



讓 AI 更安全！ 從 API 防護到 AI 應用安全的全面防護

F5, Inc.

Joshua 范茗閣

2025/04/16



Agenda

1

現代 AI 應用架構

2

AI 與 API：一體兩面的安全挑戰

3

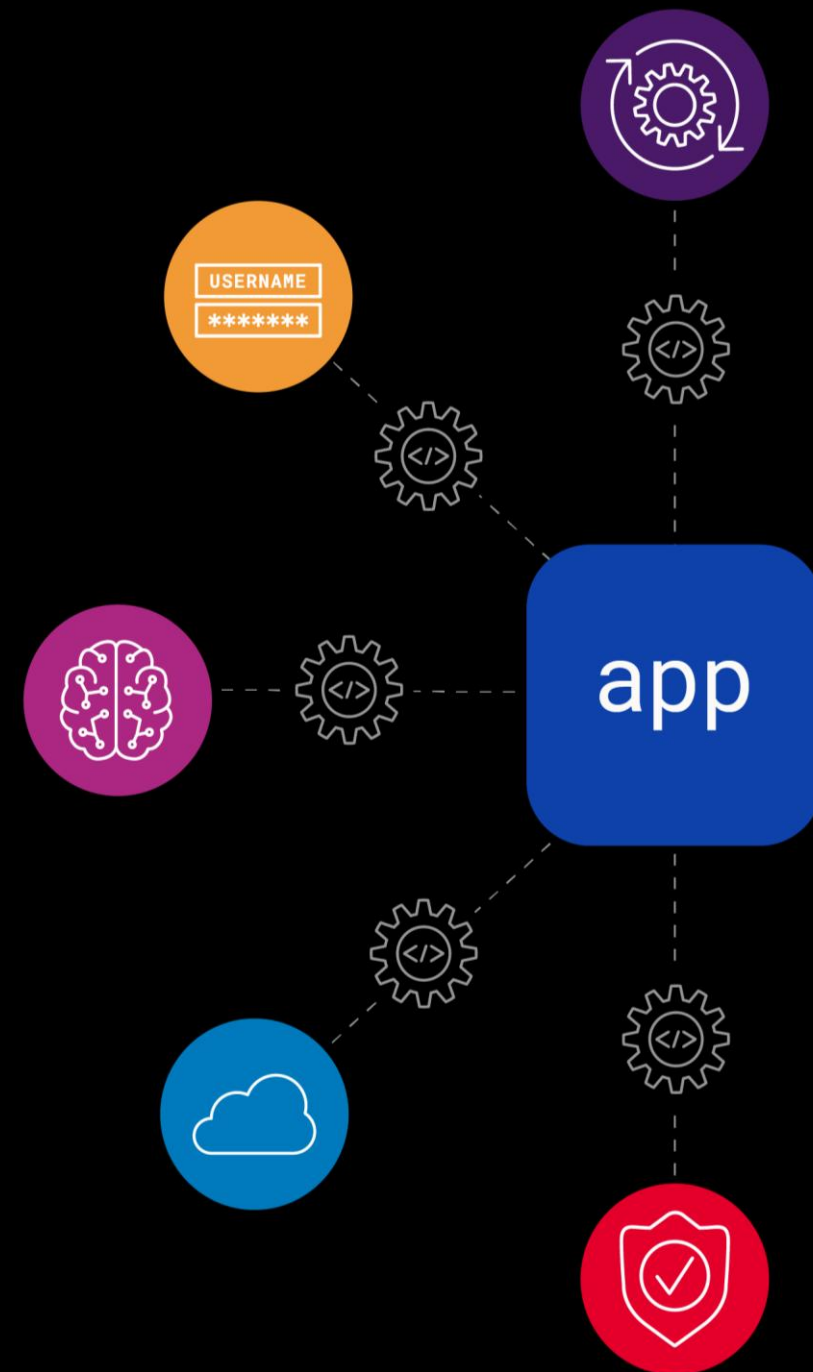
AI 驅動下的 API 防護新模式

4

AI Gateway：保護 AI 的第一道防線

5

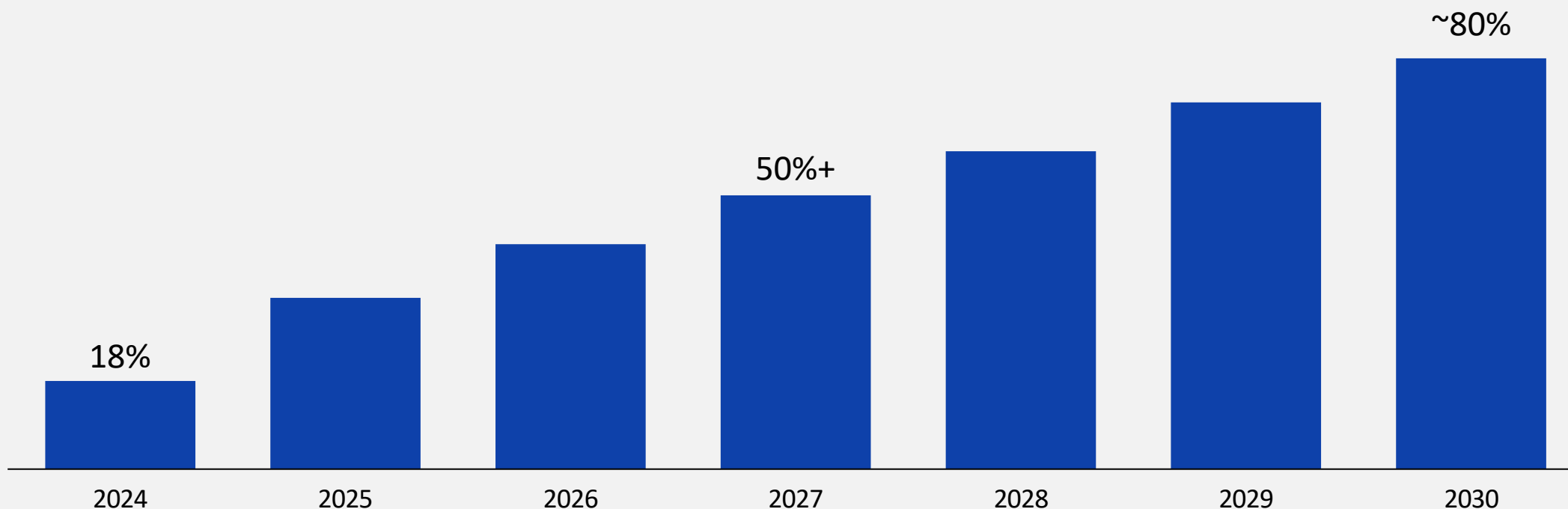
F5 AI 應用安全架構方案



現代 AI 應用架構

未來幾乎每個應用都會成為 AI 應用

Proportion of AI applications



Source: IDC, Gartner, and F5 Corporate Strategy

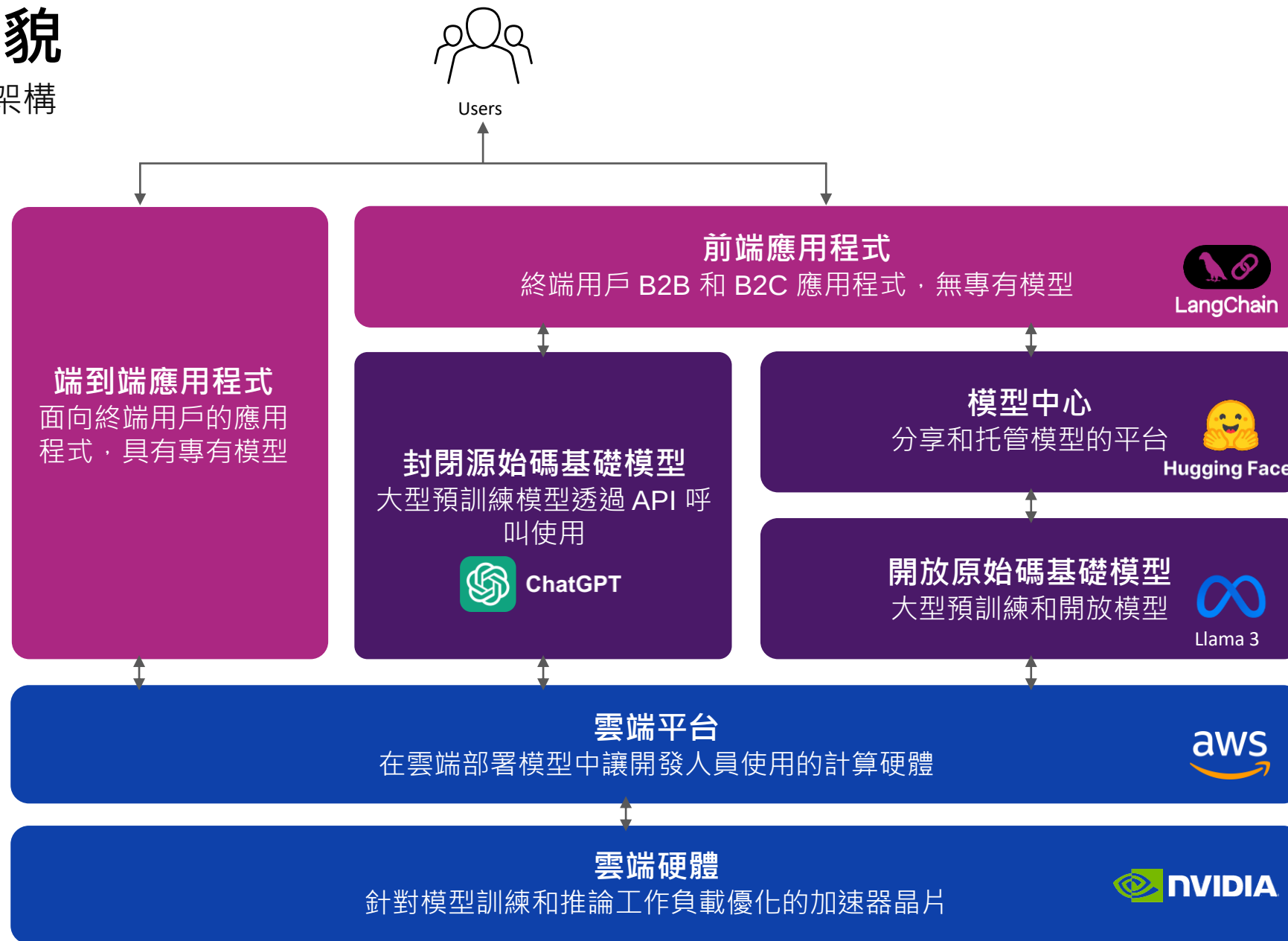
AI 應用架構全貌

應用、模型與雲端基礎架構

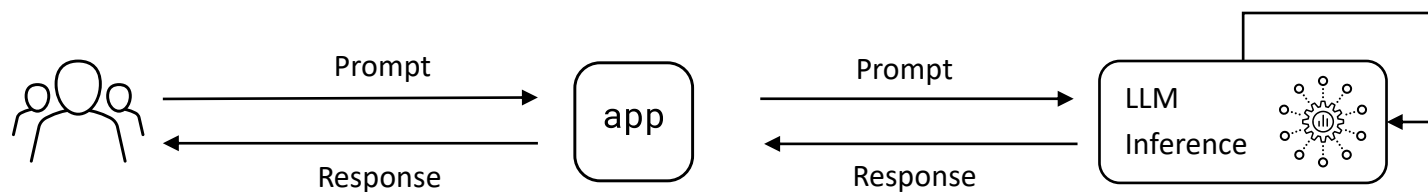


XOps

- 應用程式
- 模型
- 基礎設施



AI 應用在企業落地的基本架構



前端應用 (Application Front-end)

