解碼AI網路安全未來戰場

洞察趨勢．理解現況．實踐風控

Bruce Lan
Solution Consultant

填問卷拿好禮三步驟，加碼再抽AirPod4

Step.1
掃描QRCode 填寫問卷

Step.2截圖問卷完成畫面立即送（7F 701A）
搶先兌換聯名咖啡包

Step.3
至 Palo Alto Networks 攤位**C111**兌換精美好禮

員工使用生成式 AI 的相關應用

AI Access Security

保護GenAI使用安全

資料搜尋．上下文分析．文件生成．圖像生成．
網站生成．程式碼．腳本生成

企業建立 AI 相關應用

製程優化．品質管理．生產規劃．藥物研發．醫
療診斷．風險管理．遊戲開發．金融服務．新聞
編輯．自駕訓練．產險審核．資源配置

paloalto NETWORKS

AI Access Security

未知的 - 服務應用 · 數據流動 · 隱藏風險

Employees

AI Access enables safe AI adoption

- 影子人工智慧應用造成管理盲點
- 未經檢查的敏感資料外泄
- 嵌入在回應中的安全風險

- ChatGPT
- Copysmith
- Grammarly
- SageMaker
- Vertex AI
- Adobe Firefly
- Hugging Face
- Codeium
- Coveo
- Hypotenuse
- Elai
- Perplexity AI
- Lightning AI
- WandB
- Replicate
- NLP Cloud
- Sapling AI
- Swimm AI

未檢測到的 GenAI 插件可能會導致未經授權的資料訪問

Interconnected SaaS ecosystems with **AI-powered third-party integrations** create **security vulnerabilities**, **complicating detection and control**.

This leads to **increased risk of security breaches and non-compliance** with regulatory requirements.

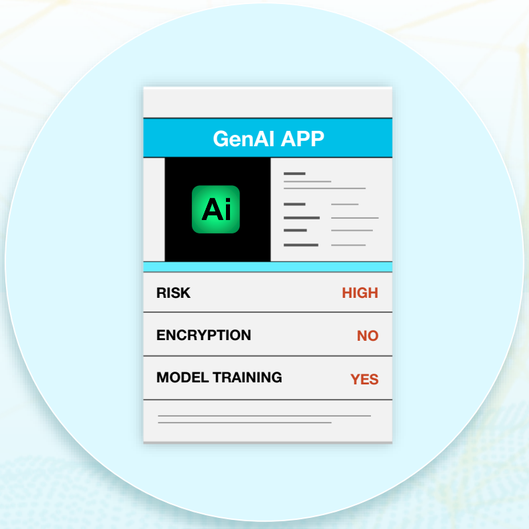
克服挑戰需要擁有的能力...

Challenges

Democratization of AI app creation has led to sprawl of insecure apps.

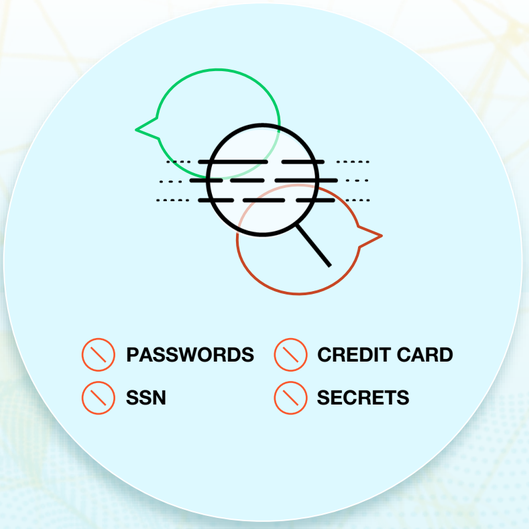
持續自動提供完整的應用可視性與自動應用分類

AI apps store, learn, and reiterate data.



基於數據進行訓練的AI應用程式的行為分析

AI apps ingest unstructured inputs and generate diverse outputs.



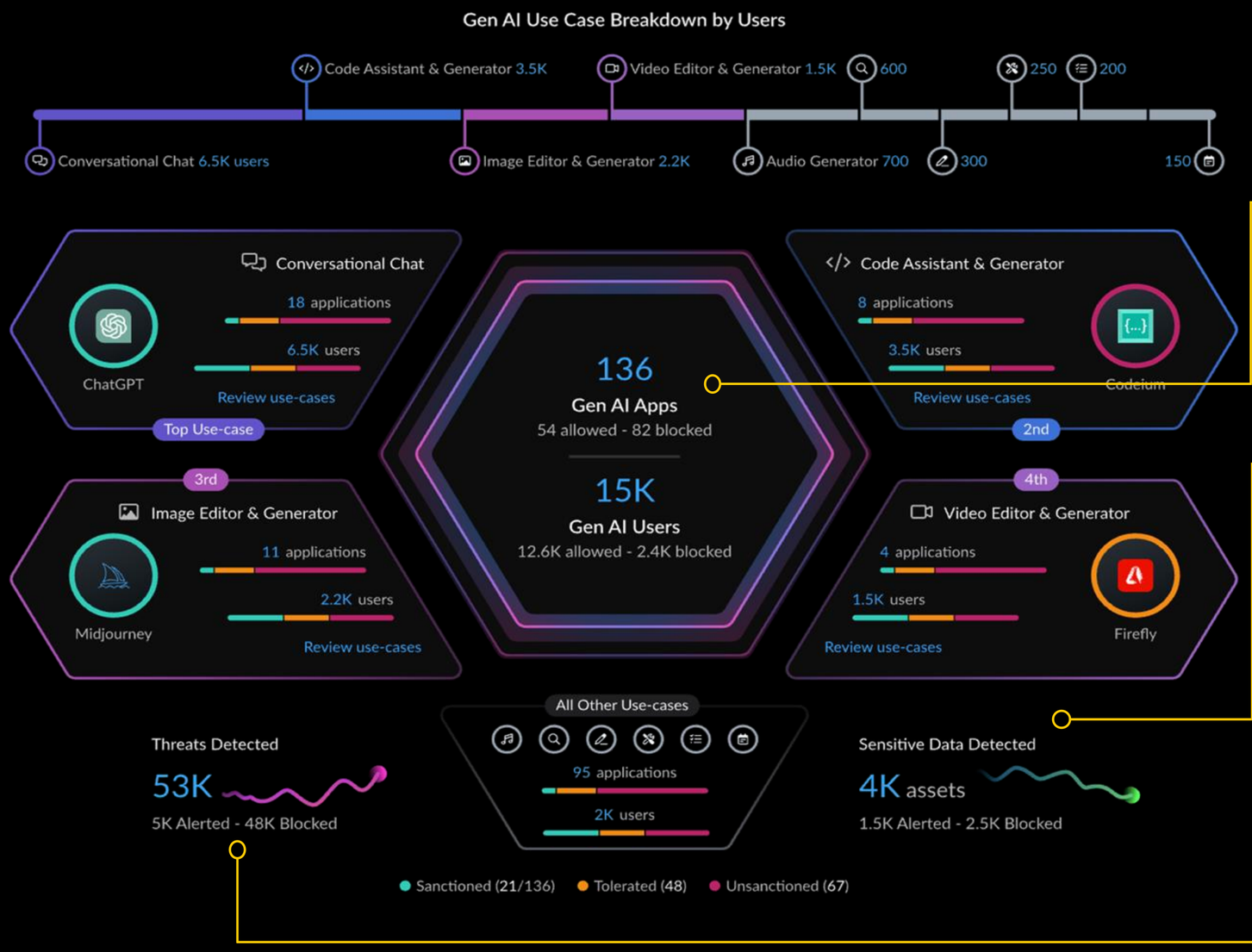
基於上下文資料的分析與威脅的偵測

Rapidly evolving, powerful apps are readily accessible through marketplaces.

