



**DefenixGPT**  
**AI Firewall**

# Mitigate AI Risks with Visibility & Control

The enterprise bridge to safe, compliant, and highly productive Public AI adoption.

## The AI Imperative



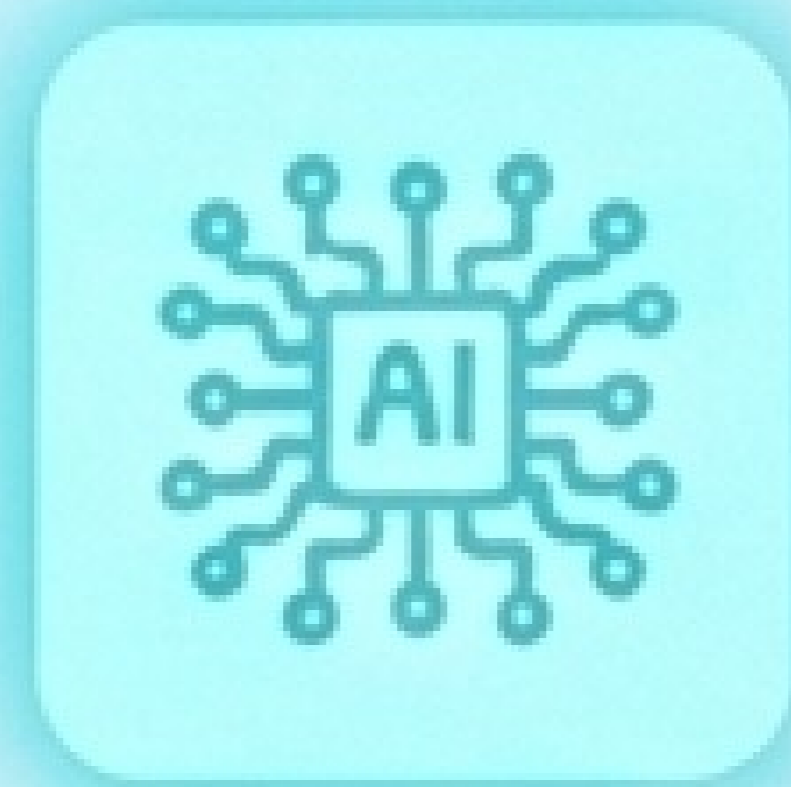
Gemini



ChatGPT



Copilot



Custom AI

Workforces demand access to Generative AI to maintain competitive productivity.

How do organisations leverage Public AI services **without sacrificing data sovereignty?**

## The Regulatory Reality



Exposing sensitive corporate data, PII, and trade secrets to public LLMs violates HIPAA, GDPR, and core infosec policies.