- 可部署於私有資料中心或雲端環境
- 可搭配既有工具與開發框架靈活整合

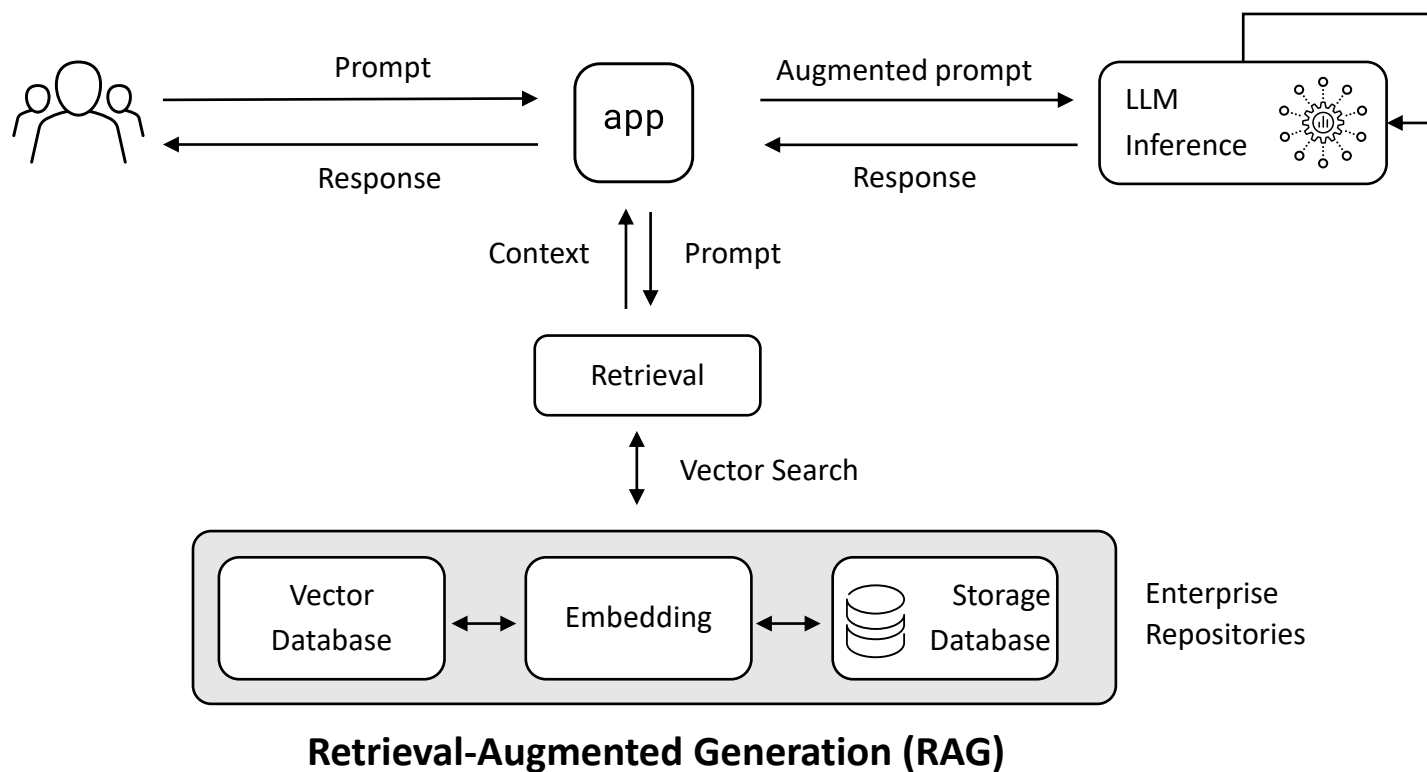


大型語言模型 (Large Language Model)

- 可透過 AI 工廠、自建私有資料中心、雲端平台或即服務 (as-a-service) 方式提供
- 可選擇使用私有 LLM 或公開模型 (如 OpenAI、Claude 等)
- 商用版本或開源模型皆可彈性應用



RAG：整合企業知識庫的 AI 應用架構



LLM 客製化 (LLM Customization)



- 預訓練 (Pre-training)
- 微調 (Fine-tuning)
- 檢索增強生成 (Retrieval-Augmented Generation, RAG)

Retrieval

檢索與嵌入 (Retrieval and Embedding)

- 可部署於私有資料中心或雲端平台，依資料敏感度彈性規劃
- 可使用主流開發框架，如 LangChain、Haystack、LlamaIndex 等進行檢索流程建構與語意嵌入處理



LangChain



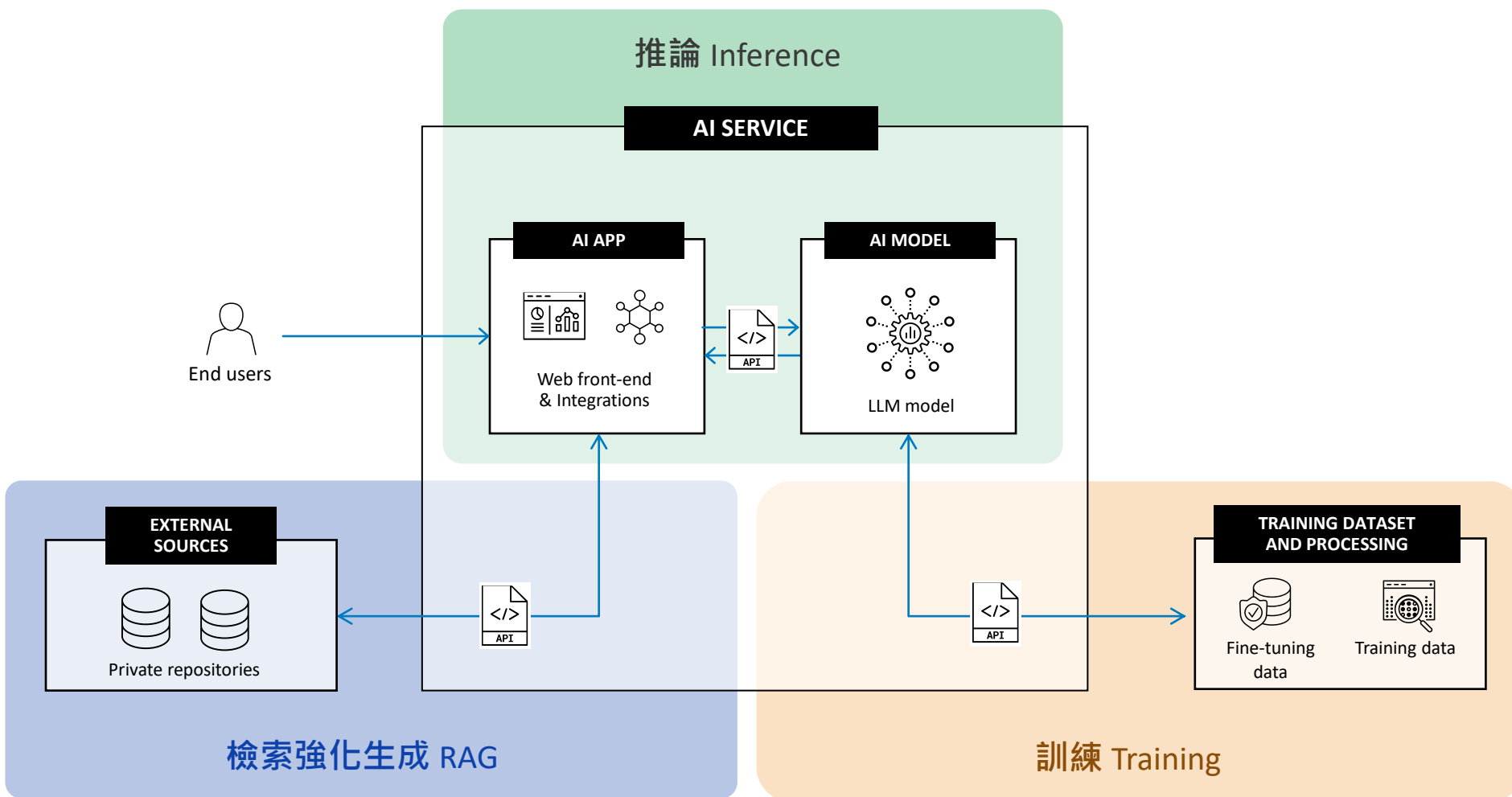
Haystack
by deepset



LlamaIndex



AI 應用架構三大核心：訓練、推論與檢索強化生成 (RAG)