第三方外掛、插件的分析與權限控制

Solutions

可視性為風險管理成敗的重要基礎



即時完整的 AI 使用可視性

檢視不同背景的員工使用了哪些 AI 應用！

完整的資料保護

掃描資料，帳密以及哪些智慧財產 IP 被分享了

彈指之間進行存取控制

攔截未經允許的應用程式，指派資安政策以及進行**資料保護、威脅過濾、合規控制...**

員工使用生成式 AI 的相關應用

資料搜尋．上下文分析．文件生成．圖像生成．
網站生成．程式碼．腳本生成

企業建立 AI 相關應用

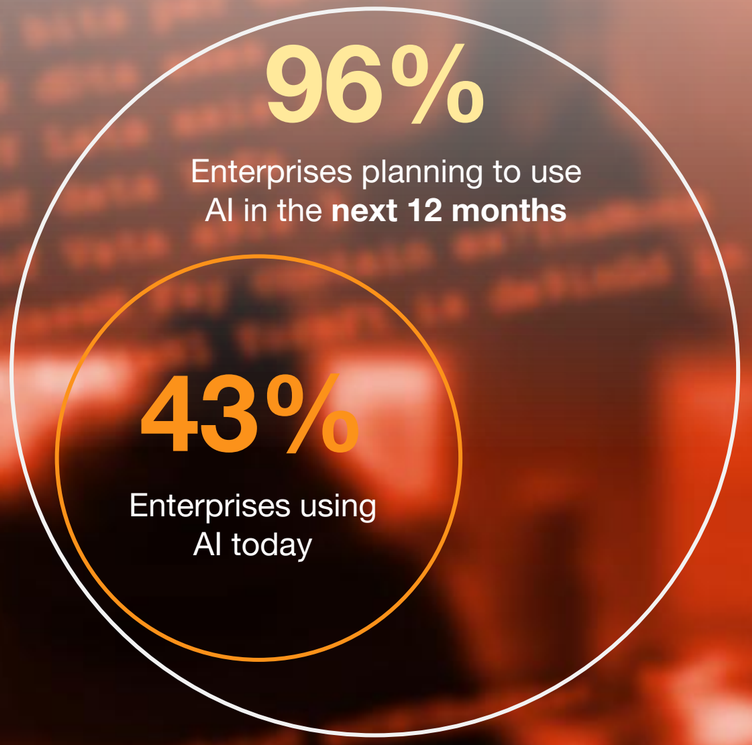
AI Runtime Security

保護AI應用系統開發與營運安全

製程優化．品質管理．生產規劃．藥物研發．醫
療診斷．風險管理．遊戲開發．金融服務．新聞
編輯．自駕訓練．產險審核．資源配置

AI App Proliferation Will Change How Enterprises Operate

應用將無處不在（趨勢）



須為全新的挑戰做好準備

TYPICAL LARGE ENTERPRISE

AI Applications

100s of sanctioned AI apps

AI Model Infrastructure

1000s of LLM, SLMs, ML scripts

AI Datasets

10s of PB of training & vector data

MILLIONS of prompts annually

現況離你的想像有多遠

50%
of employees using AI do it **without** permission

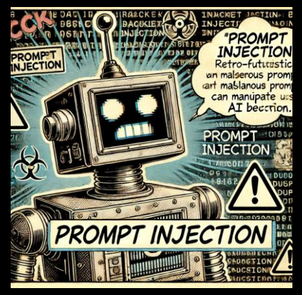
80%
of public models can be **jailbroken**

100+
malicious models available in the wild

paloalto NETWORKS | UNIT 42

從主要風險列舉的角度...

OWASP Top 10 for LLM Applications 2025

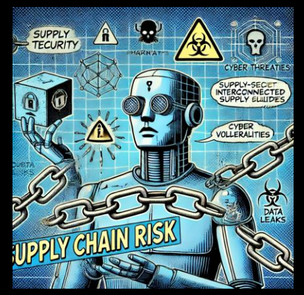


提示詞注入

AI 注入、詐騙、資料滲透、外掛程式請求偽造

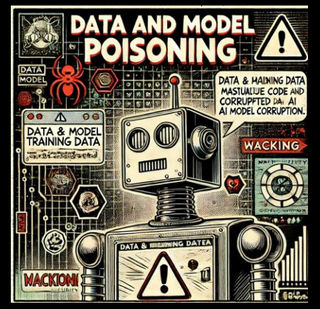
敏感資訊洩漏

LLM 在輸出中暴露機敏資料、專有演算法



供應鏈風險

LLM 供應鏈受到各種漏洞影響訓練數據、模型和部署平台的完整性



數據與模型中毒

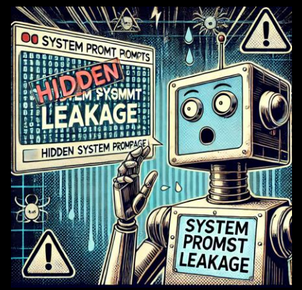
存在於訓練、微調或嵌入數據被操縱的情況，引入漏洞、後門或偏見

不當輸出處理

生成內容傳遞到其他元件和系統之前，缺乏足夠的驗證、清理和處理，導致 XSS、CSRF、SSRF、權限提升或遠端程式碼執行

過度授權

當意外、模糊或被操縱的輸出而執行損害性行動的漏洞，通常是功能過度、權限過度或自主性過度

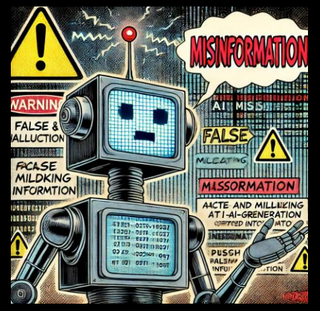


系統提示洩漏

引導模型行為的系統提示或指令可能包含非預期被發現的敏感資訊

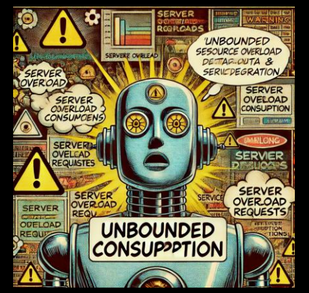
向量與嵌入弱點

在使用RAG的系統中，向量與嵌入的弱點可能被惡意行為利用，導致有害內容注入、模型輸出操縱或敏感資訊存取



錯誤資訊

生成看似準確但實際上是虛構的內容（幻覺），源於模型用統計模式填補訓練數據中的空白，而非真正理解內容

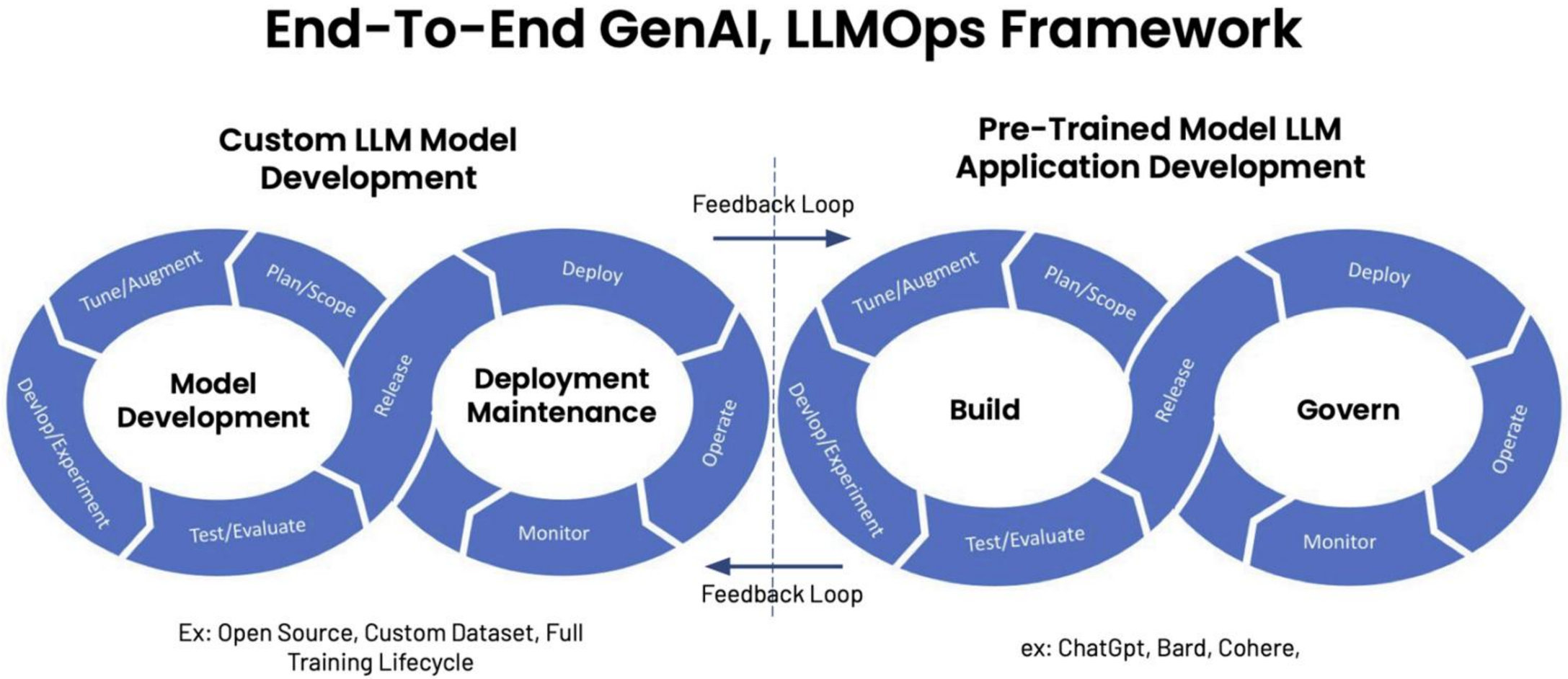


無限制消耗

允許使用者進行過度且無控制的推論，可能導致服務拒絕、經濟損失、模型盜用和服務降級等風險

從流程與安全管理框架的角度...

LLM and Gen AI App SecOps Framework



Material ,OWASP-LLM_GenAI-Security-Solutions-Reference-Guide-v1.0-1.pdf

OWASP | TOP 10 LLM APPLICATIONS & GENERATIVE AI

CHEAT SHEET

LLM & GenAI Security Landscape - 2025, Q1

https://genai.owasp.org/ai-security-solutions-landscape/

Open Source

Operate

aqua
securiti AI Security & Governance
Llama Guard Meta
ZenGuard AI
NRI SECURE AI Blue Team
Cisco AI Firewall®
LLM GUARD
BLUETEAM AI
aim security
PRISMA
TROJ.AI
paloalto

Deploy

TROJ.AI
Cisco AI Firewall®
BLUETEAM AI
Meta
paloalto
securiti AI Security & Governance
Operant
PRISMA

Release

securiti AI Security & Governance
Cisco AI Validation
CycloneDX
PRISMA
LEGIT AI-SPM
paloalto
aqua

Test & Evaluate

pillar SECURITY
Cisco AI Validation
Prompt Fuzzer
promptfoo
aqua
Enkrypt AI
Giskard
Synack
LangCheck
MODELSCAN
securiti AI Security & Governance
CYBERSECEVAL Meta
ZenGuard AI
LLM vulnerability scanner
HARMBENCH
Cobalt
PROTECT AI
WATCHTOWER
PRISMA
MINDGARD
TROJ.AI
Citadel AI
AIShield
Cymetrics