## Unrestricted Access

**Productivity:** High

**Shadow AI Risk:**  
**Critical**

**Regulatory Compliance:**  
**Failed**

**Outcome:** Immediate exposure to data exfiltration and OWASP LLM vulnerabilities.



## Complete Block / Ban

**Productivity:** Zero

**Shadow AI Risk:** High  
(Employees bypass controls via personal devices)

**Regulatory Compliance:**  
**Passed** (Technically)

**Outcome:** Loss of competitive edge, high workforce frustration.



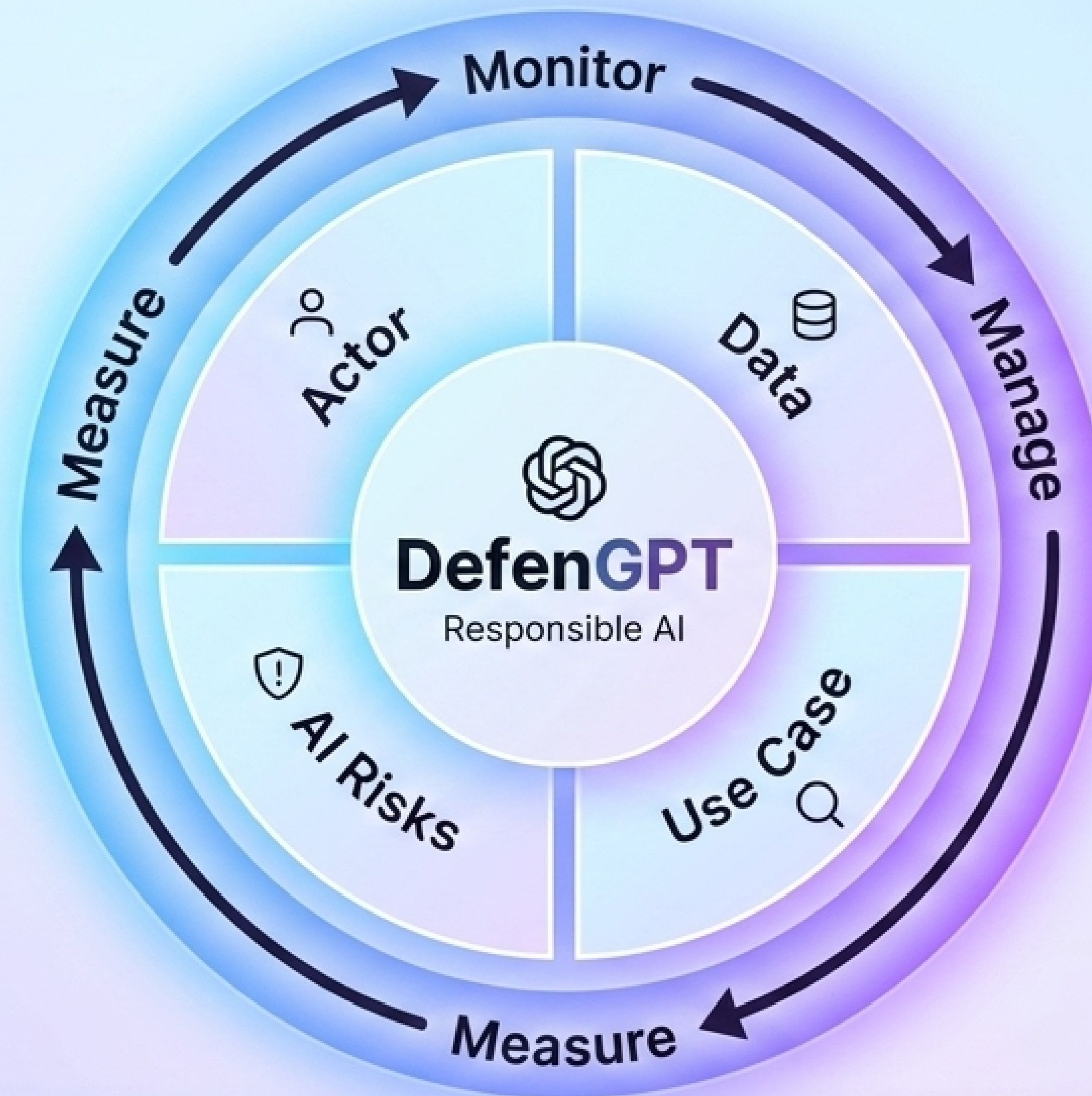
## DefenGPT Governed Access

**Productivity:** **High**  
(Frictionless access to ChatGPT, Gemini, Copilot)

**Shadow AI Risk:** **Neutralised**  
(Full visibility)

**Regulatory Compliance:**  
**Verified** (Real-time PII redaction and policy enforcement)

**Outcome:** Unmatched resilience and safe innovation.



DefenGPT acts as an **intelligent perimeter**—sitting **seamlessly between your workforce** and Cloud AI Services. It ensures every prompt and response is **measured against your unique organisational policies** before leaving your network.

