusted Partner in Cybersecurity



AI AND MACHINE LEARNING SECURITY

How a threat model mapped the AI risks in a multi-cloud platform at design stage

A five-phase threat model of a generative AI and machine learning platform at a large financial services company.

CLIENT
Large financial services company

SERVICE
AI/ML threat modelling

SCOPE
30+ components across AWS and Azure

REMEDIATION
All recommendations implemented

AT A GLANCE

CLIENT
Large financial services company

PLATFORM
Generative AI service with RAG pipelines, plus an ML model hub

SERVICE
AI security assessment

SCOPE
30+ components across AWS and Azure, and the data flows between them

METHOD
NIST SP 800-30, the NVIDIA AI Red Team framework and STRIDE

INPUT
Interviews with 8 stakeholders and the platform's documentation

REMEDIATION
All recommendations implemented by the client

DELIVERED BY
ioSENTRIX AI security team

EXECUTIVE SUMMARY

Prompt injection, missing access control in retrieval and open inference paths, found on paper first

A large financial services company had built an AI and machine learning platform for document processing, customer analytics and generative AI question answering. It spanned more than 30 components across AWS and Azure, and traditional assessments did not cover its AI-specific risks.

ioSENTRIX built a threat model of the whole platform in five phases. It identified more than ten threats, five of them High, including prompt injection, missing role-based access control in retrieval pipelines and unprotected inference paths, each with a fix. The client implemented every recommendation.

30+	5	10+
components mapped across two clouds	High-severity threats identified at design stage	threats, each with a specific fix

The models were new. The questions were old: who can ask, what can they reach, and who checks?

Details in this case study are anonymised and generalised. Methods that could harm other products are withheld.

© 2026 ioSENTRIX LLC

01 What we found

HIGH THREAT 1

Prompts could steer the AI into misusing its access

Both direct and indirect prompt injection were possible: instructions placed in a user's input, or hidden in a document the system retrieves, could change what the language model does.

Business impact: the platform could be pushed to reveal or act on sensitive financial data and customer information.

HIGH THREAT 2

Retrieval did not check who was asking

Role-based access control was not enforced in the retrieval-augmented generation (RAG) pipelines, and some ML inference paths were unprotected. The model also found data exposure risks in logging, storage and document uploads.

Business impact: a user could receive answers drawn from documents their role should never see.

Figure 1. Two threat scenarios the model ranked High



HOW WE FOUND IT

Our team interviewed the people who built and ran the platform, mapped every component, trust zone and data flow across both clouds, then worked through how an attacker would use each one.

© 2026 ioSENTRIX LLC

02 Results

Figure 2. Threats by severity (10+ threats)[1]



Figure 3. What the client implemented

All recommendations implemented by the client

AREA	THREAT FOUND	WHAT THE CLIENT IMPLEMENTED
Prompt injection	Direct and indirect injection could steer the model	• Defences against direct and indirect injection
Retrieval access	No role-based access control in RAG pipelines	• Access control for retrieval and inference paths
Data handling	Exposure in logging, storage and document uploads	• Fixes for logging, storage and uploads

03 Lessons for teams building on AI

AI needs its own threat model

Prompt injection ranked High, yet earlier assessments did not look for it. Treat AI as a new attack surface.

Threat 1 · OWASP Top 10 for LLM Applications (2025), LLM01: Prompt Injection

Enforce access at retrieval

Without RBAC in the RAG pipeline, the model returned what the pipeline handed it, whoever was asking.

Threat 2 · OWASP Top 10 for LLM Applications (2025), LLM08: Vector and Embedding Weaknesses

Map across cloud boundaries

Across 30+ components in AWS and Azure, trust broke where data crossed providers.

Approach

[1] Severity rated on a five-level likelihood and impact matrix based on NIST SP 800-30 Rev. 1. Levels shown: High, Moderate, Low and Very Low.

ioSENTRIX

Building on generative AI? See a sample AI threat model or book a strategy session before your platform scales.

iosentrix.com/contact-us

contact@iosentrix.com

+1 (888) 958-0554

13800 Coppermine Rd, Suite 190
Herndon, VA 20171

CREST-accredited · ISO/IEC 27001 certified · SOC 2 Type 2 attested

© 2026 ioSENTRIX LLC. General information only. Client details anonymised.