Develop & Experiment

Meta
Mend.io
TRAIL of BITS
aqua
Cisco AI Validation
snyk
securiti AI Security & Governance
pangea
Operant
PRIVACY Raven
Cloaked AI
TROJ.AI
PRISMA

Augment, Fine Tune Data

Cloaked AI
securiti AI Security & Governance
UNSTRUCTURED.IO
BLUETEAM AI
PRISMA

Scope & Plan

MITRE ATLAS
BLUETEAM AI
Seezo
paloalto
pillar SECURITY
securiti
PRISMA
STRIDE GPT

Monitor

aqua
Cisco AI Validation
aim security
BLUETEAM AI
AIShield
HIDDENLAYER
LAKERA
Meta Prompt Guard
Operant
paloalto
PROTECT AI
securiti
splx AI
TROJ.AI
ZenGuard AI

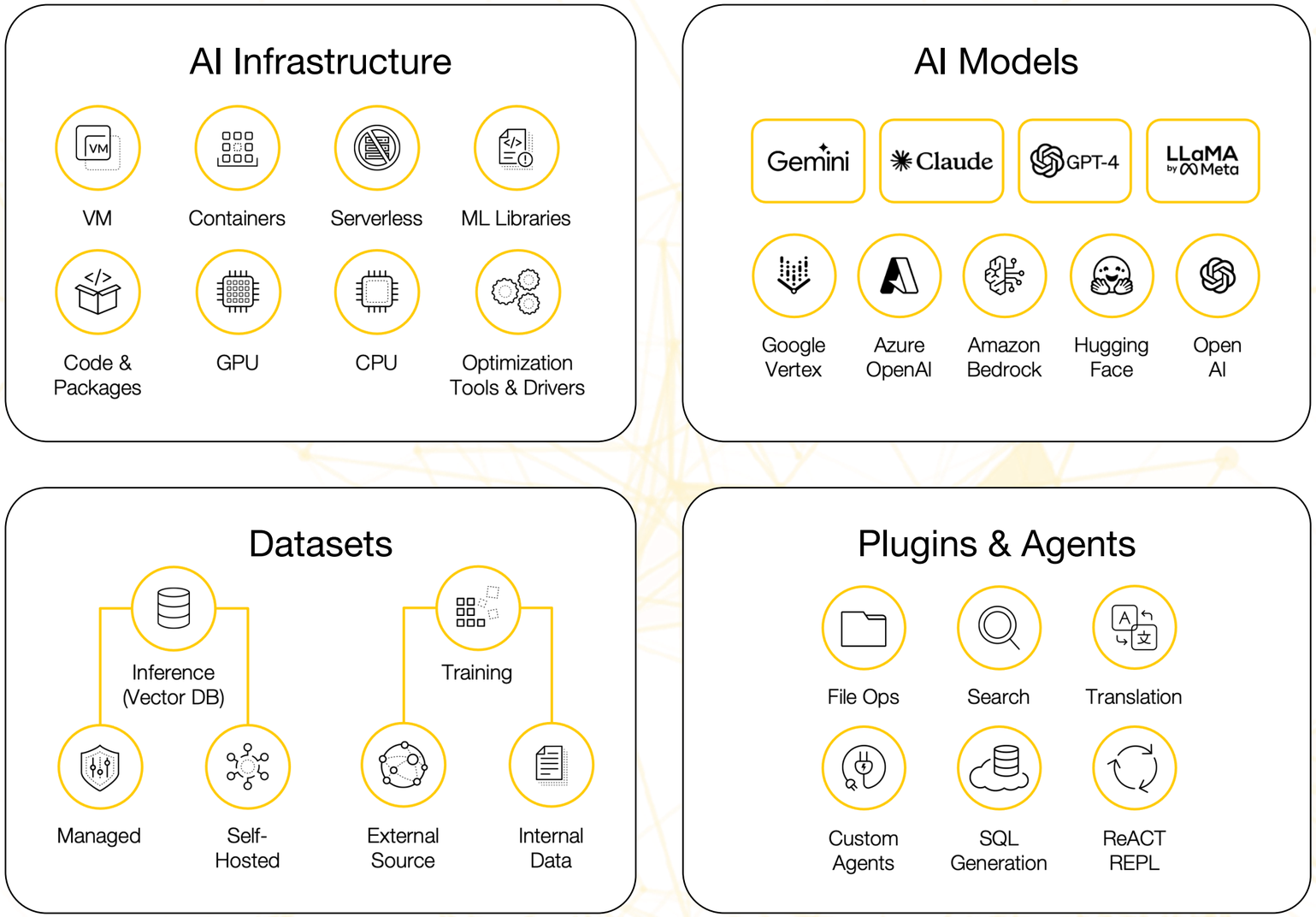
Govern

aim security
Cisco AI Validation
BLUETEAM AI
Lasso
paloalto
PRISMA
Prompt:
securiti
Tumeryk AI GUARD
unbound security
AI VERIFY FOUNDATION

Source; OWASP Gen AI Security Solutions Landscape Guide 2025. Q1

幾個需要保護的重要組成

Enterprise AI Apps Use an Extended Tech Stack . .



顯而易見的風險不容忽視

...Which Introduces New Runtime Risks

填問卷拿好禮
三步驟，加碼
再抽AirPod4

AI Infrastructure

VM Containers Serverless ML Libraries

Code & Packages GPU GPU Optimization Tools & Drivers

Insecure outputs

Remote code execution

Lack of segmentation

AI Models

Gemini GPT-4 LLaMA by Meta

Google Vertex Azure OpenAI Amazon Bedrock Hugging Face Open AI

Prompt injection

Model DoS attacks

Datasets

Inference (Vector DB) Training

Managed Self-Hosted External Source Data

Plugins & Agents

File Ops Search Translation

Custom Agents SQL Generation ReACT REPL

Unrestricted DB queries

Sensitive data leaks

Unrestricted communication

完整覆蓋服務運行時的安全防護

Any App. Any Threat. Any Cloud. One Solution.
AI Runtime Security

Comprehensive AI threat protection

- AI Model Protection
- AI Data Protection
- AI Application Protection

+

Precision AI security services

ADV TP · ADV WF · ADV URL · ADV DNS · DLP

+

Inline **Network** Intercept

OR

API Detection as Code (DaC)

Built-in protection for AI-specific threats

Support for **40+ popular AI models** across Google Cloud, Microsoft Azure, Amazon Web Services and OpenAI or any model via API

Cloud-Delivered Security Services (CDSS)

Industry's **most complete coverage** in a single, all-in-one portfolio.

Multiple form factors

Virtualized **ML-powered NGFWs** and API* deployment options allow best-in-class security with cloud speed, agility and scale.

*Including post-GA functionality.

paloalto NETWORKS | STRATA™

將保護融合其中的二個方式

Empower Your Security Teams

Network
security approach

- **AI-specific threat** detection & prevention
- **Centralized** security for **entire** network
- **Perimeter** & **East/West** protection
- **Comprehensive** container security

Use one or both approaches.

Configure policies once & apply consistently.

Developer
security approach

- AI Detection as Code (**DaC**)
- **User-specific** security policies
- Granular protection with custom error **behavior**
- **Simple** setup in minutes

保護組織的網路架構免受 AI 特定和傳統網路攻擊

API Intercept 通過將“安全即代碼”（Security-as-Code）嵌入到您的程式碼中

在 AI 工作流中嵌入安全檢查（API）

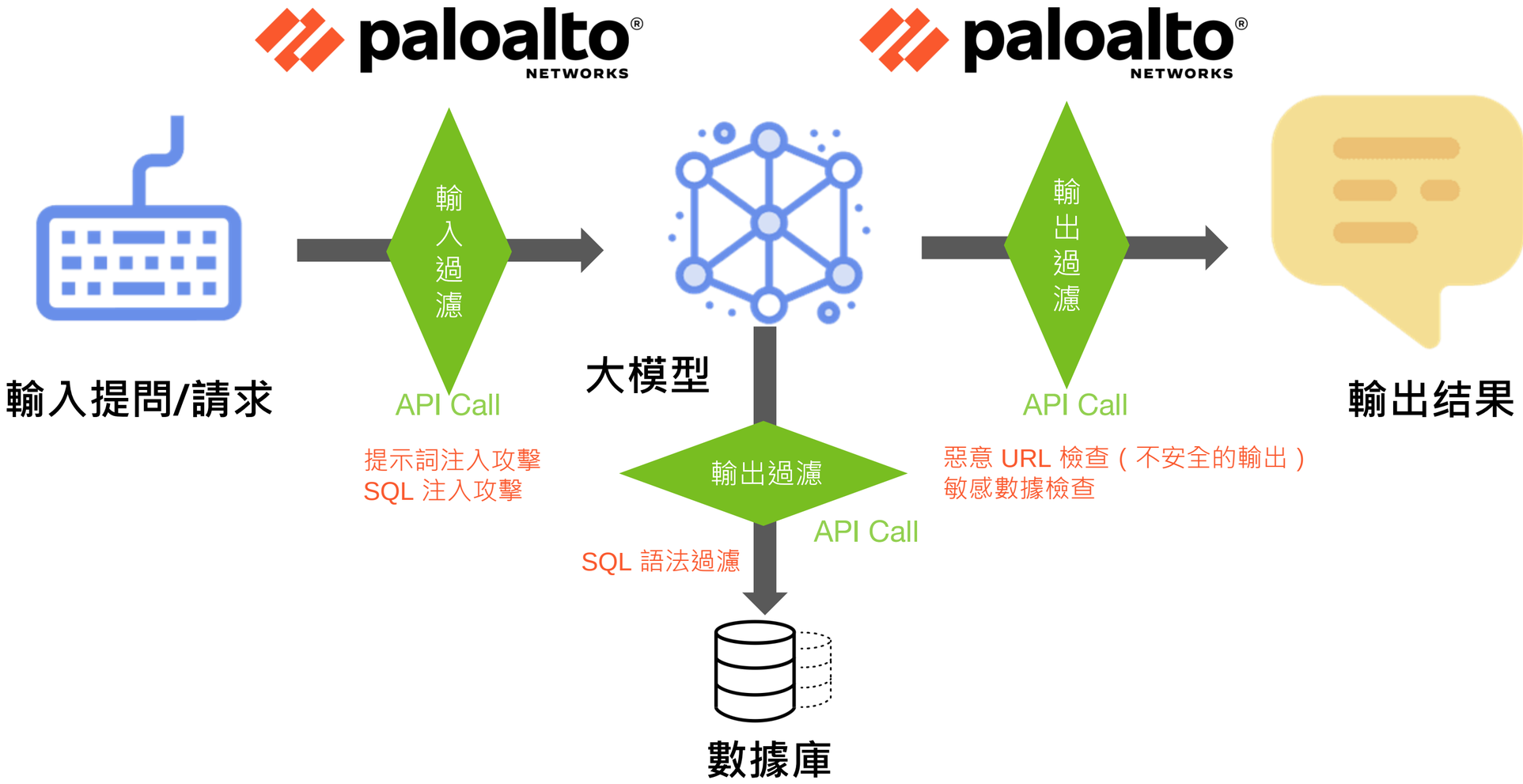
How can I protect my **AI apps, models, and datasets** from AI specific threats like **prompt injection attacks, insecure model outputs, and sensitive data loss**?