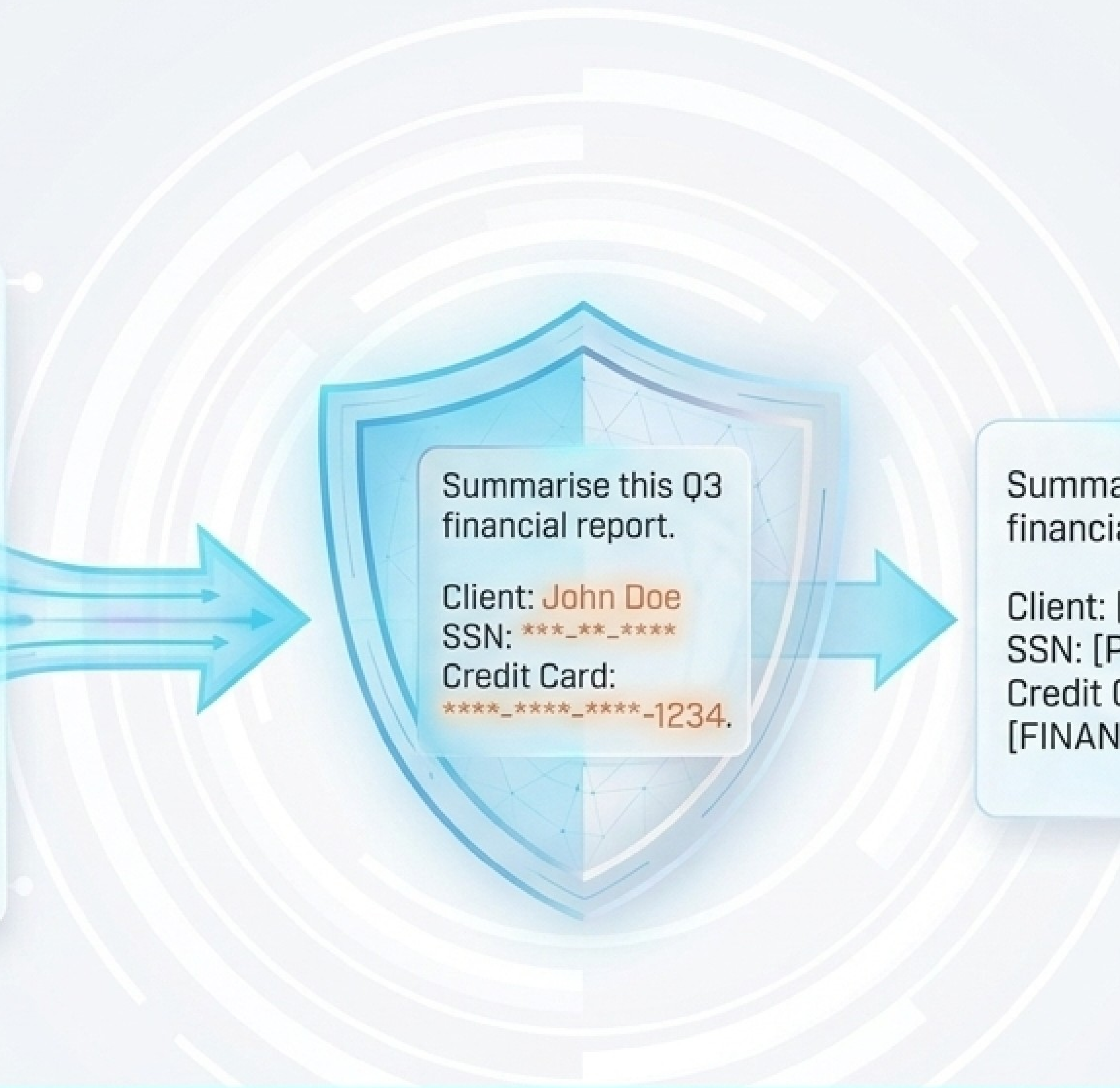
User chat



Summarise this Q3 financial report.

