Stop AI based Attacks with AI

Roman Borovits
Solution Engineering, f5

Agenda

AI for business

New attack surface, new threats

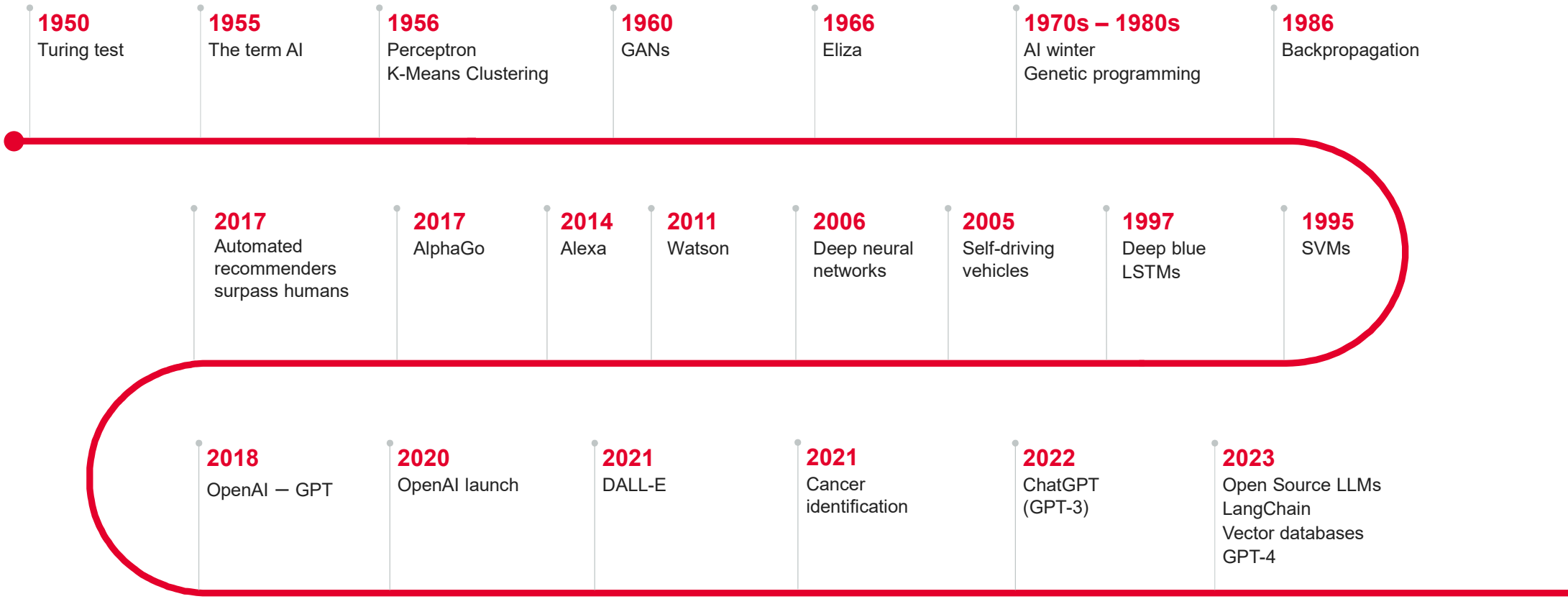
Leveraging AI for Security

Protecting your AI applications

AI Gateway

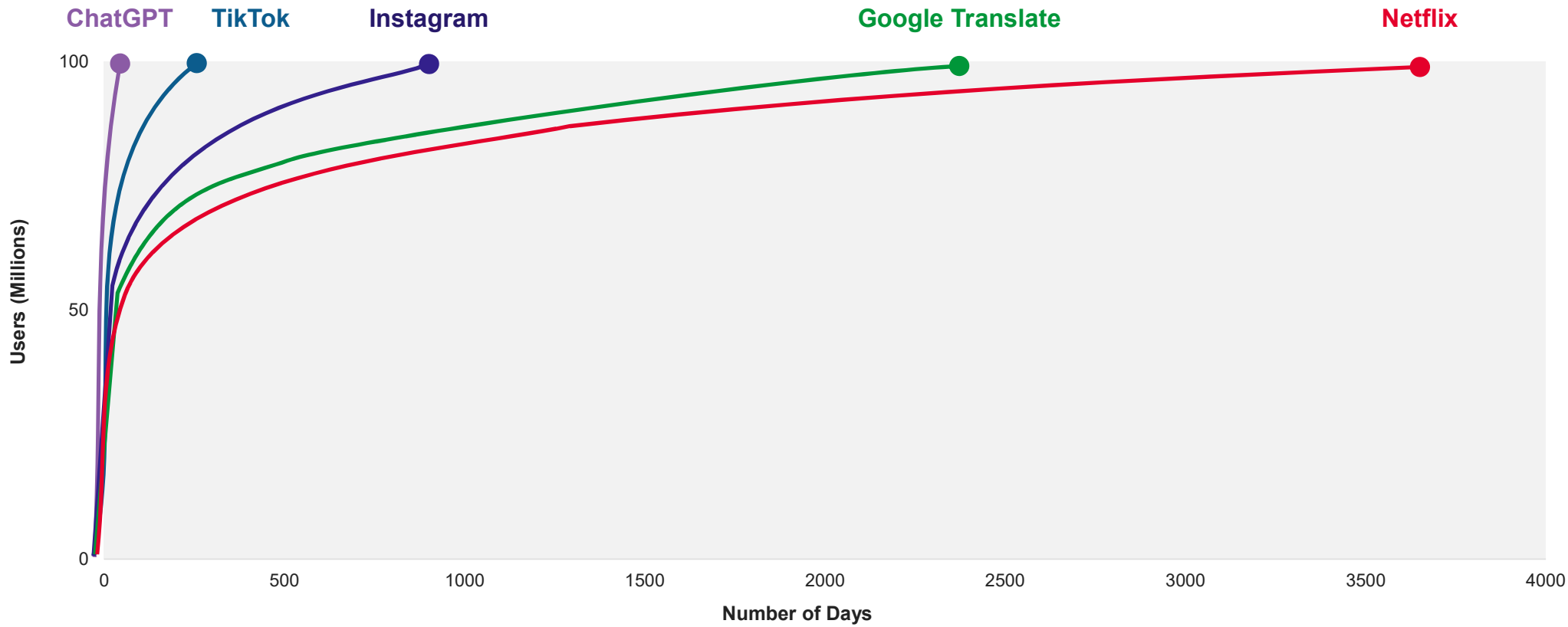
AI is not new

History of AI



The pace of adoption of Generative AI has been astounding

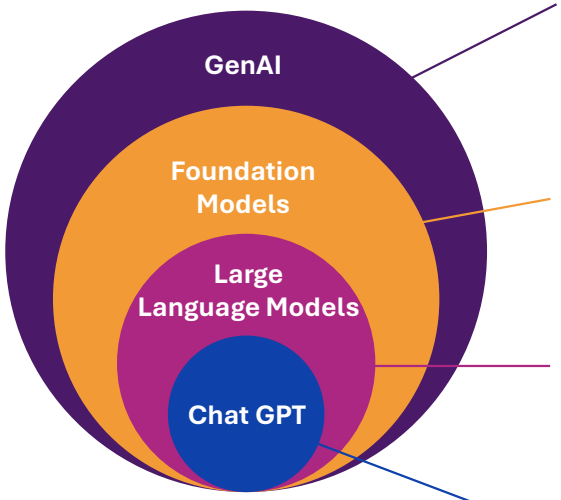
Time it took companies to reach 100 million users:



Sources: Global X ETFs with info derived from: BBC News. (2018, Jan 23). Netflix's history: From DVD rentals to streaming success; Cerullo, M. (2023, Feb 1). ChatGPT user base is growing faster than TikTok. CBS News.