Detect and prevent threats at the network layer **without modifying app code**

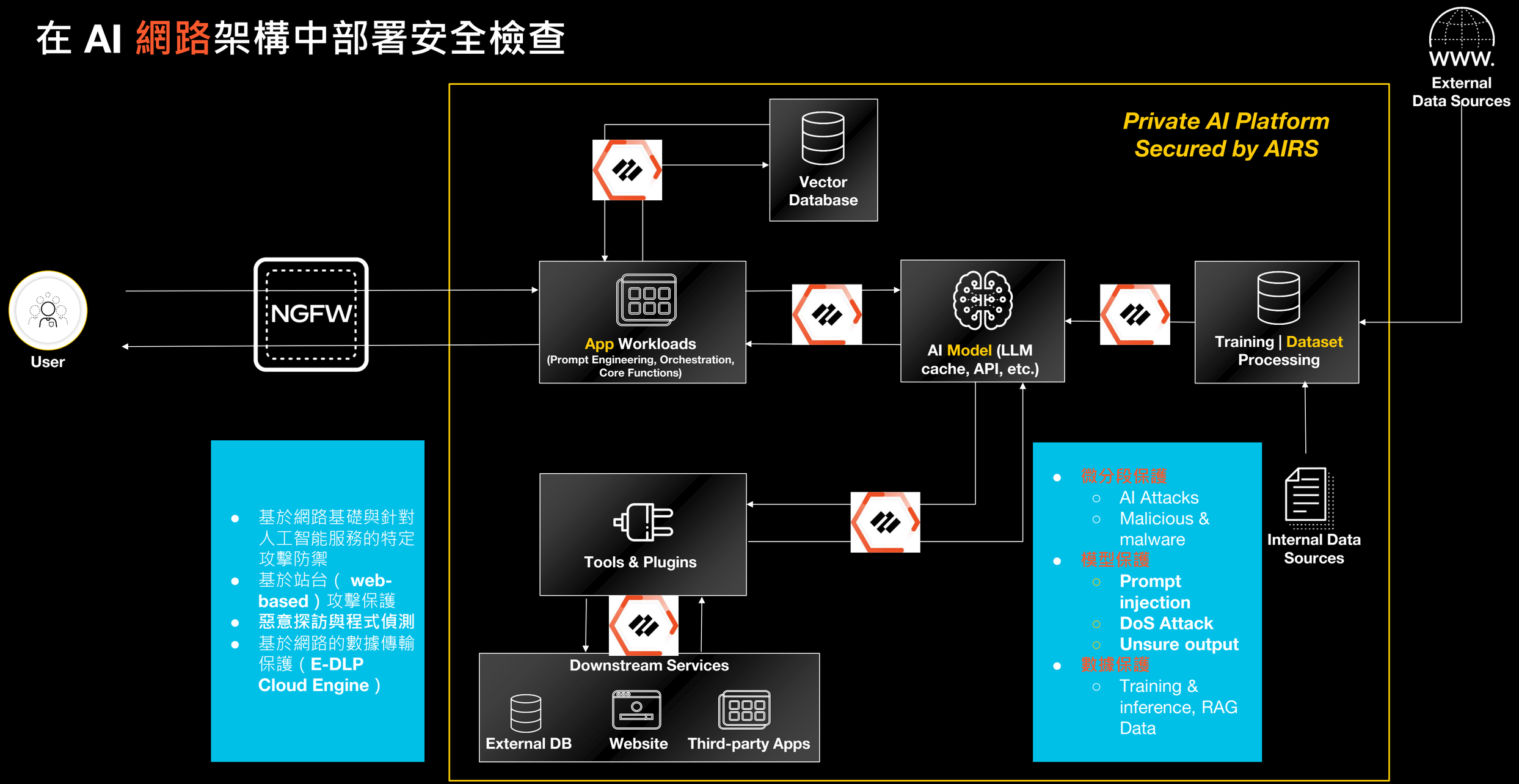
Detect and prevent threats in app code **without modifying the network**