Client: John Doe  
SSN: \*\*\*\_\*\*\_\*\*\*\*  
Credit Card: \*\*\*\*\_\*\*\*\*\_\*\*\*\*-1234.

User chat



Summarise this Q3 financial report.

Client: [CLIENT\_NAME]  
SSN: [PII\_REDACTED]  
Credit Card: [FINANCE\_REDACTED].



Real-time, rule-based interception ensures PII, HIPAA-regulated data, and Prompt Injection attacks never reach the public cloud.

## Enterprise Governance

Monitor usage, enforce responsible AI, and maintain absolute compliance.

## Fortified Security

Prioritise data privacy with classification, sensitivity control, and threat mitigation.

A unified platform empowering regulated industries to control how Public AI is utilised, without inhibiting productivity.

## Complete AI Oversight

### Business Safeguarding

Proactively prevent  
trats and meoffal  
and procidors

### Advanced Risk Rules

Set granular, custom  
maranmentous and  
risk deficienty

### AI Governance Policies

Enforce responsible AI  
to elements AI and  
government

### AI Monitoring

Track and analyse  
to eam monitel and  
monitoring

### Data Taxonomy

Automatically  
categorise and  
insenge data

### Shadow AI Discovery

Understand exactly  
how to exctand  
optimizations

# INTELLIGENT PERIMETER: STRATIFIED SECURITY LAYERS

## Layer 1: Data Classification

Automatically classify user input data for immediate risk analysis before processing.

## Layer 3: Rule-based Enforcement

Execute rigid, automated rules to mitigate identified AI risks on the fly.

## Layer 5: OWASP LLM Top 10

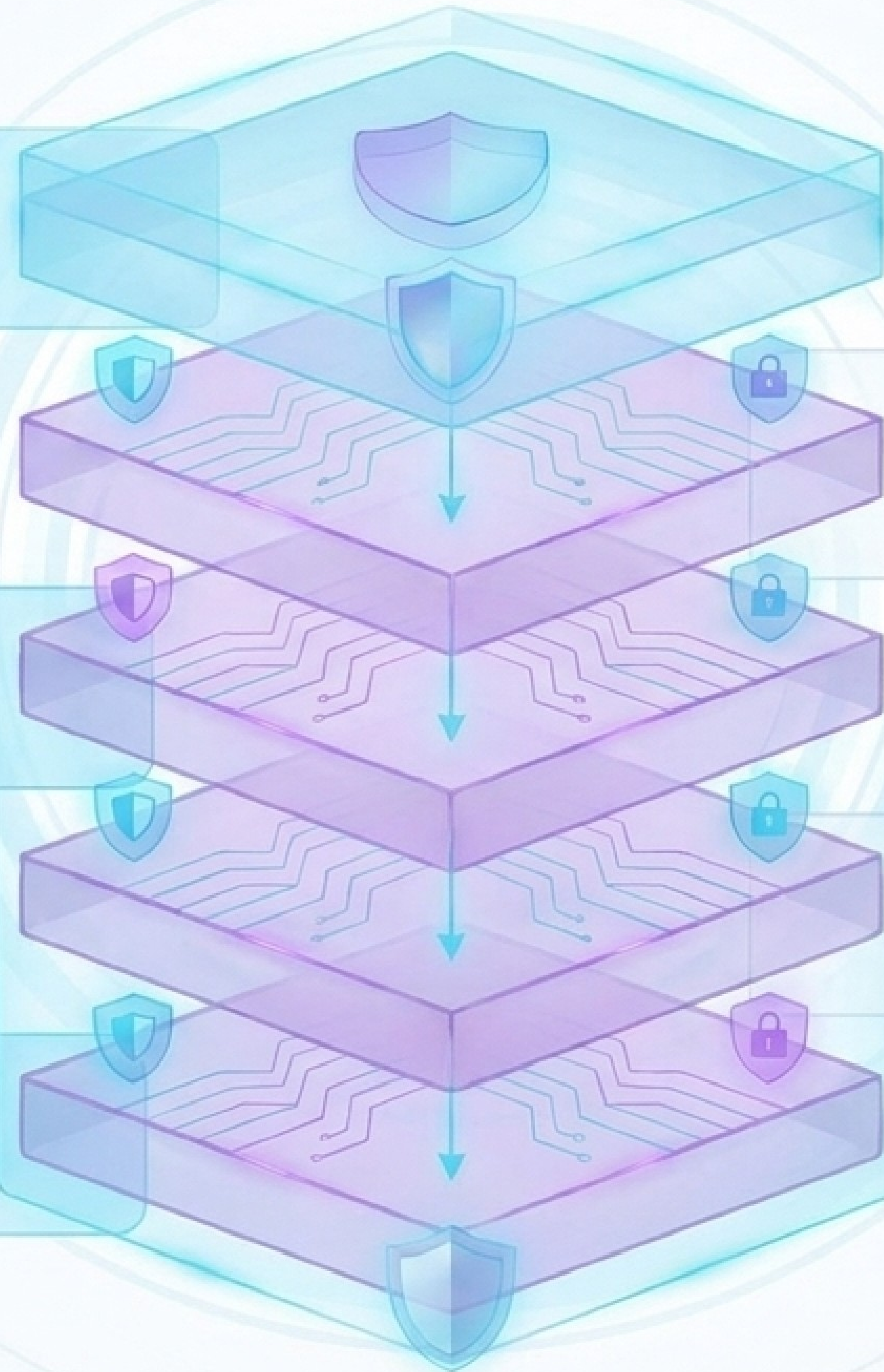
Address and resolve the most critical vulnerabilities standard to Large Language Models.