Generative AI and Large Language Models (LLM)

Setting the context

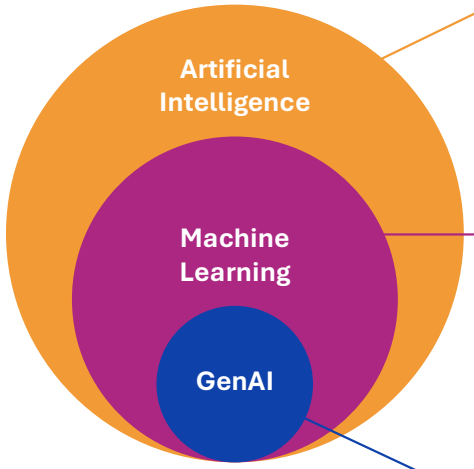


A form of machine learning which uses large quantities of data to train models to produce content

A general purpose model trained on large quantities of data

A model trained on large quantities of text to produce plausible content

A publicly available service which provides a chat interface to an LLM



Technology which emulates human or beyond-human intelligence

A form of AI which uses data to train models to make predictions or produce insights

A form of machine learning which uses large quantities of data to train models to produce content

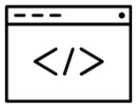
AI for business

Generative AI has democratised technology

Anyone, technical and non-technical, can unlock its power

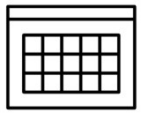
1,000+

Plugins available with **GPT-4**



Create an app when you've never coded before

Appeal an insurance denial



Write Excel formulas

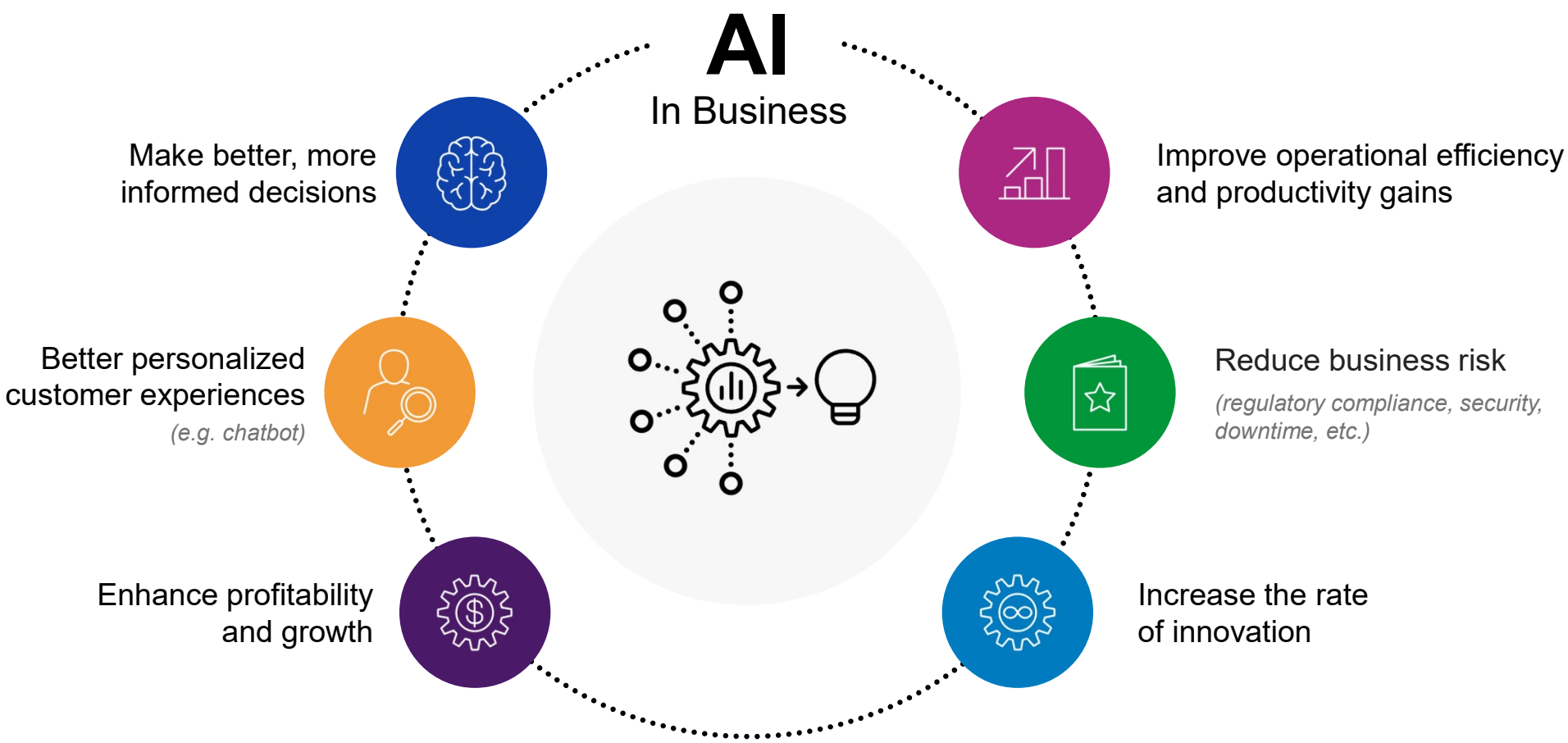
Design a personal shopping assistant

Build new games

Write marketing blogs

Source: OpenAI

AI is a general Purpose Technology



Adoption of AI in business is starting to ramp up

75%

of organisations consider AI a core business focus

AI models are

Hybrid

with 40% of orgs deploying AI models on premises and 65% in public clouds

AI budgets expected to grow

94%

from 2024 to 2026

Most businesses consider

Security and Compliance

as one of the top challenges that complicate deployment of AI models and applications

Source: F5 2024 State of Application Strategy Report

New attack surface, new threats

For cyber security, AI is a double-edged sword

Creates new attack surfaces and new attack capabilities

Moving at alarming speed, no reservations

Enhances cyber security

Exercising caution with deferment to human expertise

AI – With Rewards Come Risks

By 2025, Gartner predicts **sustainable and ethical use** of AI will be among their top concerns for enterprises[1]

AI's **delusional hallucinations** can augment reality - wreaking havoc and breaking trust