在 AI 工作流中嵌入安全檢查（API）

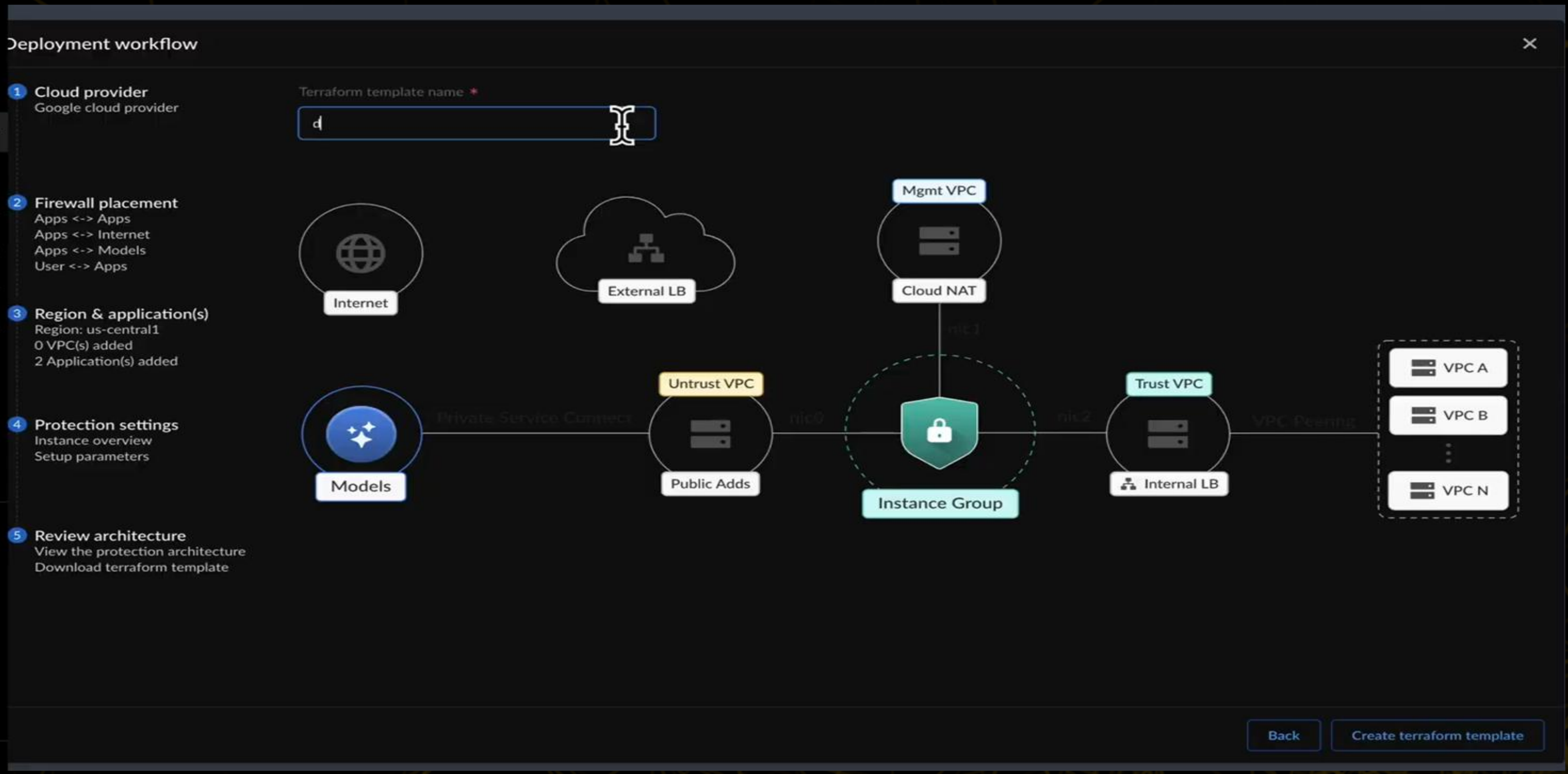
without modifying app code, the network



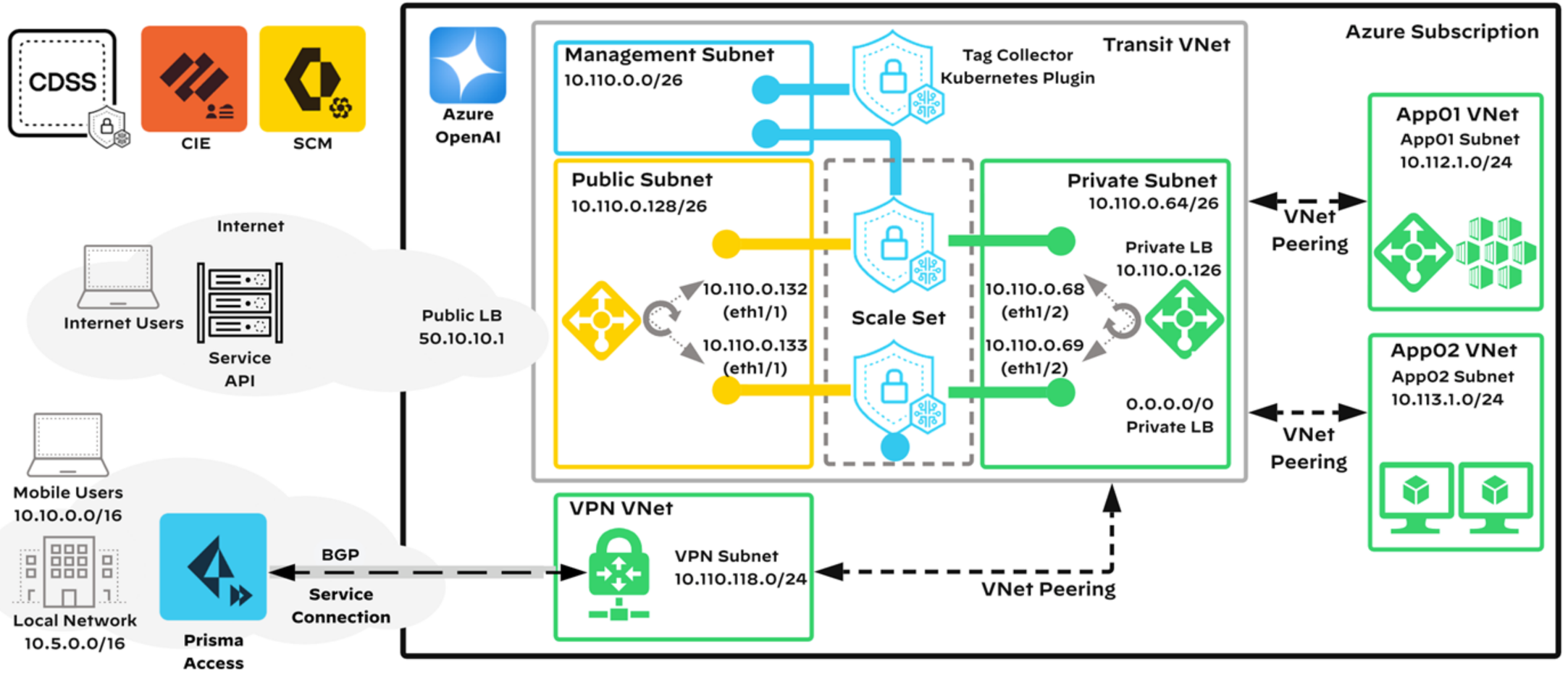
在 AI 網路架構中部署安全檢查



在 AI 網路架構中部署安全防護

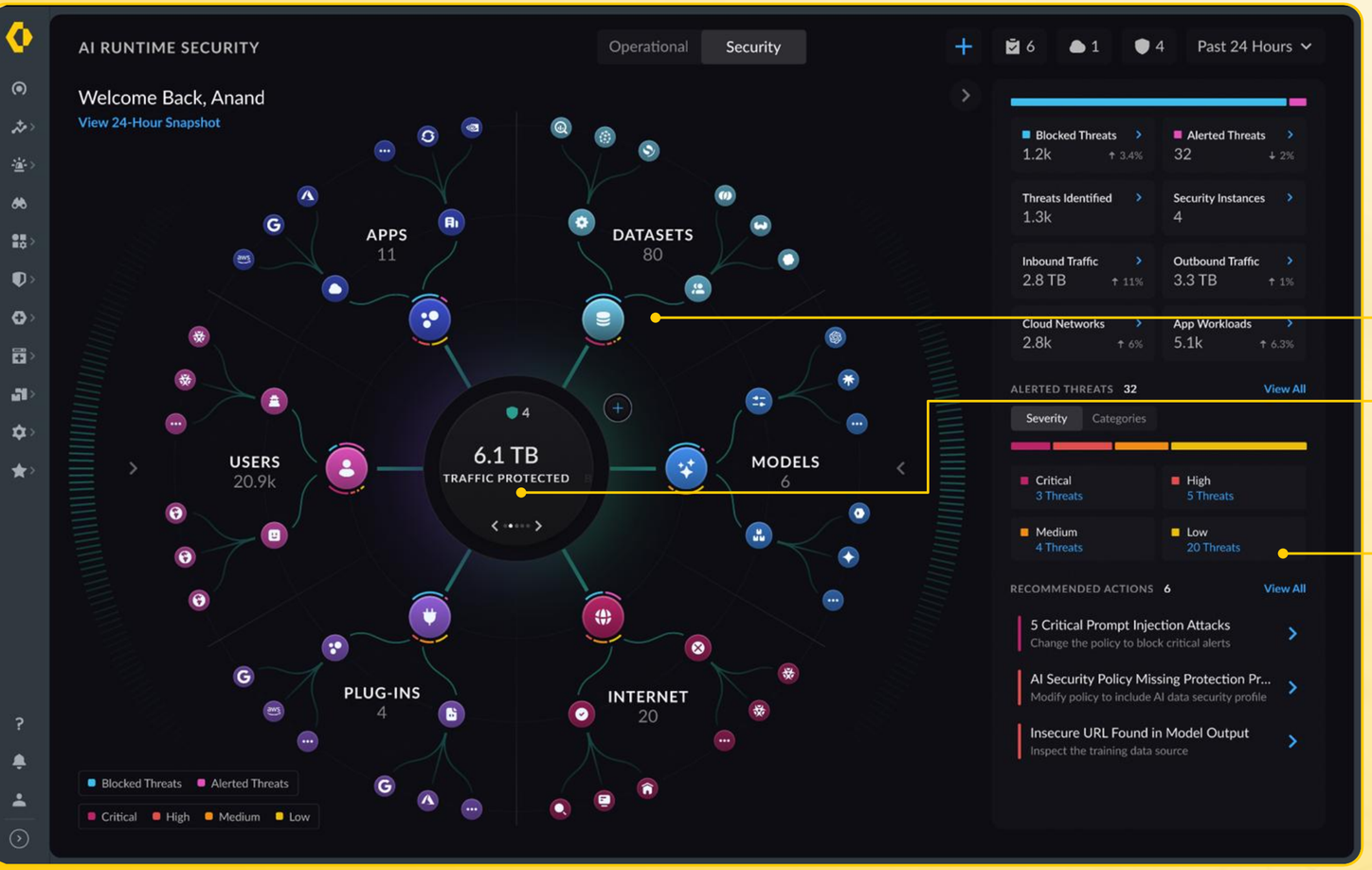


在 AI 網路架構中部署安全防護



再次獲得完整的管理可視性

Our approach to securing your enterprise AI applications.



Discover

all AI applications, models, users and datasets automatically.

Protect

models, data and apps against AI-specific and foundational attacks.

Monitor

the runtime risk exposure of your AI ecosystem continuously.