AI 與 API： 一體兩面的安全挑戰

API：連接 AI 模型、資料與應用的關鍵橋樑

串接訓練資料，驅動模型學習流程

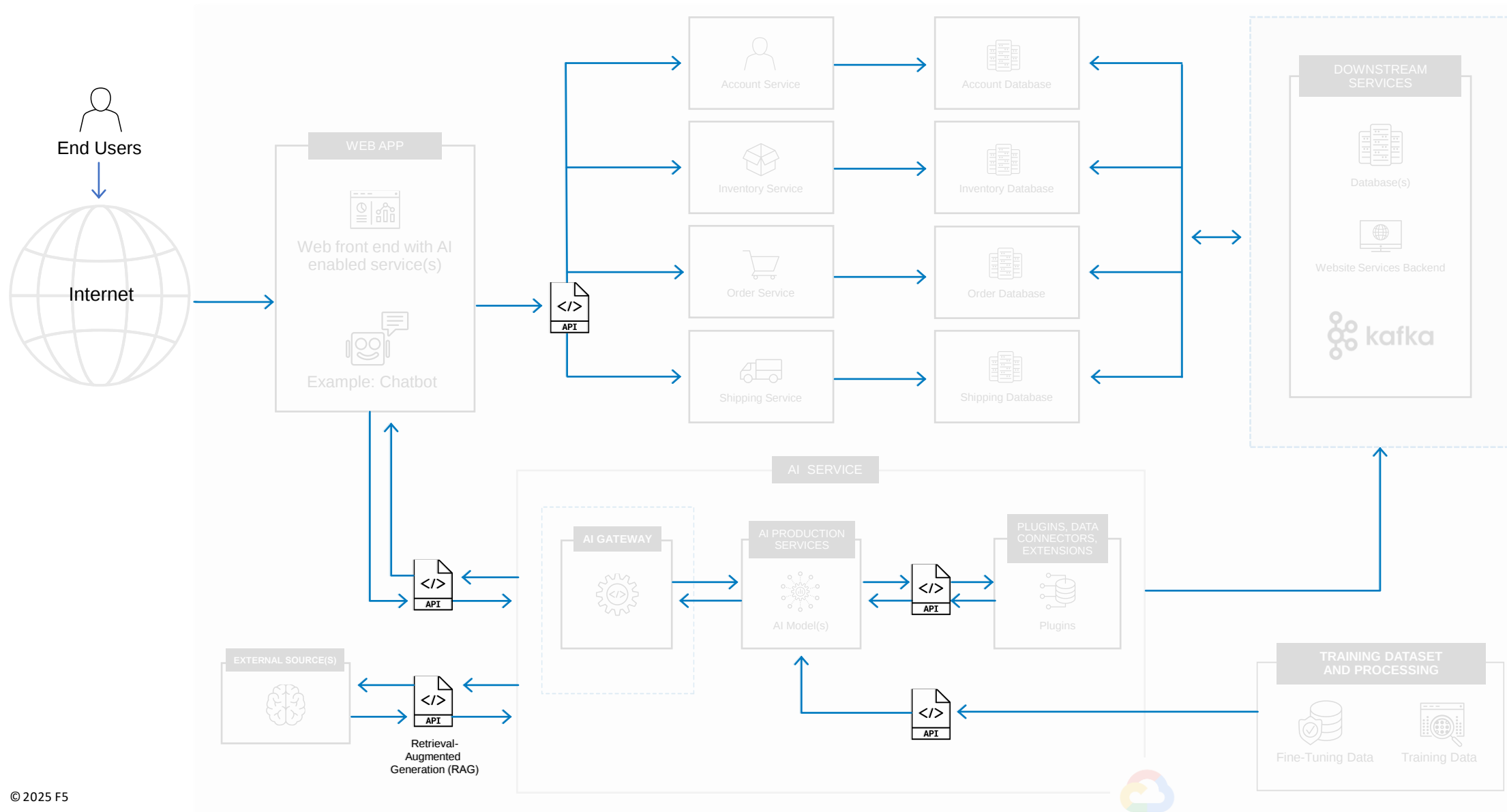
將訓練完成的模型整合進實際推論應用中

串接外部資料來源，擴充 AI 模型
語意能力（如 RAG）

將推論結果整合進 AI 應用與
使用者體驗中

API 是貫穿 AI 架構開發與應用部署的核心連接通道，無論是模型訓練資料的輸入、推論模型的存取，還是最終將 AI 能力交付給使用者，所有環節幾乎都仰賴 API 來完成整合與串接。

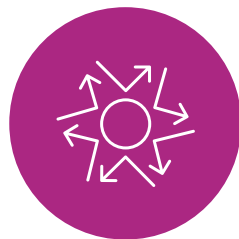
API：連接 AI 模型、資料與應用的關鍵橋樑



API 可能讓 AI 應用 與資料暴露在多種 威脅之中



輸入注入攻擊 (Injection Attacks on AI Input)



阻斷服務攻擊 (DoS Attacks Impacting AI Functionality)



資料外洩 (AI Data Exfiltration)



中間人攻擊 (Man-in-the-Middle on Model Communication)

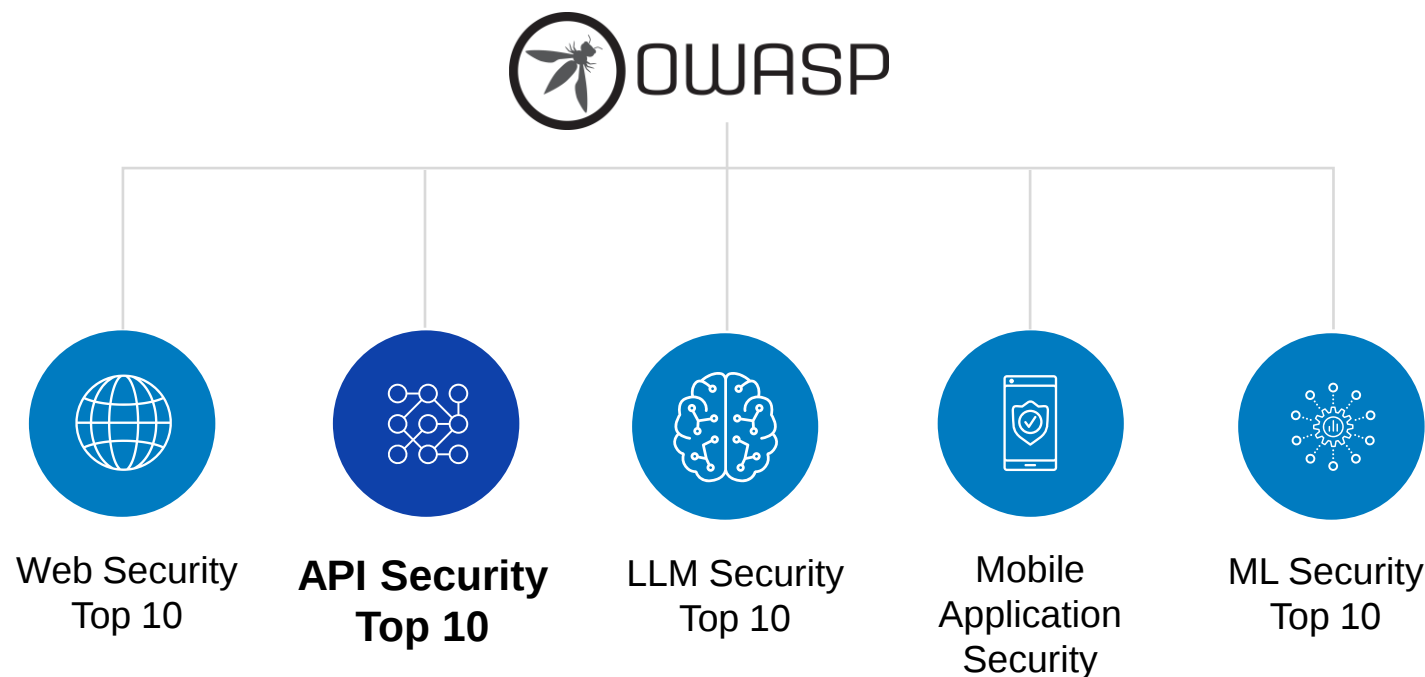
AI 安全與 API 安全，本質上是一體兩面

API 是攻擊者的主要入口，若無法保護提供模型服務的 API，就無法真正保護 AI 本身

“ ”

無論你如何使用生成式 AI，
最終你都是通過 SDK、函
式庫或 **REST API** 來呼叫一
個端點

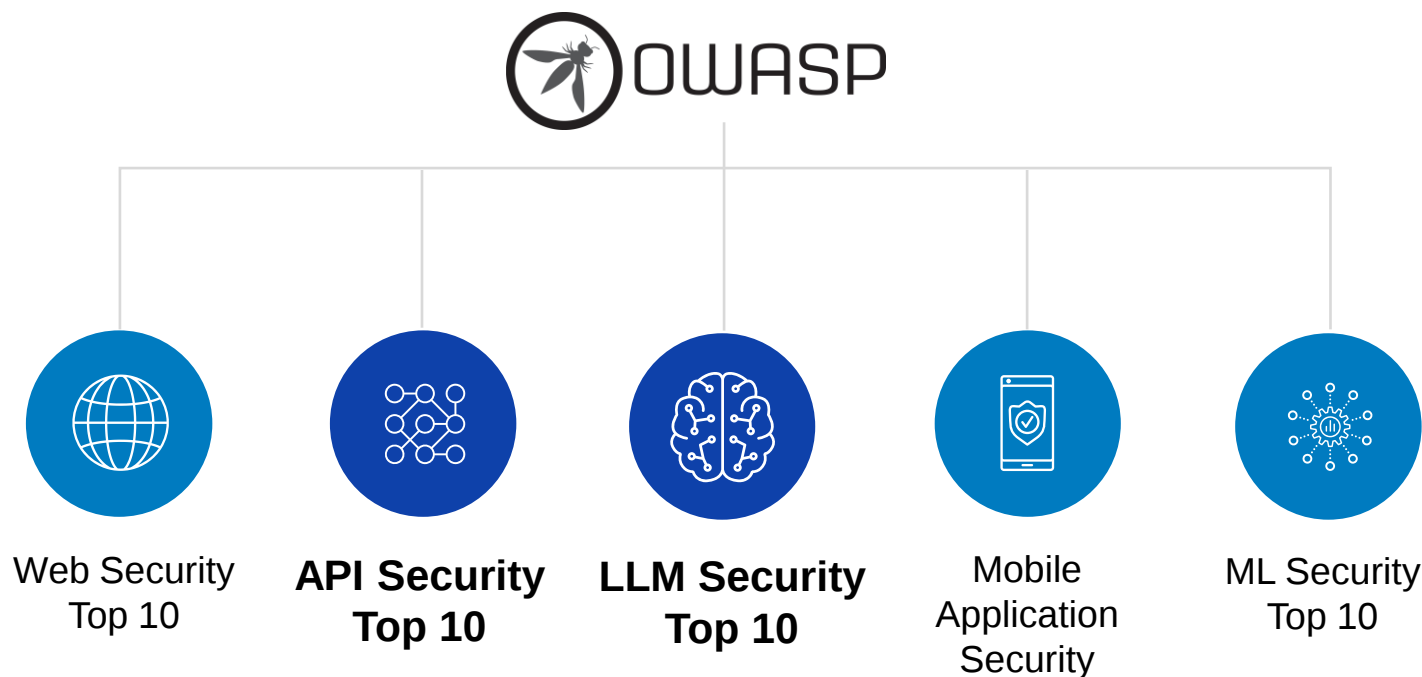
Mete Atamel
Developer Advocate, Google Cloud