## Layer 2: Data Sensitivity & Protection

Actively manage and prevent the exposure of sensitive corporate data, including strict PII and HIPAA compliance.

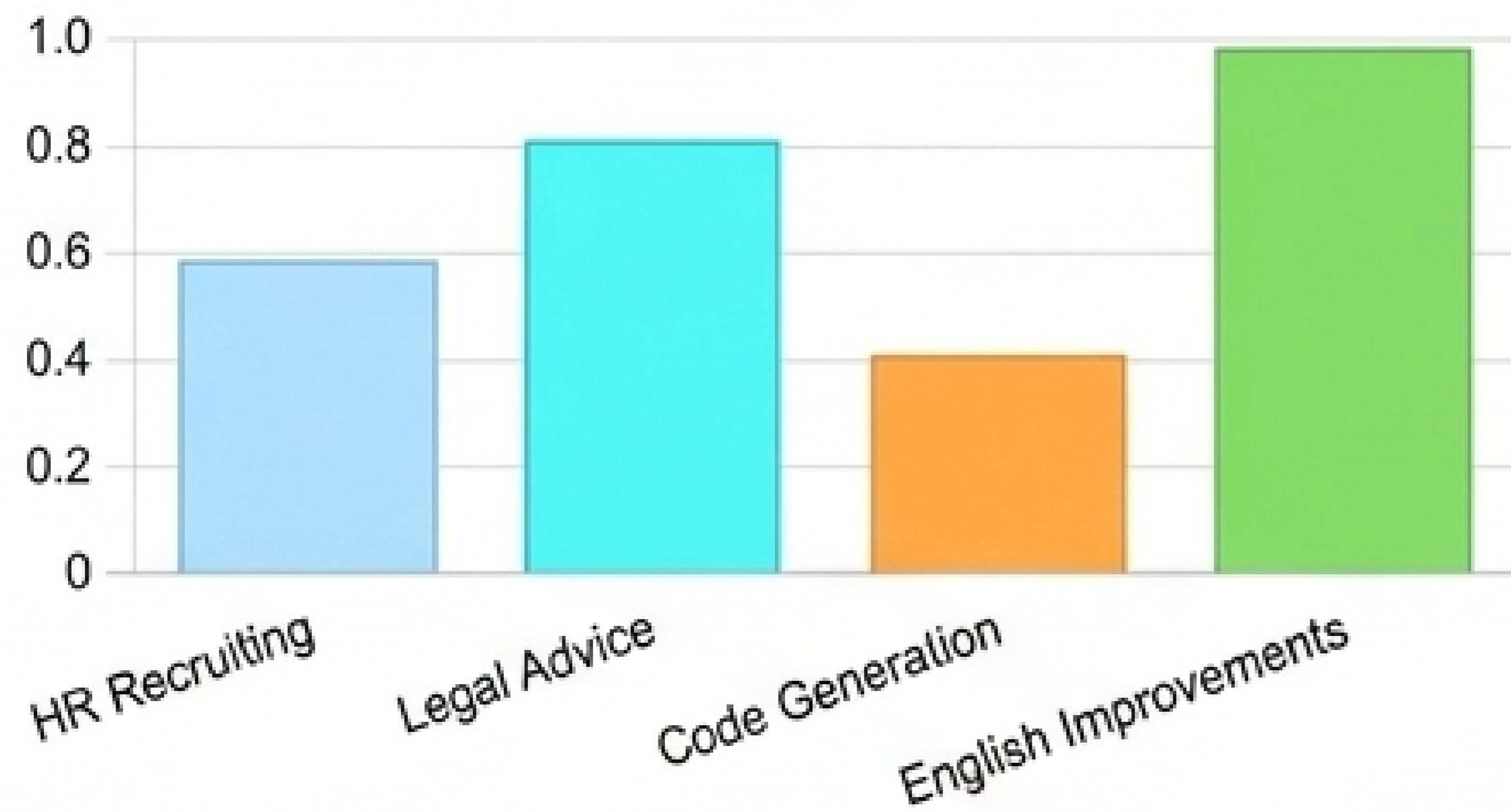
## Layer 4: Input Validation

Intercept and neutralise hostile threats, including sophisticated prompt injection attacks.