Protect

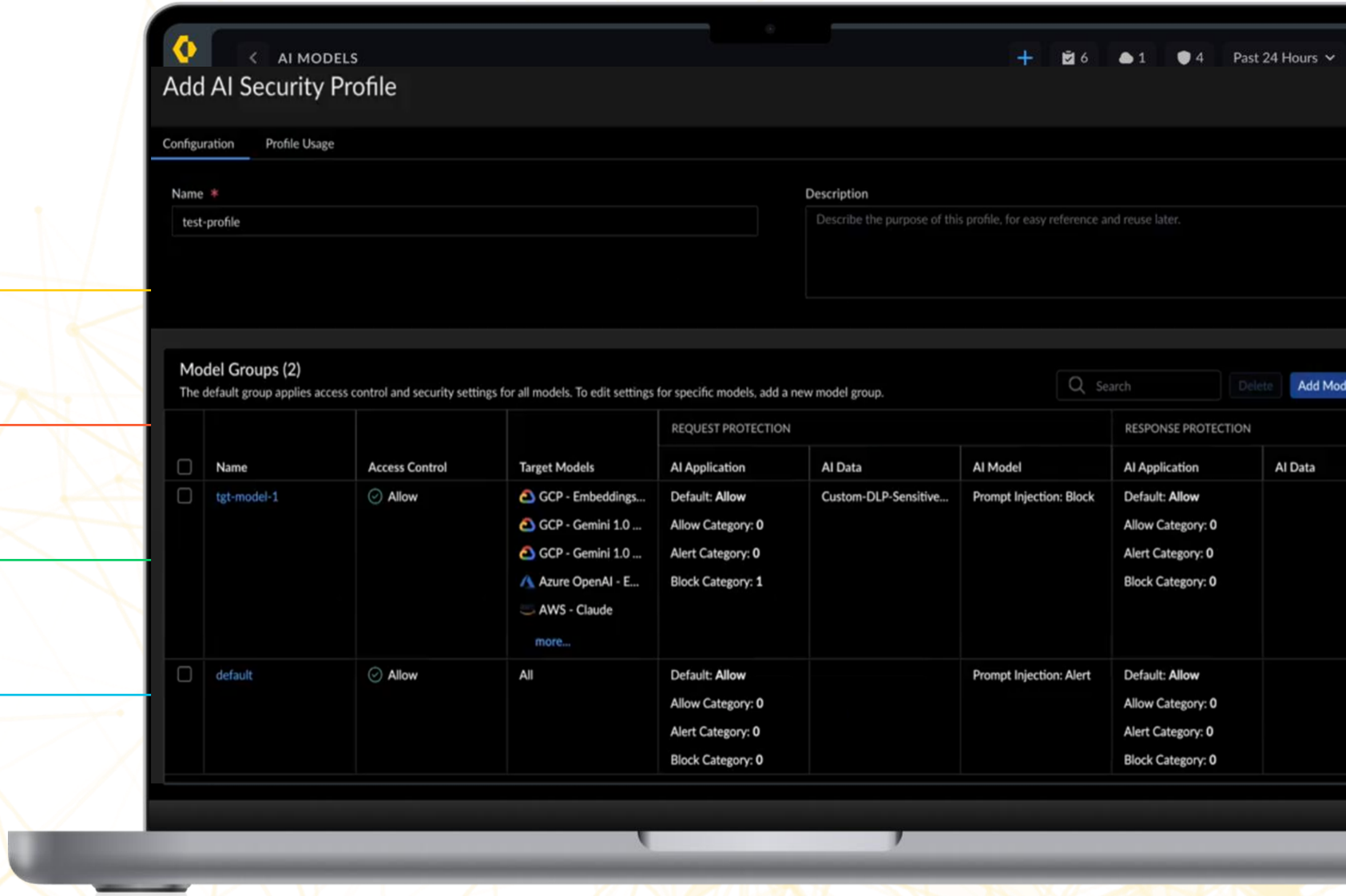
Your **AI models**

40+ models protected across GCP, Azure, AWS and OpenAI.

Block **12 types** of indirect & direct prompt injections and obfuscations.*

Prevent **Model DoS** attacks.*

Filter **unmoderated** content* (biased, sexual, harmful, hateful, criminal).

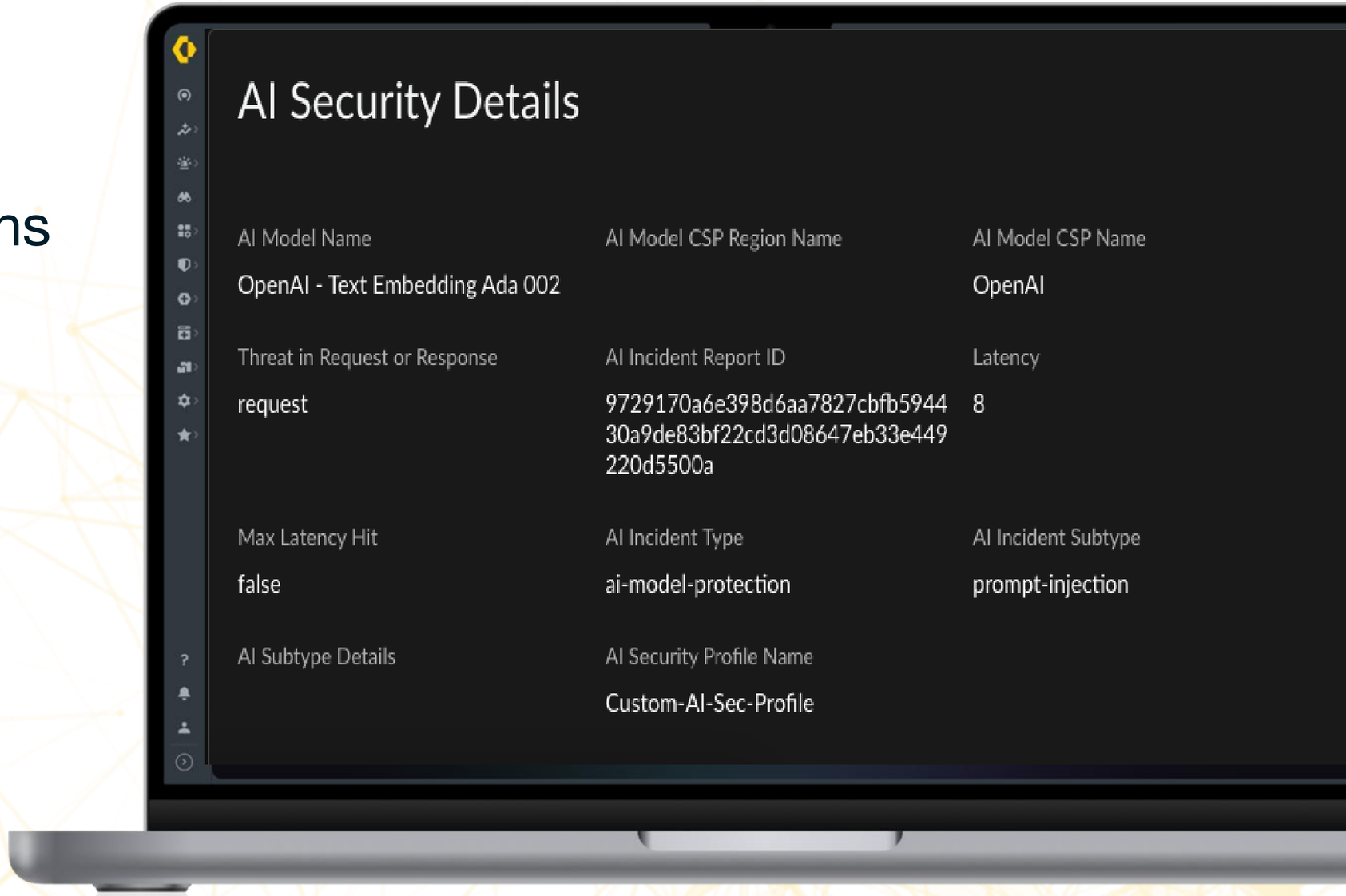


*Including post-GA functionality.

Monitor

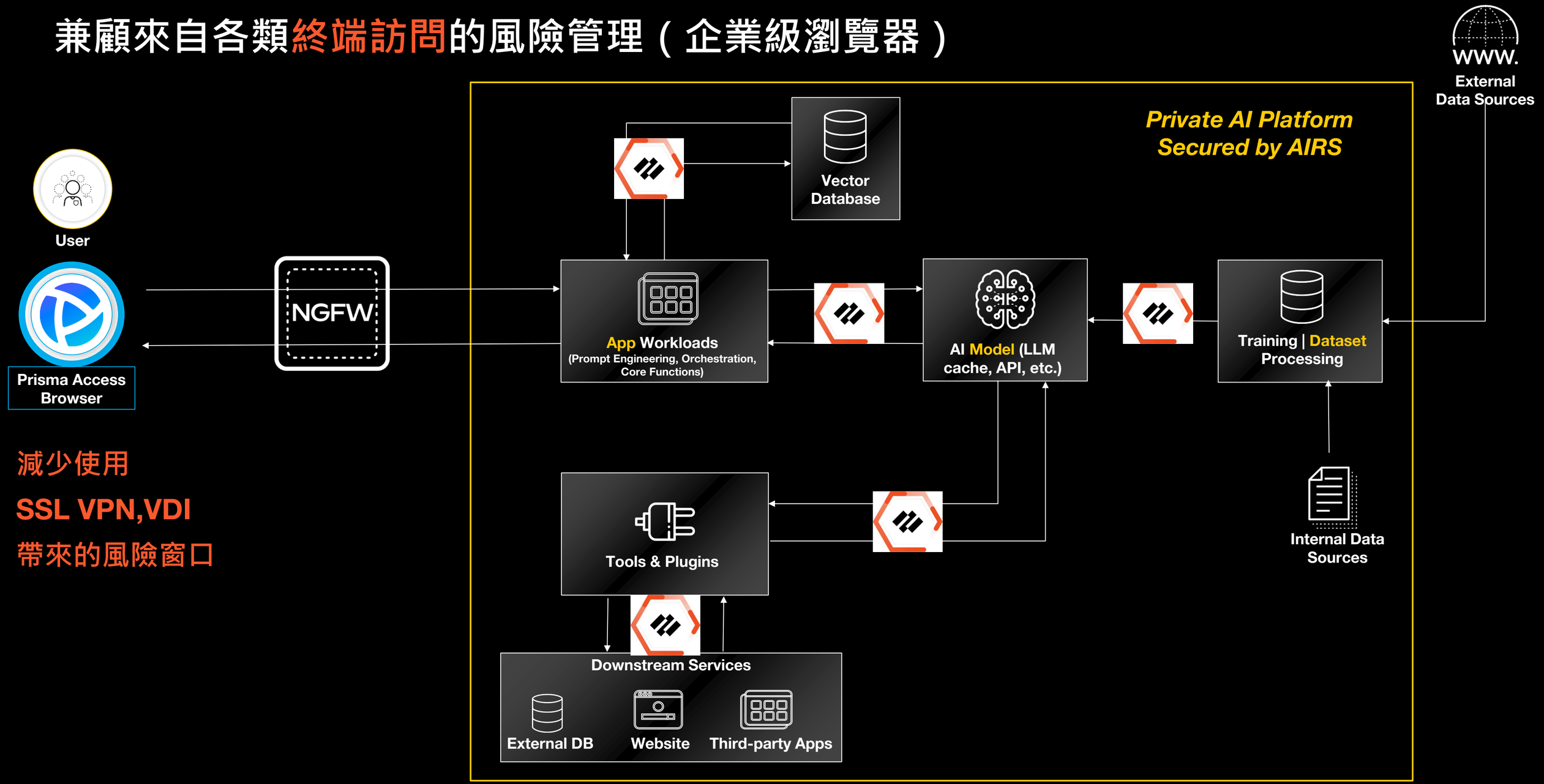
New threats and unprotected applications