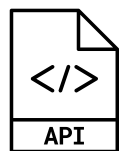
AI 為原已複雜的資安問題帶來更多全新威脅

隨著 AI，特別是大型語言模型 (LLM) 的應用普及，企業的風險面將進一步擴大，暴露於前所未見的攻擊手法與資安挑戰之中：

- 提示詞注入 (Prompt Injection)
- 輸出處理不當 (Improper Output Handling)
- 無限制的資源消耗 (Unbounded Consumption)
- 敏感資訊洩漏 (Sensitive Information Disclosure)
- 過度自主權 (Excessive Agency)



Web、API 到 AI：新一代 AI 應用的三重安全挑戰



AI / LLM Apps

應用安全

WEB APPLICATION FIREWALL
DDOS MITIGATION
BOT DEFENCE

API SECURITY

防護 OWASP API Top 10 攻擊
API 自動發現與掃描
API 管理政策與法規遵循
執行期間的 API 行為防護

AI SECURITY

防護 OWASP LLM Top 10 攻擊
提示詞注入 (Prompt Injection) 防禦
訓練資料汙染 (Data Poisoning) 偵測
模型竊取 (Model Theft) 防護

應用交付

**APPLICATION DELIVERY
CONTROLLER (ADC)**

API GATEWAY API MANAGEMENT

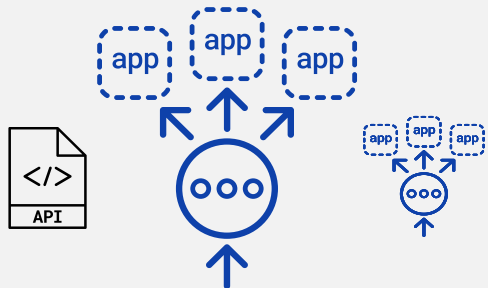
API 請求路由控制
開發者入口網站 (Dev Portal)
請求速率限制

AI / LLM PROXY

深度整合大型語言模型 (LLM)
提示詞控管與最佳化
完整可觀測性 (Rich Observability)
Token 計算與資源追蹤

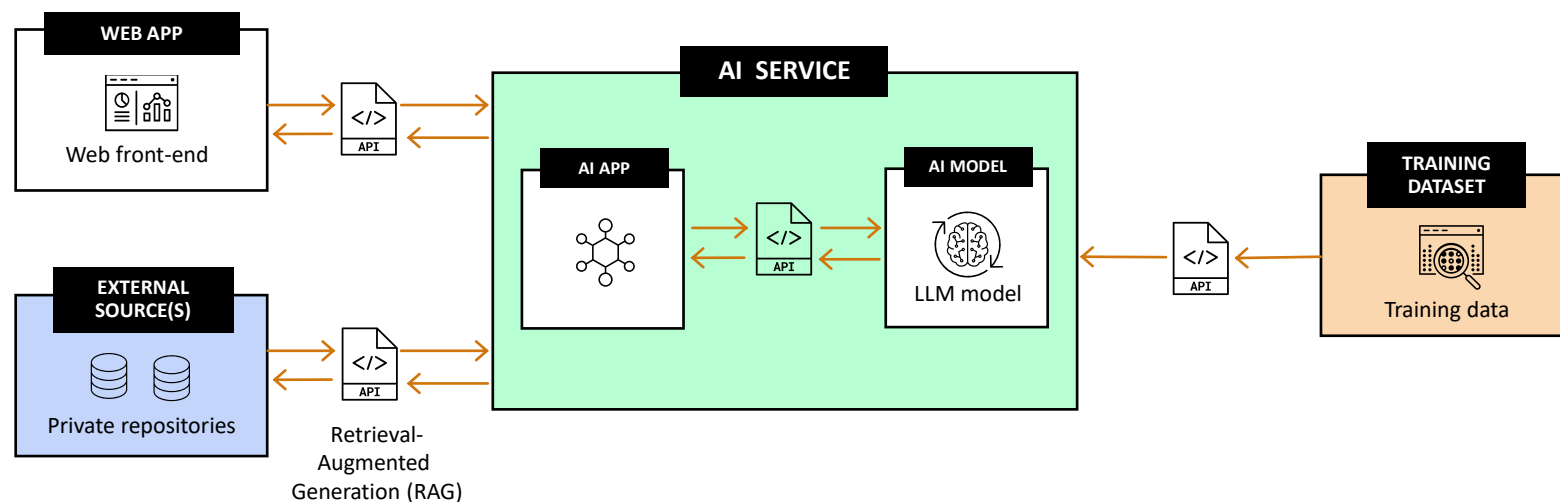
AI 驅動下的 API 防護新模式

API Gateway

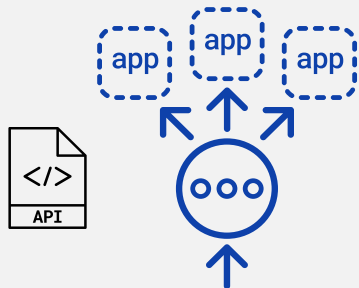


- API 路由策略
- API 流量負載平衡
- 憑證加解密處理
- 存取控制
- 身分驗證與權限控管
- 請求速率控制與限制
- 監控與可視化儀表板

現有的 API Gateway 的角色與功能



API Gateway

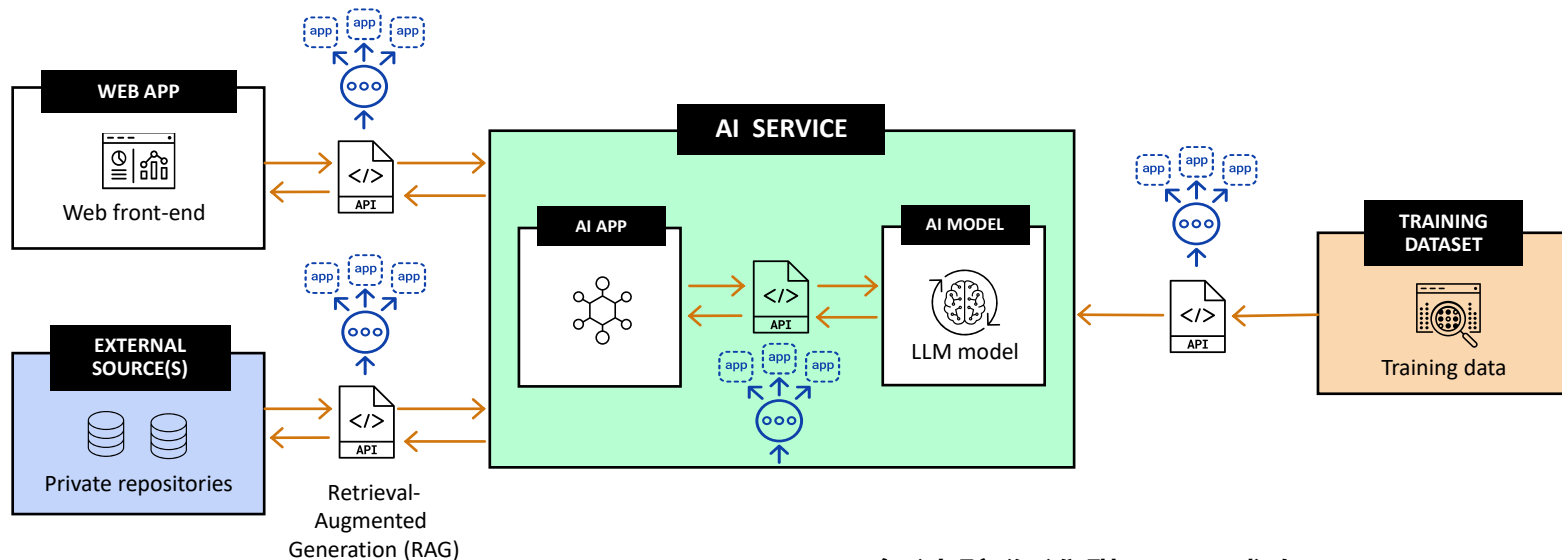


- API 路由策略
- API 流量負載平衡
- 憑證加解密處理
- 存取控制
- 身分驗證與權限控管
- 請求速率控制與限制
- 監控與可視化儀表板

現有的 API Gateway 安全防護能力有限...