# AI Firewall at Work

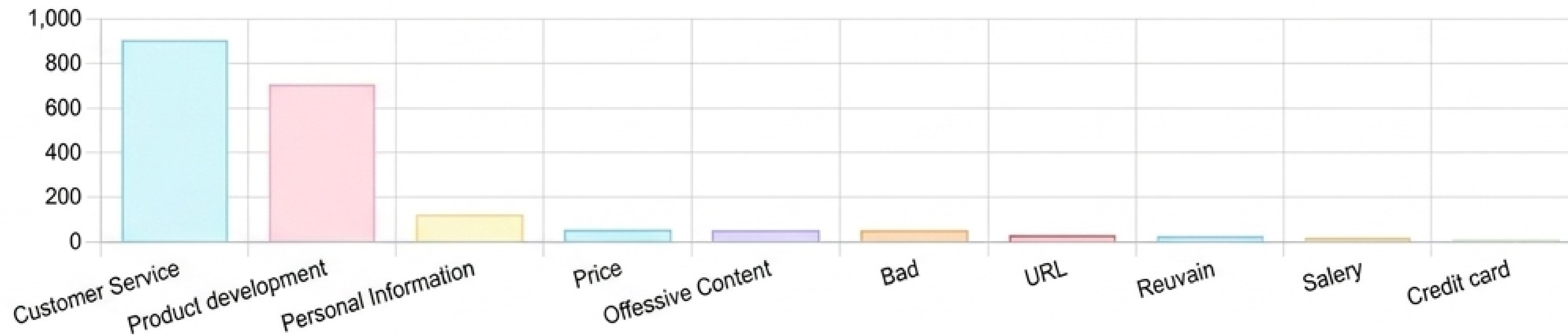
### Activity Usage



### AI Activity Risk By Level



### Top data classification rule



Abstract control translated into actionable oversight. Your entire AI landscape visualised in real-time.