- Analyze your AI Runtime **risk posture**.
- Evaluate **unprotected** AI applications.
- Identify **risky communication pathways** from AI applications



paloalto NETWORKS | STRATA

兼顧來自各類終端訪問的風險管理（企業級瀏覽器）



減少使用

SSL VPN,VDI

帶來的風險窗口

企業級瀏覽器
Prisma Access Browser

填問卷拿好禮
三步驟，加碼
再抽AirPod4

創建在任何裝置
上的安全區域

保障最後一哩的
數據安全

身份控制

打造適用於每個
角色的作業環境

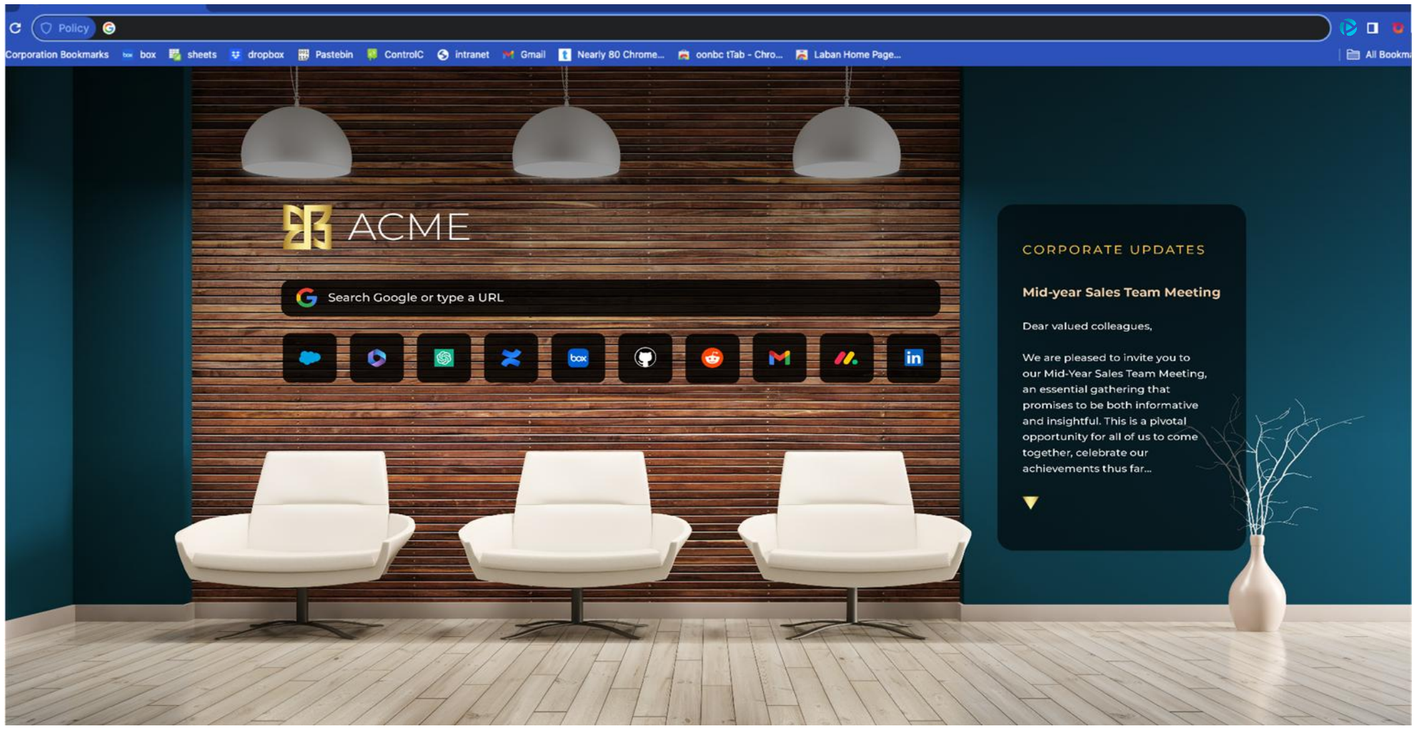
安全區域的實現機制

阻絕非監管裝置的各類風險

Isolates the workspace.

- **Hardens the browser** with additional layer of encryption.
- **Isolates environment** against:
  - Keyloggers
  - Screen scrapers
  - Untrusted certificates
  - Public networks

Checks device posture every 90 seconds.



完全可控的瀏覽器安全

Reduces attack surface by disabling sensitive browser components.

Defends against malicious extensions.

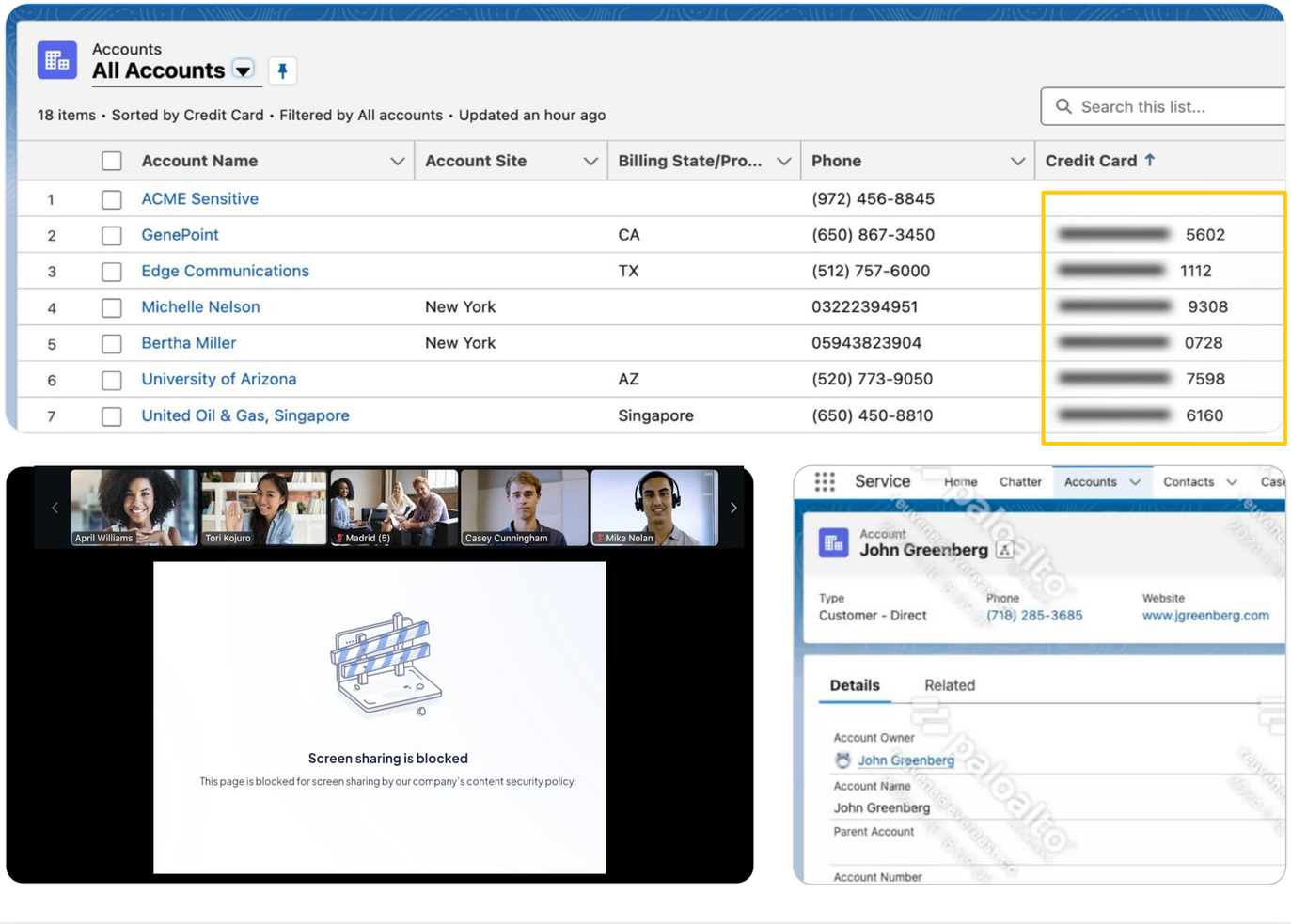
Collects web insights for threat hunting and forensics.

Activity auditing **User timeline** **Session recording**

數據訪問控制與安全

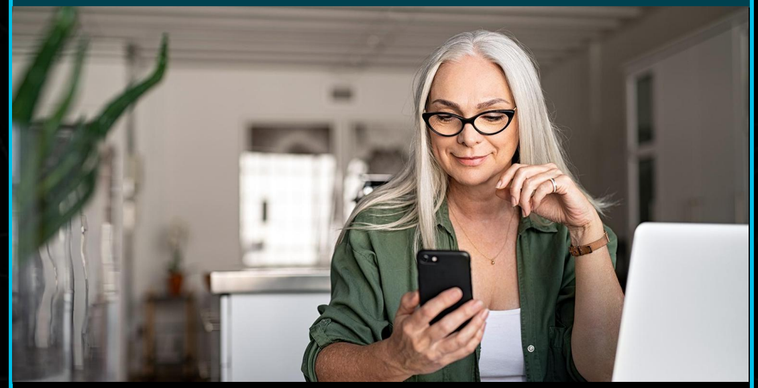
ALL user actions and last-mile.