只提供基本的安全和存取授權控制



無法防禦進階 API 威脅

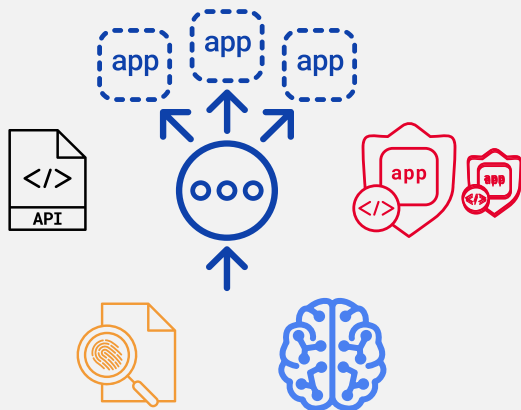
無法量化 API 風險

無法發現和監控所有 API 資源

無法延伸防護到原始碼與開發階段

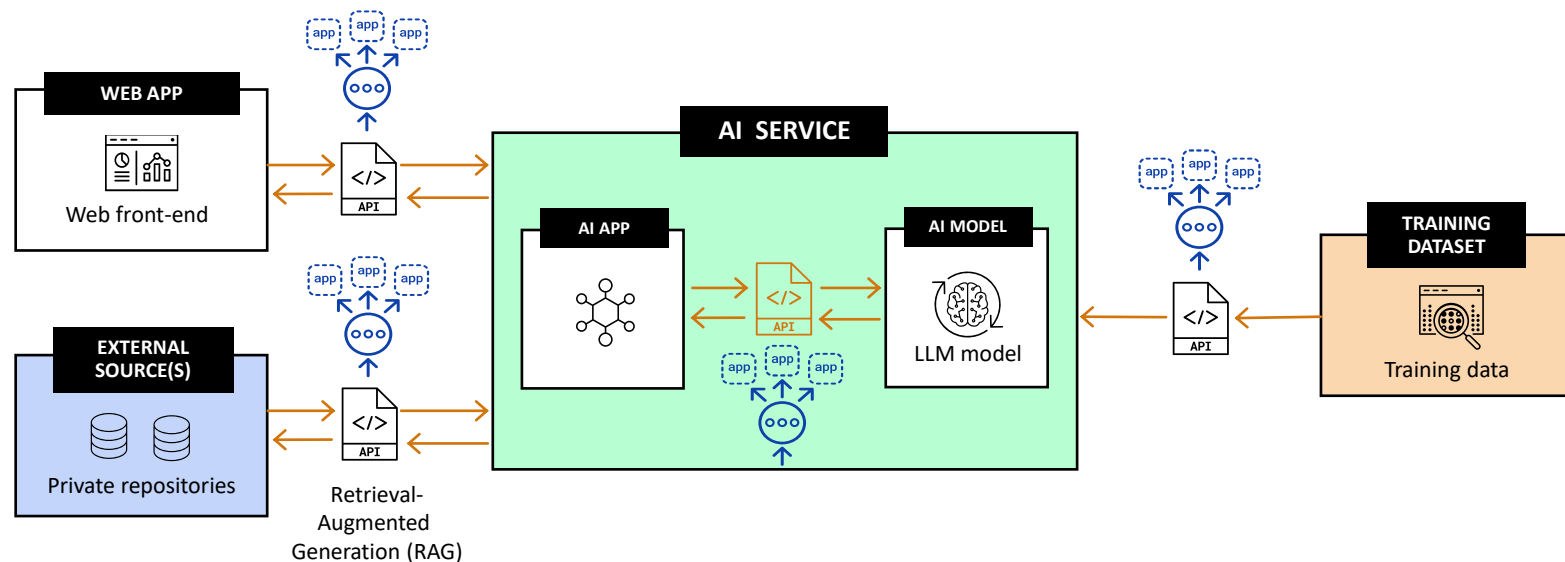
API Security

API Gateway

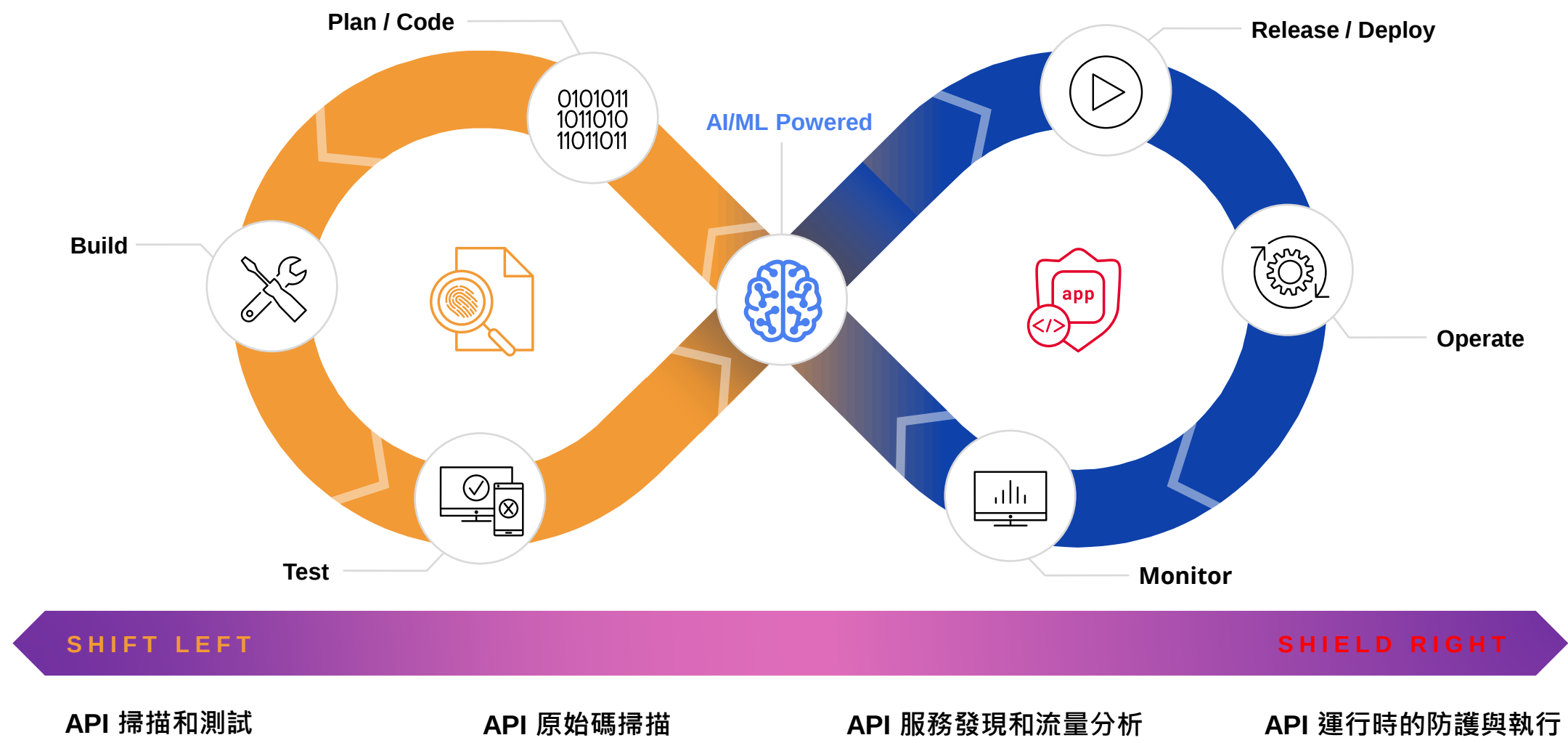


- 依 URI 控制允許的 Method
- 針對個別 API 進行不同策略防護
- 參數檢查與合規控管
- 常見弱點防護機制
- DDoS 防禦與機器人攻擊防護
- 流量盤點和安全態勢分析
- 效能和安全可視化
- 弱點掃描和探測
- 原始碼掃描
- AI 助理協助安全維運

API Security 是實現全面防護的關鍵核心



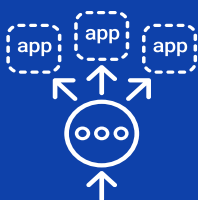
從開發到運行，貫徹 API 安全防護



AI Gateway : 保護 AI 的第一道防線

API 與 AI Gateway 的功能領域與進化

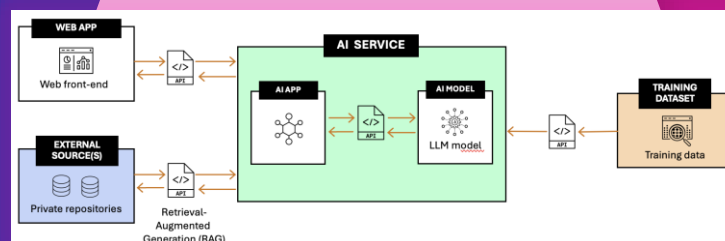
API Gateway



以服務為中心，是支撐 AI 應用的核心元件

- 面向服務
- 面向流量
- 安全與性能導向
- OWASP API TOP 10

AI Application Architecture



AI Gateway



以模型為中心，具備可語意感知與自然語言處理能力

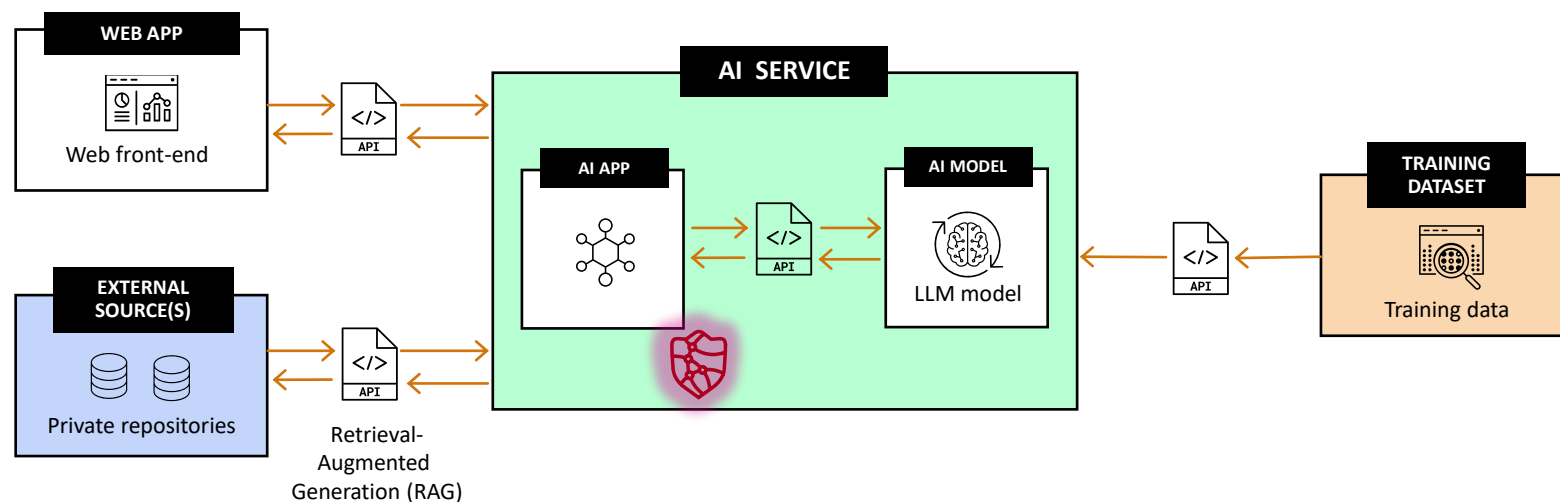
- 面向模型
- 面向成本
- 安全與智慧導向
- OWASP LLM TOP 10