## Account Activity Auditing

[Click here](#) to learn more about Activity Auditing

Advanced search

**Granular Search:** Filter by user, date range, risk level, or specific data classification rules.

Name or user email

Enter user name or email

Prompt

nda

Content name

Enter content name

From Date

45 — 7777

Response

Search in response

Risk

Select risk

To Date

45 — 7777

Topics

Enter topic

Usage Classification Rules

Use usage classification

Chat Name

Enter chat name

Detailed usage

Enter detailed usage

Data Classification Rules

Data data classification

Search

Clean/Reload

REFRESH

+ REANALYZE

FILTER RTAT

Policy Risk	Sensitivity level	Action	Policy Name	Chat Name	Prompt	User Name	Created Time
<input type="checkbox"/> Low	<input checked="" type="radio"/> Low	<input type="radio"/> Monitored	Code writing	Outlook Keyboard Shortcuts Guide	Switch to mail and switch to calendar. Teg...	AOV-Virtual Assistant	30/04/2024 13:01:58
<input type="checkbox"/> Uigh	<input type="radio"/> High	<input type="radio"/> Monitored	Legal advice for all	NDA - Generate a summary for NDA...	Generate a summary for NDA	Your Combie	20/04/2024 12:21:17
<input checked="" type="checkbox"/> Mgh	<input checked="" type="radio"/> High	<input type="radio"/> Monitored	Legal advice for all	NDA - coding to this NDA is copying ...	coding to this NDA is copying an idea a s...	Your Combie	11/04/2024 15:38:38

**Complete Prompt & Response Logging:** Retain a full history of exact user prompts and LLM outputs for compliance audits.

**Violation Tracking:** Instantly see exactly which governance policy was triggered, the sensitivity level, and the automated action taken.

Policy applied

High risk coily Legal advice for all

Prompt

- coding in the NDA is copying an idea a violation of the contract?
  - Data classification:  
The sentant did not match any data classification rules
  - Usage classification:  
A. Name: Legal Advice (AI)
  - Specific usage:  
Legal contact interpretation

Response

- According to the Non-Disclosure Agreement (NDA), andited in the document, cnpying an idea easid Italy is i nard l kсад a vicieller of the centrent. The sentriment definss "Coofidential Information" brnady is include verious forms of data, materials, and discloses disclosed by this Receiving Party to the Receiving Party. The Receiving Party is triggered to heflit and manking the Confidential Information in data confidence. For ose or regtriation of the ideas of Information eathon euthamiation would Italy mach the dsein in the RDOA, potantly, leselling to legal requieremeres.  
Data classification:  
A. Name: Personal Information (AI) Sensitivity Level: NGA

Data sources

- Document: NDA [🔗](#)  
Data classification:  
A. Name: Personal Information (AI) Sensitivity Level: NGA

User email

yazeik@gettachware.com

Chat topics

Violation of contract, Confidential Information, Copying on Ideas, Non-Disclosure Agreement

Firewall type

BusinessCPTAy.