- Unauthorized login
- File upload/download
- Copy/Paste
- Text typing
- Text masking
- Printing
- Screenshotting/Sharing
- Camera/Microphone
- Watermarks



適用於不同角色 - 輕量．高效．好管理

承包商和第三方



承包商

併購和收購

客服中心

業務單位

第一線和現場工作人員

員工自有裝置

BYOD 計畫

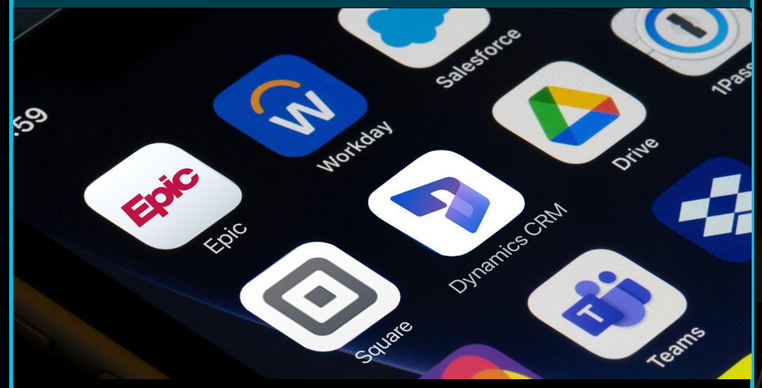
行動裝置

下班時間

運送筆記型電腦替代品

VDI,RDP,VNC 替換

以往無法想像的使用案例



進階 DLP

保護具有權限的使用者

啟用 **GenAI**

帳戶接管

影子 **IT**

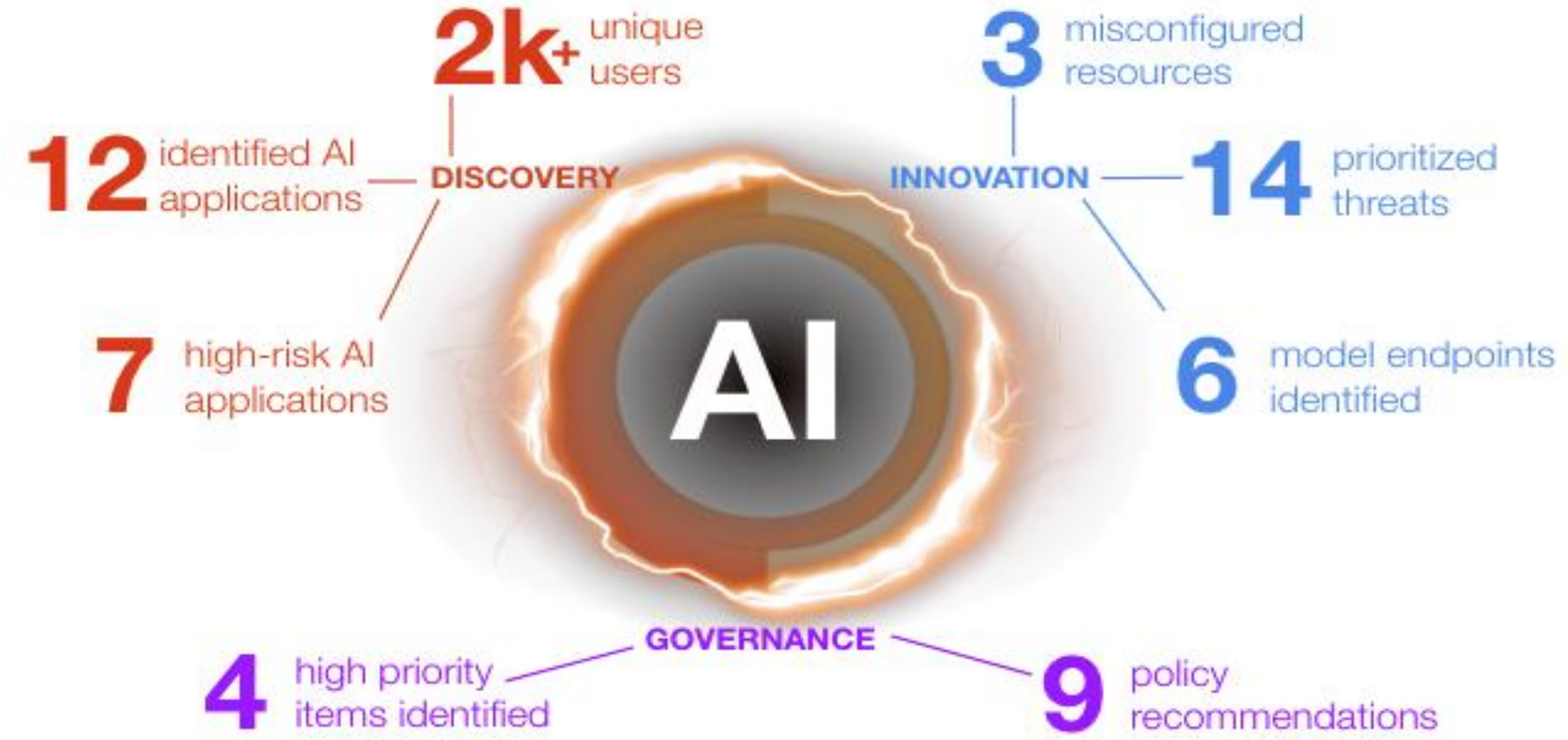
查看所有網路活動

允許未經授權的 **SaaS**

支援不可解密的流量

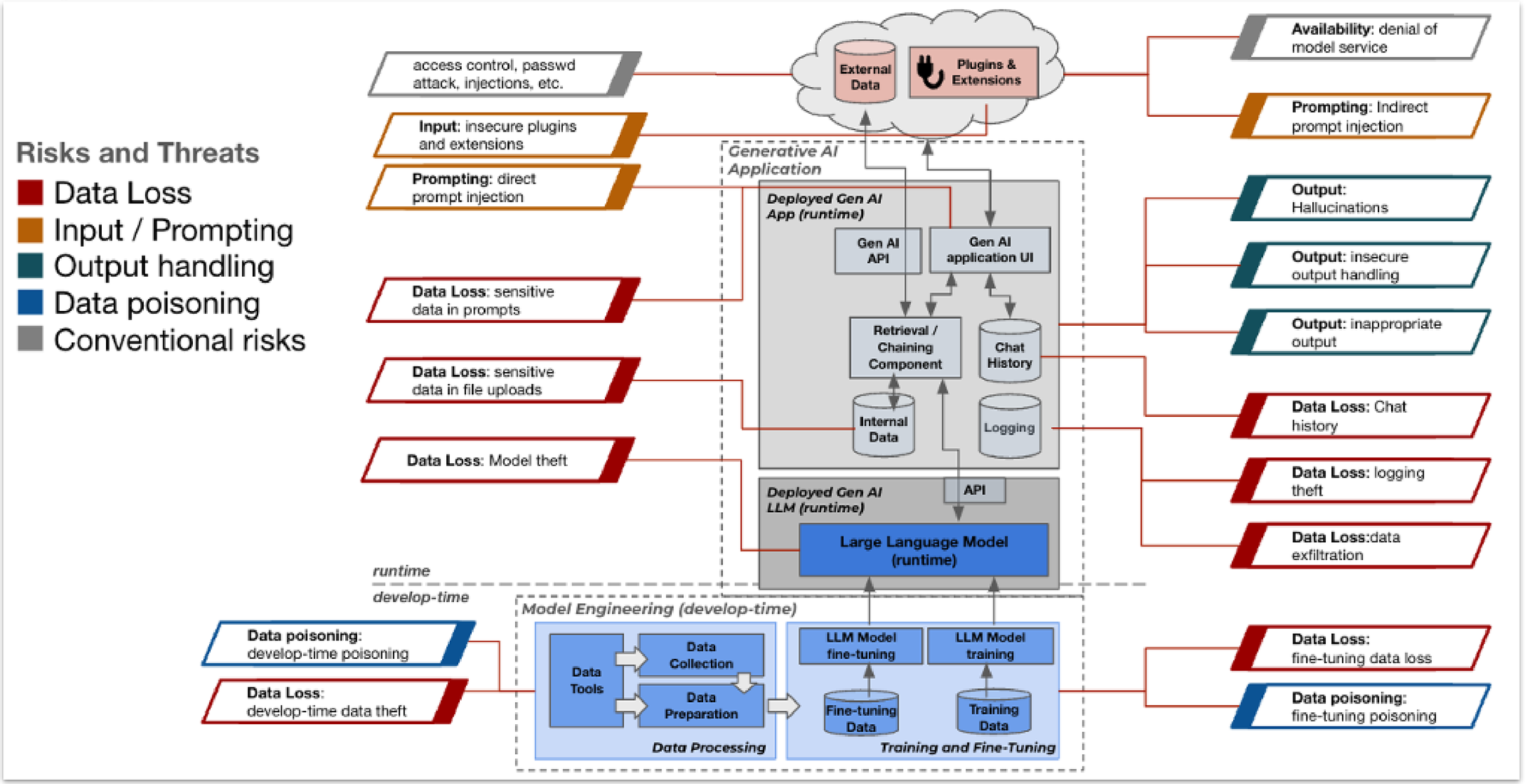
滿足新型態需求的專家服務

AI Security Assessment Service



認識維運環境的新面貌

AI security blueprint



Blueprint aggregates best practices across industry-leading frameworks (e.g., Google SAIF, NIST AI RMF, OWASP) into a practitioner-friendly format to enable rapid time-to-insight

填問卷拿好禮三步驟，加碼再抽AirPod4

Step.1
掃描QRCode 填寫問卷

Step.2截圖問卷完成畫面立即送（7F 701A）
搶先兌換聯名咖啡包

Step.3
至 Palo Alto Networks 攤位**C111**兌換精美好禮

Thank You

paloaltonetworks.com