AI Gateway



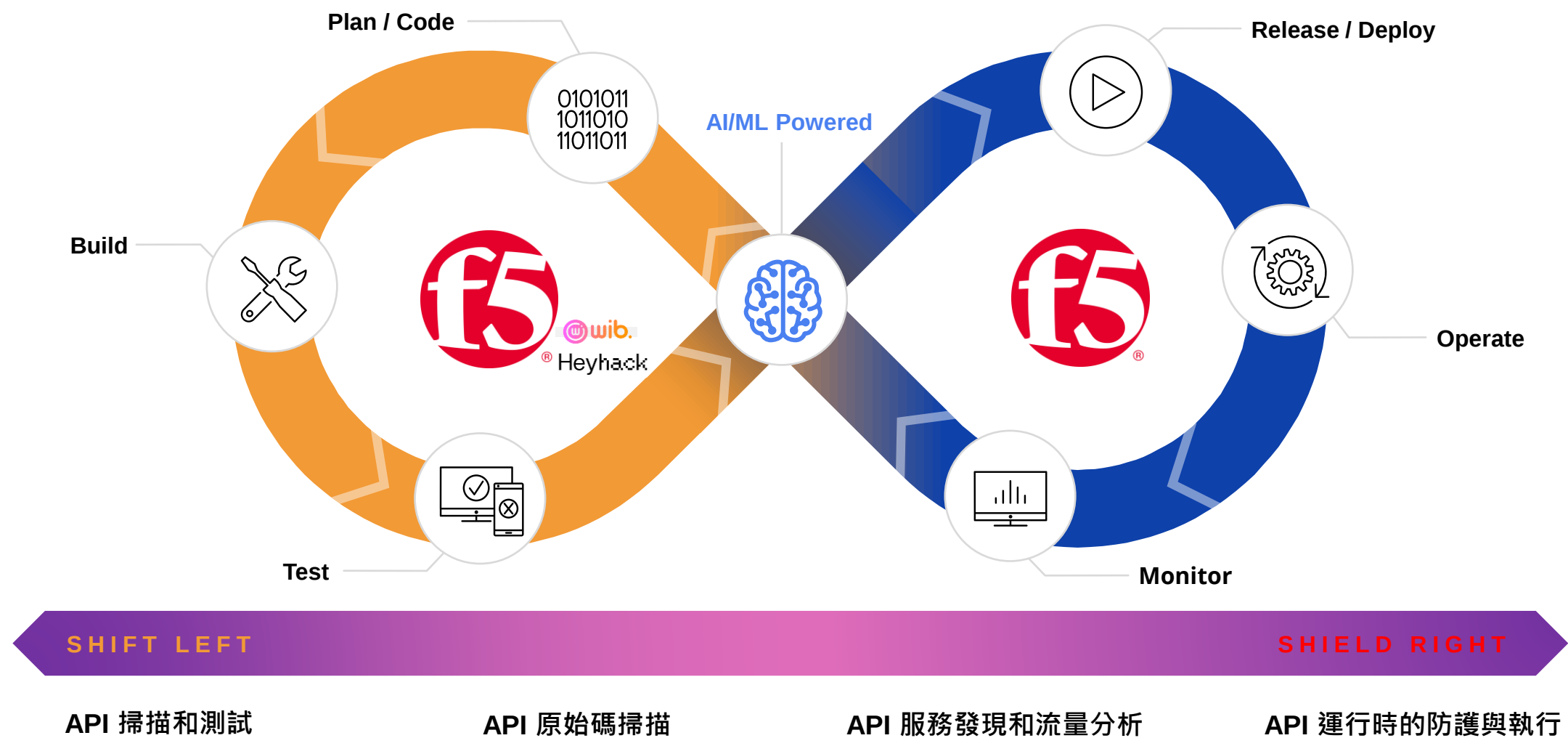
- 連線管理與失敗重試機制
- 安全防護與注入攻擊阻擋
- 成本觀測與資源使用管理
- 提示詞優化與範本管理機制
- 語意理解與語境意圖辨識
- 語意快取與重複請求優化
- 幻覺抑制與模型輸出控管

AI Gateway 的角色與核心能力

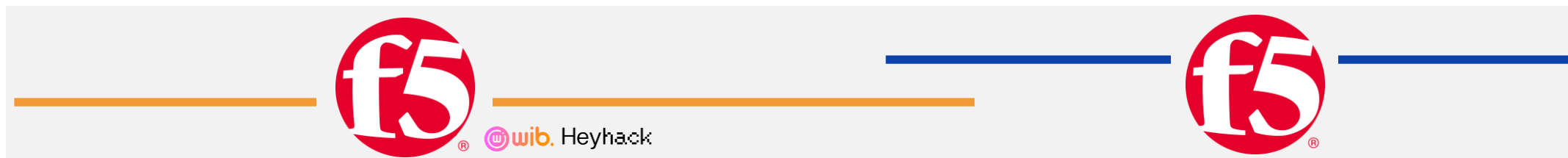


F5 AI 應用安全架構方案

從開發到運行，F5 打造最完整的 API 安全防護



從開發到運行，F5 打造最完整的 API 安全防護



- 程式碼安全分析（原始碼檢測）
- 預生產環境漏洞掃描
- API 安全測試與驗證
- 透過爬蟲技術全面識別攻擊面
- 產出報告、工單與即時通知

- OWASP Top 10 安全風險防護
- DDoS 攻擊防禦與緩解
- API 自動探索與辨識
- API 安全驗證與合規檢查
- 全方位可視化監控

SHIFT LEFT

SHIELD RIGHT

API 掃描和測試

API 原始碼掃描

API 服務發現和流量分析

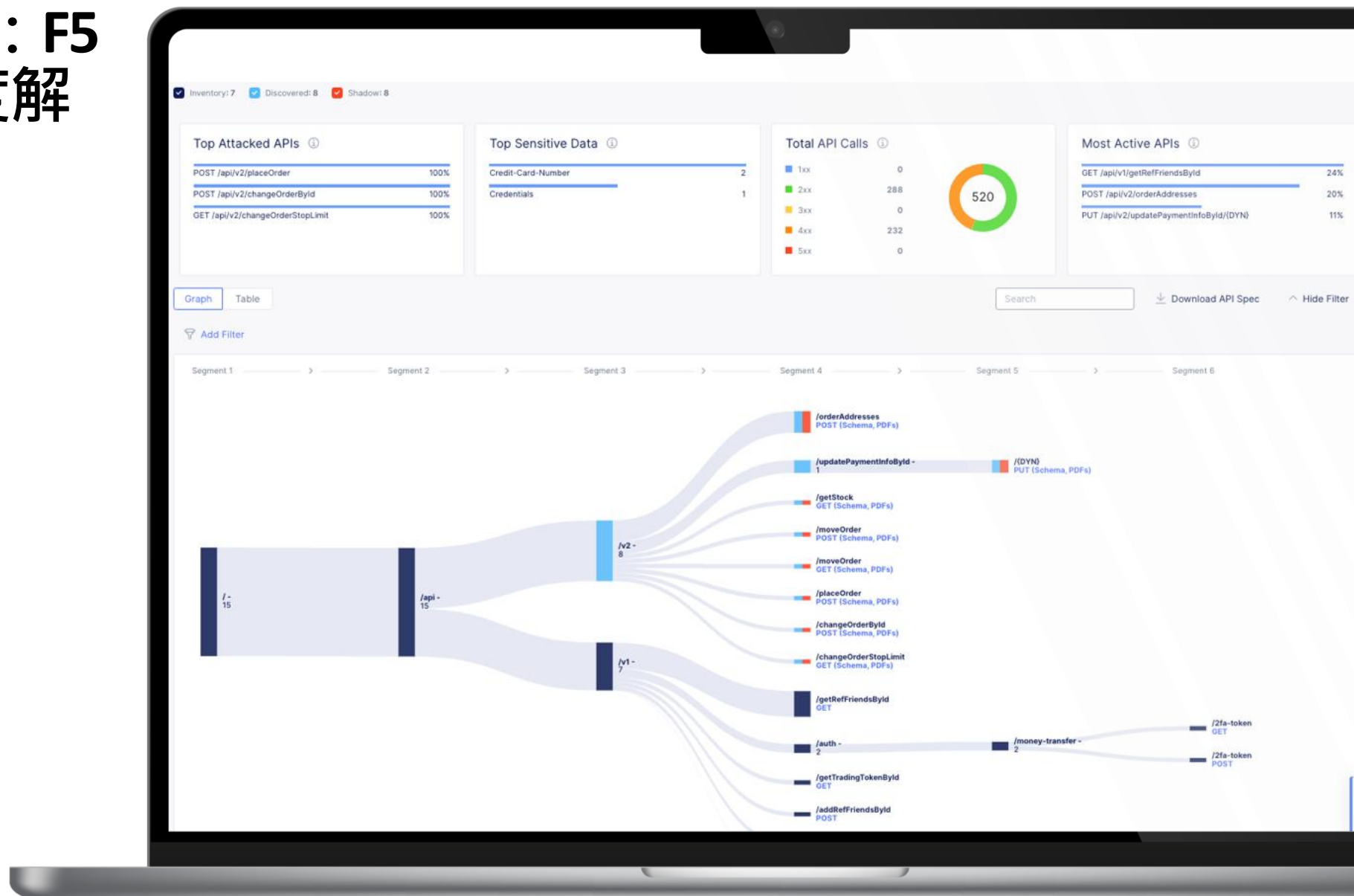
API 運行時的防護與執行

API 可視性再進化：F5 提供多重視角深度解析

F5 結合多種探索技術，協助完整掌握所有 API 狀態：

- 透過原始碼進行 API 掃描與發現
- 依據流量進行即時分析與檢測
- 利用自動爬蟲技術探索隱藏 API

可自動識別 PII 等敏感資料暴露風險，並根據原始碼與流量生成 OpenAPI (OAS) 規格檔案



API 端點威脅評分與安全分析

即時掌握每個 API 端點的威脅狀況與風險評分，快速聚焦需要優先處理的高風險項目

TableGraph

Search

Add Filter

21 Items

CurrentArchived

☐	API Endpoint ⓘ	Method	Sensitive Data ⓘ	Threat Level ⓘ	API Category	Discovery Source ⓘ	Risk Score ⚡	Schema Status ⓘ
☐	/api/adjectives	POST	email location-data +2 View All	High	DiscoveredInventory	Traffic, Code Analysis	90	OpenAPI
☐	/api/animals	POST	sentence-api-france-social-n... social-security +4 View All	High	DiscoveredInventory	Traffic, Code Analysis	90	OpenAPI
☐	/api/colors	GET	location-data	High	DiscoveredShadow	Traffic, Code Analysis	70	Discovered
☐	/api/locations	GET	location-data	High	DiscoveredInventory	Traffic, Code Analysis	0	OpenAPI
☐	/api/locations	POST	—	High	DiscoveredInventory	Code Analysis	0	OpenAPI
☐	/api/locations	DELETE	—	None	DiscoveredShadow	Code Analysis	0	Discovered
☐	/api/sizes	GET	—	None	DiscoveredShadow	Code Analysis	0	Discovered
☐	/api/sizes	POST	—	None	DiscoveredShadow	Code Analysis	0	Discovered
☐	/api/sizes	DELETE	—	None	DiscoveredShadow	Code Analysis	0	Discovered
☐	/api/locations/{id}	GET	—	None	Inventory	—	0	OpenAPI