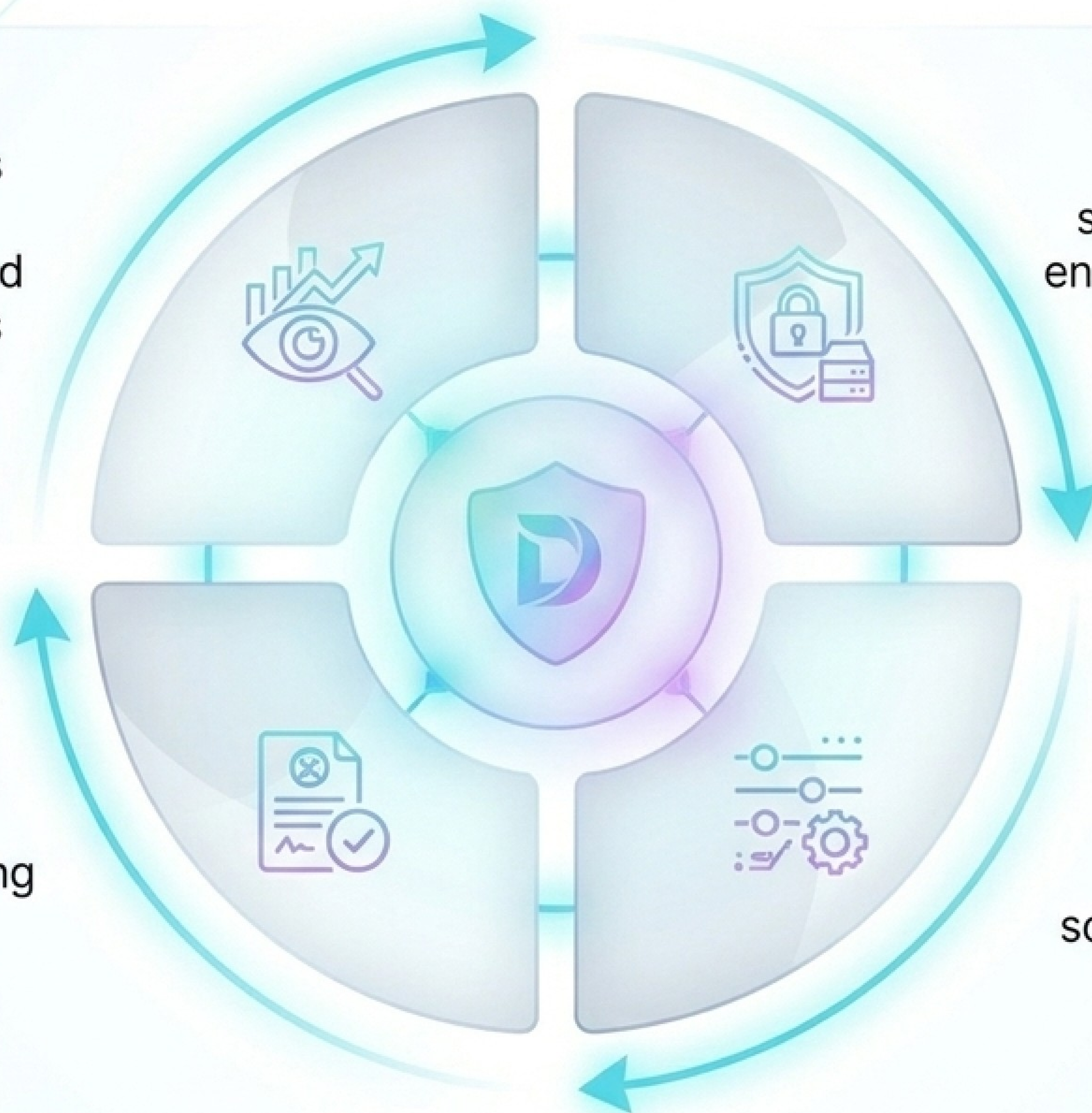
# Unmatched Resilience: The Defenix Advantage

**Increased Visibility:** Offers detailed tracking and reporting on AI adoption and usage, enabling businesses to optimise AI strategies.

**Enhanced Compliance:** Mitigates risks of non-compliance by providing AI governance tailored to regulatory standards.

**Data Privacy:** Protects sensitive data from exposure, ensuring that AI usage adheres to company policies and security protocols.

**Risk Management:** Allows for granular policy enforcement, making the solution adaptable to different departments or user needs.



# Try DefenGPT AI Firewall for free today.



Don't choose between AI innovation and enterprise security. Deploy the intelligent perimeter.



[Sales@Defenix.AI](mailto:Sales@Defenix.AI)

Website: [www.defenix.ai](http://www.defenix.ai)