10

50

100

Items per page

Endpoint Details

API Endpoint/rest/basket/0somerreally...
Base Path/rest
API Type/REST
Risk Score/Low 30
Method/GET
API Category/Inventory, Discoverd

Authentication/Authenticated, JWT: Header
Discovery Source/Traffic, Code
API Compliance/PCI, GDPR, HIPPA
Protection Rule/Configured Edit
Rate Limit/Configured Edit
OpenAPI Validation/Configured Edit

OverviewDiscoveredInventory OpenAPISecurity Posture

ActiveArchived

Vulnerabilities

Query parameter contains sensitiv... High
State: Open
Created: 10:59 AM, Jan 21
Last Observed: 08:34 AM, Jan 25

Authenticated API accessed witho... High
State: Under Review
Created: 10:59 AM, Jan 21
Last Observed: 08:34 AM, Jan 25

HTTP is supported and Strict-Tran... Medium
State: Open (Manually)
Created: 10:59 AM, Jan 21
Last Observed: 08:34 AM, Jan 25

Error handling reveals stack trace... Low
State: Open
Created: 10:59 AM, Jan 21
Last Observed: 08:34 AM, Jan 25

Sensitive data in API response wit... Low
State: Open
Created: 10:59 AM, Jan 21
Last Observed: 08:34 AM, Jan 25

API parameter contains URL... Low
State: Open
Created: 10:59 AM, Jan 21
Last Observed: 08:34 AM, Jan 25

State
Open
Change State

Category
Weak authentication / authorization

Description
A request without Authorization header was sent to an API endpoint that requires authentication and response code "200 OK" was returned. Attackers may call this API endpoint successfully without authentication and read or modify application data.

Risk score
Attack impact: 9.0 (Critical)
Authorization bypass may allow read/write (depending on the endpoint). This leads to loss of confidentiality and integrity

Evidence
[References to requests that lack HSTS header](#)

Remediation
To mitigate this issue, use the following solutions:
1. Change application code to enforce authorization.
2. Configure Authorization header to be mandatory and specify the expected authorization scheme.
Expected authorization method and example of a valid request GET /api/protected-resource HTTP/1.1
Authorization: Bearer <access_token>

Malicious activity

API 資產盤點與風險評分

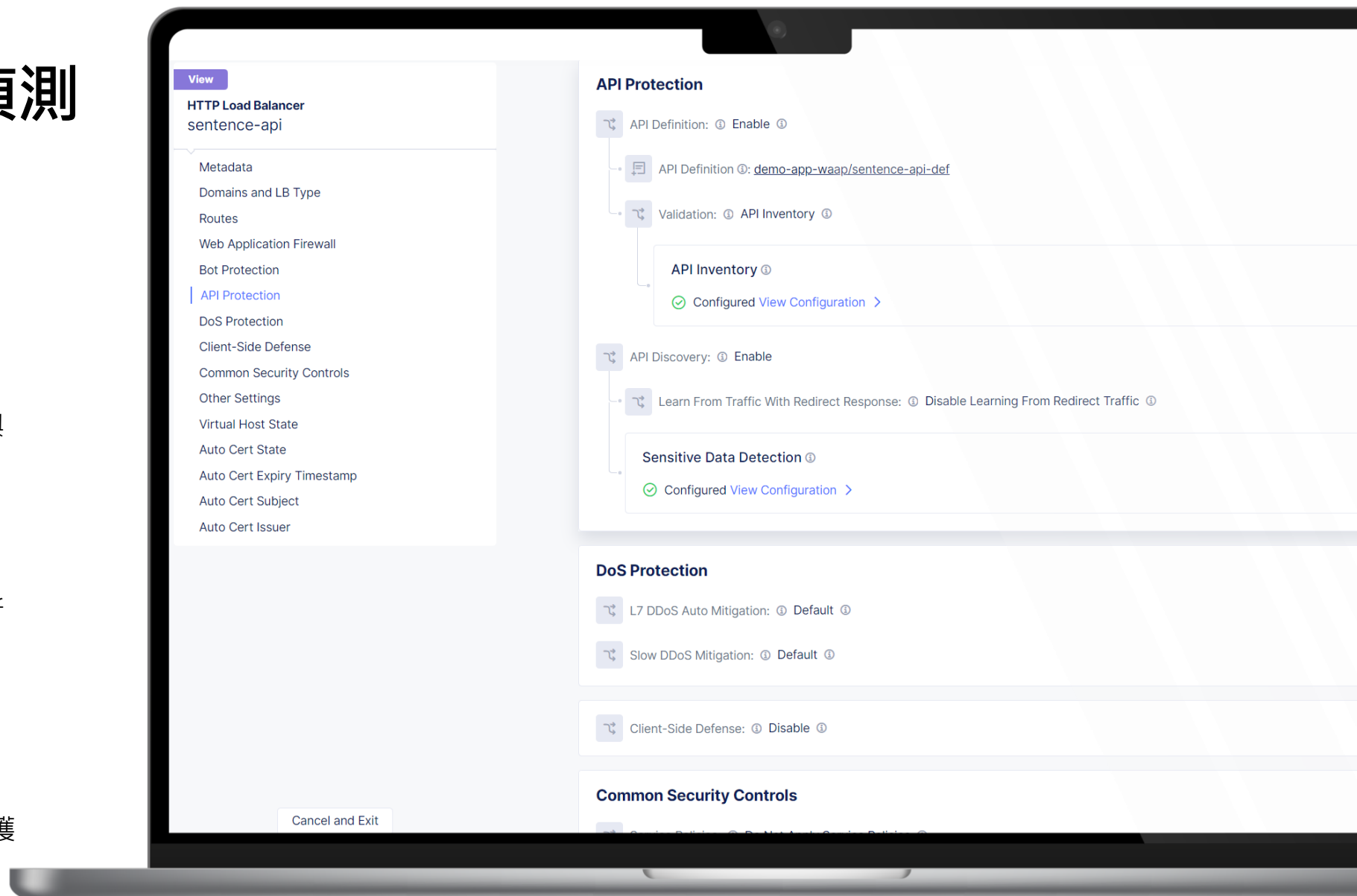
API 端點細節資訊與安全態勢

28 © 2025 F5

API 防護新層次： F5 提供從預防到偵測 的完整機制

透過即時執行的防護與控制機制，確保 API 行為正確，避免濫用與攻擊風險

- API 上線前測試：在進入執行階段 (runtime) 之前，先行驗證其正確性與安全性
- AI/ML 行為分析與異常偵測：運用人工智慧辨識異常行為與潛在威脅
- 持續流量監控與漏洞識別：即時分析 API 流量，偵測攻擊與弱點
- Schema (OAS) 驗證：以正向安全模型比對 API 請求與定義一致性
- API 執行階段的全面流量防護：涵蓋 WAF、API 安全規則、限流與資料防護等機制

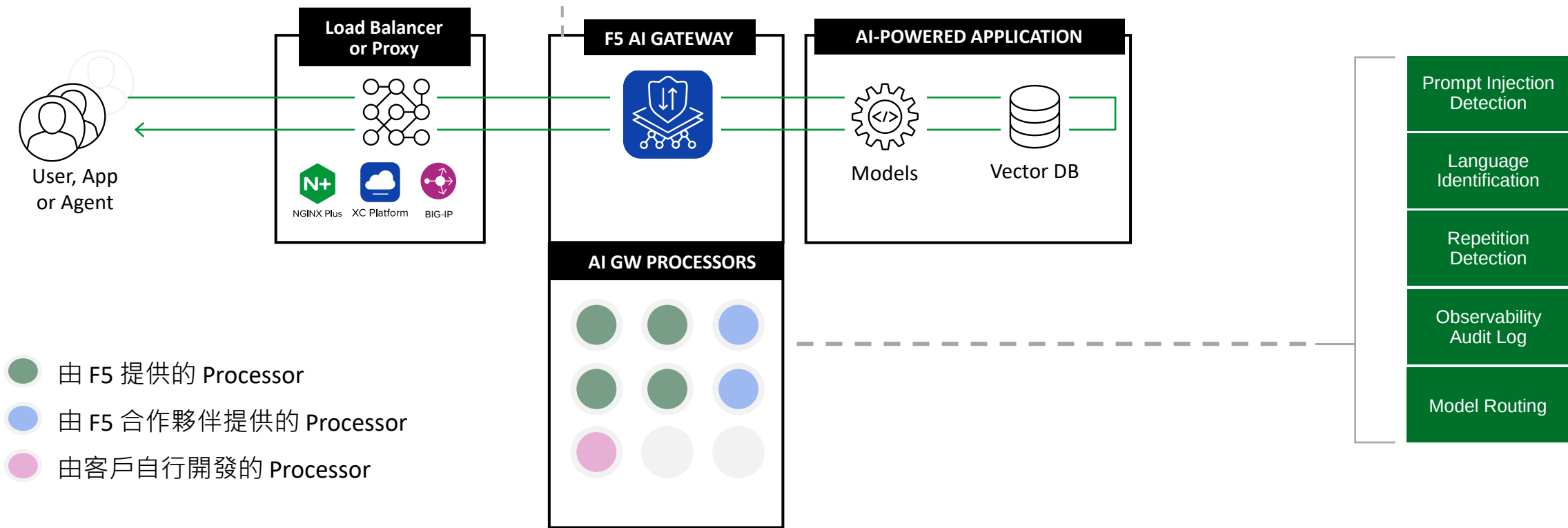


全新 AI 解決方案：F5 AI Gateway

全面應對 AI 應用在模型安全、易用性與語意風險等關鍵挑戰

整合與支援

- 支援 Kubernetes 部署
- 提供 Processor SDK (Python)
- 可整合 OpenAI API、Azure OpenAI、Anthropic、Ollama 等主流 LLM 平台



AI in F5 : AI 助理驅動 API 流量管理與安全事件智能處理

The screenshot displays the F5 Multi-Cloud App Connect interface, specifically the Security Analytics section. The left sidebar shows the navigation menu with options like Overview, Performance, Discovered Services, Manage, and Security. The main content area is divided into several panels:

- Security Analytics:** A bar chart showing 17 items for the event type 'waf_sec_event' on Nov 21. Below the chart is a table with columns 'Time' and 'Country,city'.
- NGINX One:** A panel showing the configuration for the 'silentcartographer' instance. It includes a file tree on the left and a configuration editor on the right.
- AI Assistant:** A floating window that provides explanations and guidance for the configuration snippets.

Table Data (Security Analytics):

Time	Country,city
21 Nov 22:20:13	CH,Rumlang
21 Nov 22:20:12	CH,Rumlang
21 Nov 22:20:12	CH,Rumlang

NGINX One Configuration Snippet:

```
server {
    # pass the PHP scripts to FastCGI server listening on 127.0.0.1:9000;
    #location ~ \.php$ {
    #    root           html;
    #    fastcgi_pass   127.0.0.1:9000;
    #    fastcgi_index  index.php;
    #    fastcgi_param  SCRIPT_FILENAME /scripts$fastcgi_script_name;
    #    include        fastcgi_params;
    #}

    # deny access to .htaccess files, if Apache's done
    # this for the httpd's security, it must precede all other
    # location directives.
    #location ~ /\.ht {
    #    deny all;
    #}

    # enable /api/ Location with appropriate access
    # to make use of NGINX Plus API
    #location /api/ {
    #    api write-on;
    #    allow 127.0.0.1;
    #    deny all;
    #}

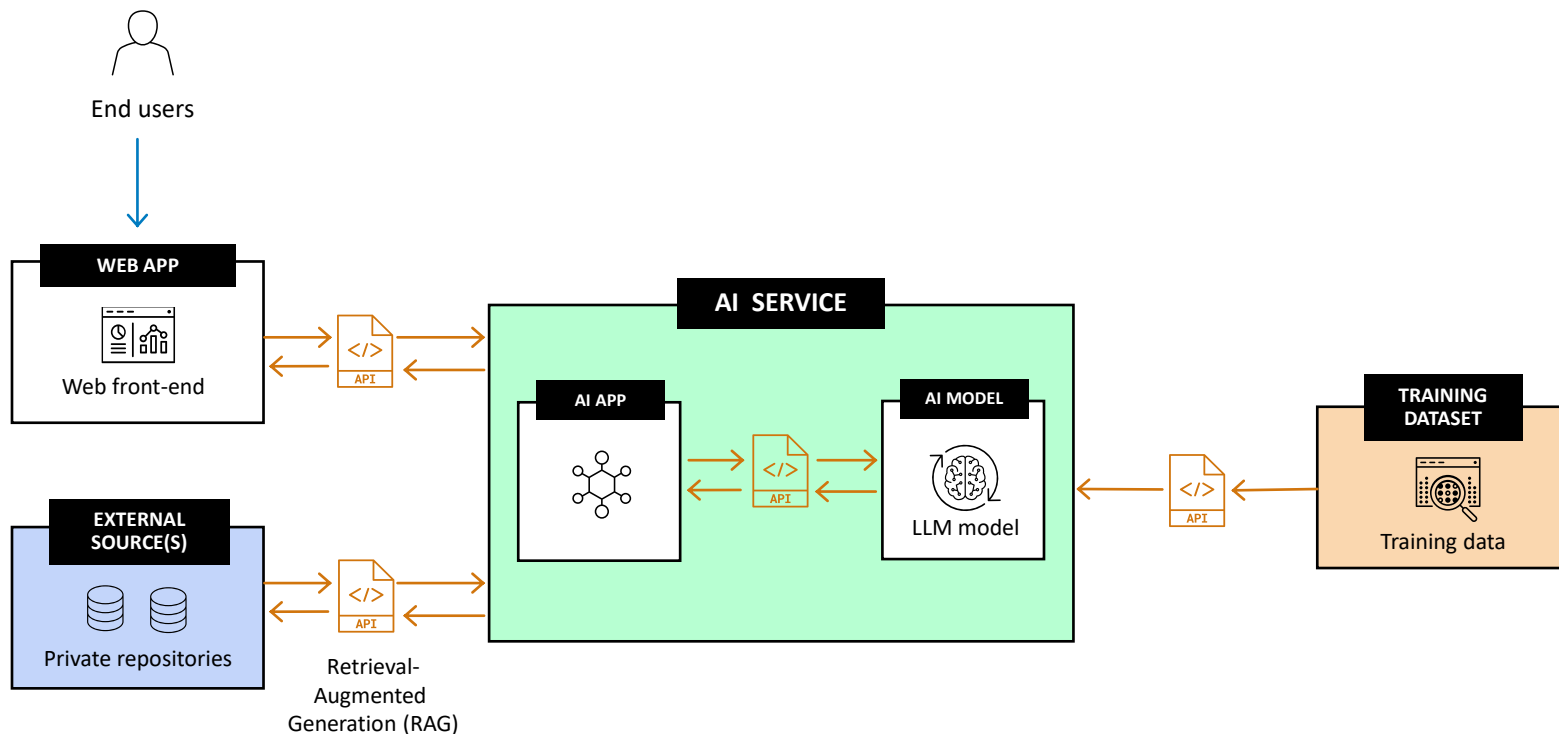
    # enable NGINX Plus Dashboard; requires /api/ Location
    # enabled and appropriate access control for remote
    # access.
    #location = /dashboard.html {
    #    root /usr/share/nginx/html;
    #}
}
```

AI Assistant Explanations:

- Explain this NGINX configuration snippet:** The AI Assistant explains the purpose and guidance for the configuration snippet, highlighting the security implications of the API access control.
- Purpose:** This NGINX configuration snippet enables the NGINX Plus API for read/write access, allowing dynamic reconfiguration of the server, such as managing upstream servers on-the-fly or key-value storage. It also restricts access to the API to only the localhost address (127.0.0.1) for security purposes.
- Guidance:** To modify this snippet, consider adjusting the IP addresses allowed to access the API based on your specific network requirements. You can add more IP addresses to the allow directive if needed, and you can remove or modify the deny all directive if you want to allow access from other sources as well. However, it is crucial to ensure that only authorized administrators have access to the NGINX Plus API for security reasons.

F5 AI 安全架構：全面守護 API 與 AI 應用

從流量防護、API 閘道，到語意層的 AI 防線，F5 建構橫跨應用與模型的整合式安全架構



要保護 AI 應用，需要涵蓋以下防護要素

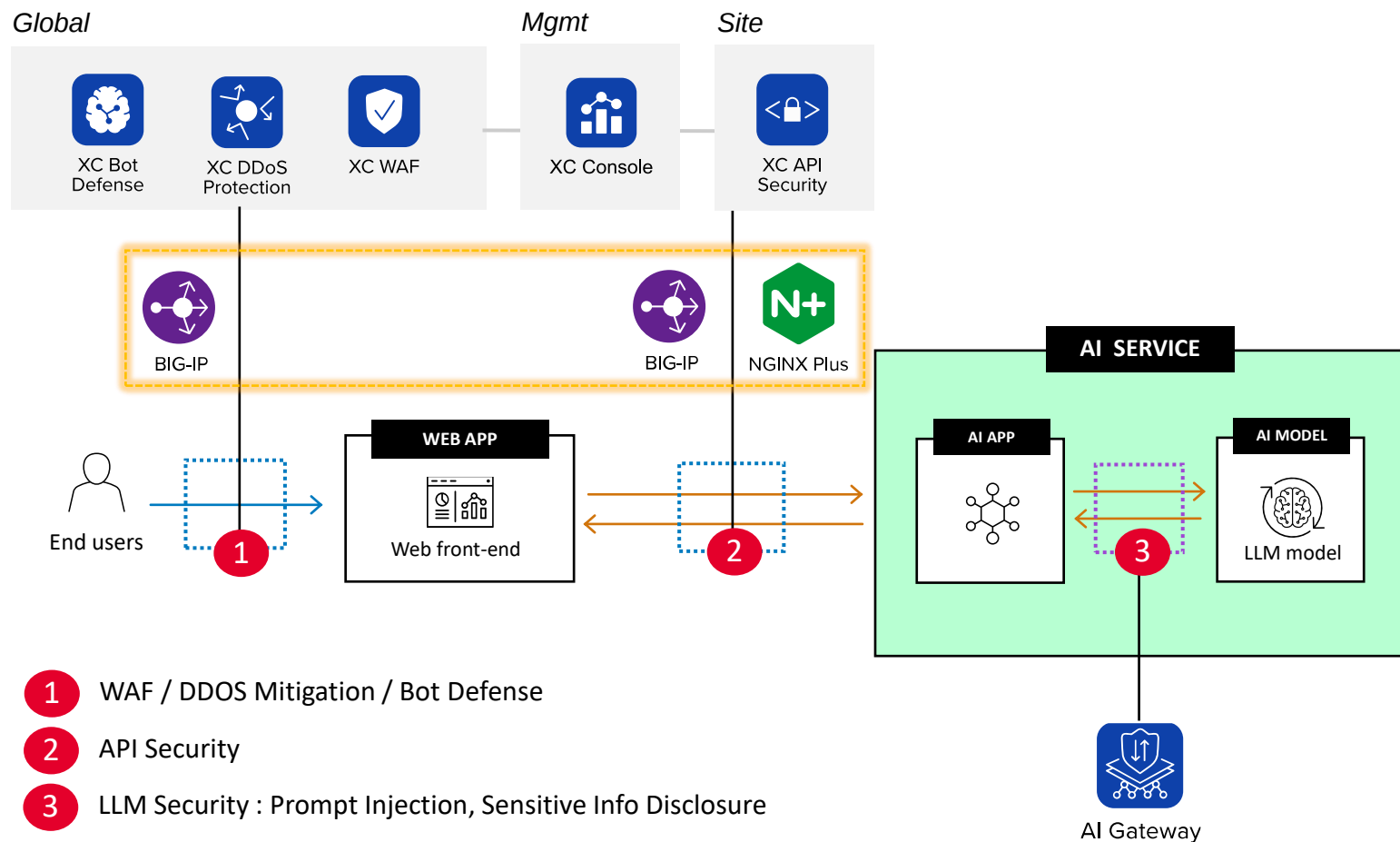
- Web Application Firewall (WAF)
- DDoS Mitigation
- Bot Defense
- API Security
- AI/LLM Security

F5 提供的安全解決方案

- WAAP Solutions
- API Security
- New : F5 AI Gateway (Security)

F5 AI 安全架構：全面守護 API 與 AI 應用

從流量防護、API 閘道，到語意層的 AI 防線，F5 建構橫跨應用與模型的整合式安全架構



要保護 AI 應用，需要涵蓋以下防護要素

- Web Application Firewall (WAF)
- DDoS Mitigation
- Bot Defense
- API Security
- AI/LLM Security

F5 提供的安全解決方案

- WAAP Solutions
- API Security
- New : F5 AI Gateway (Security)