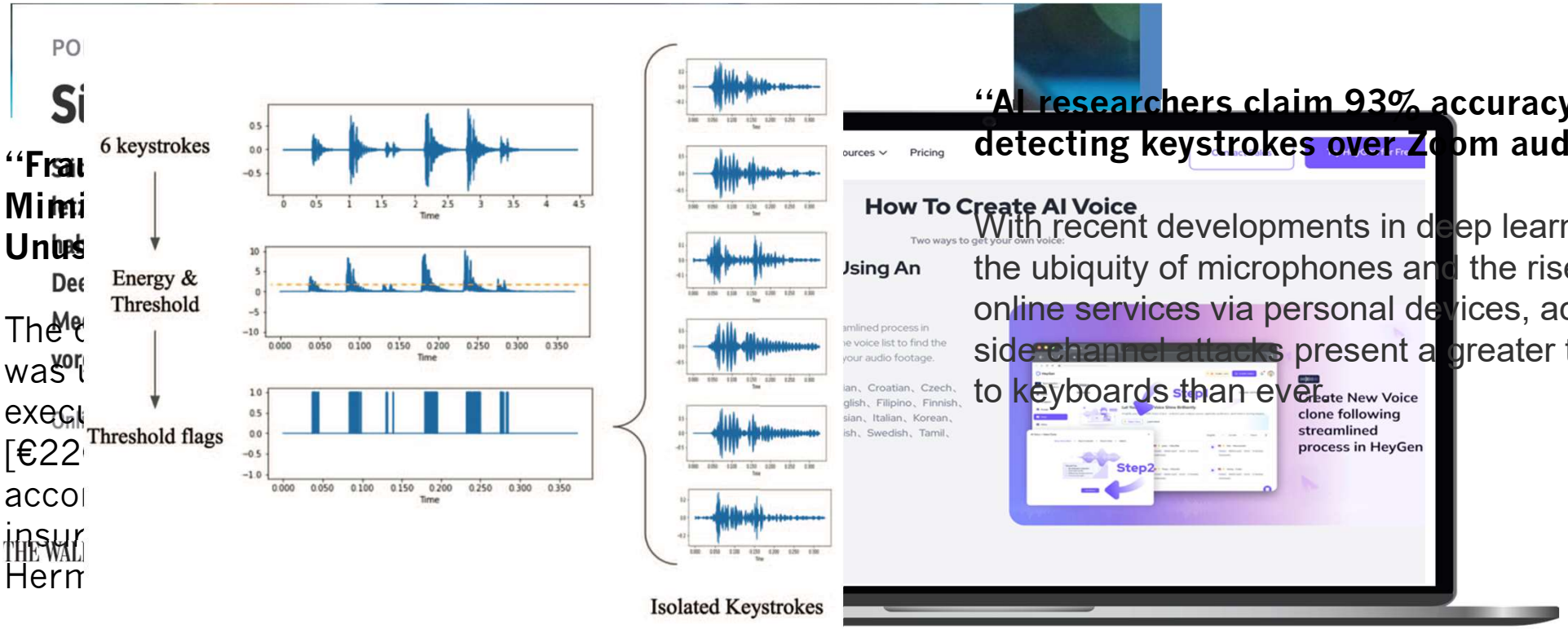
Input manipulation attacks – **attacker alters input** to manipulate models

Lack of visibility and control – too much autonomy can result in unintended consequences

Hackers leverage AI to conduct attacks - **AI-generated phishing attempts and social engineering scams** easily bypass traditional security measures

Model inversion attacks – **attackers reverse engineer a model** to extract information

Deepfakes, Vishing, Side Channel



“AI researchers claim 93% accuracy in detecting keystrokes over Zoom audio”

With recent developments in deep learning, the ubiquity of microphones and the rise in online services via personal devices, acoustic side-channel attacks present a greater threat to keyboards than ever.

Leveraging AI for Security

Use AI to Stop Automated Attacks

Key Areas:

- **Learning**
- **Pattern detection & recognition**
- **Analysis & Conclusions**
- **Mitigation & Policy Building**
- **Assist the Admin**

Traditional machine learning models and analytics are a core capability of F5's security offerings.

Primarily used as insight analytics that inform new rules and policy postures in response to patterns or anomalies identified by the ML models.

AI is not new to F5

2006
Policy Builder to automatically create WAF security policy

2011
Machine Learning (ML) to optimise BIG-IP network stack

2012
ML-based learning mode to improve WAF efficacy

2014
Behavioural anomaly analysis to improve WAF efficacy

2015
ML in iHealth to identity customer BIG-IP issues and misconfigurations

2016
Behavioural DoS to mitigate attacks using dynamic signatures

2019
Stage 2 retrospective analysis to deter bots and automated attacks

2020
Shape acquisition AI powered anti-bot and fraud

2021
Dynamic API Discovery to mitigate risk from rogue APIs

2021
False Positive Suppression to reduce burden on SecOps teams

2022
Malicious User Detection to automatically mitigate bad actors

2023
API Endpoints Risk Score to measure risk of discovered APIs

AI Assisted F5 Security Solutions

DDoS Mitigation – L7 attack detection and mitigation

Bot Defense without CAPTCHA – AI detects most modern Bots without CAPTCHA

WAF & API Protection – Signature Tuning, Malicious User Detection, Fingerprinting

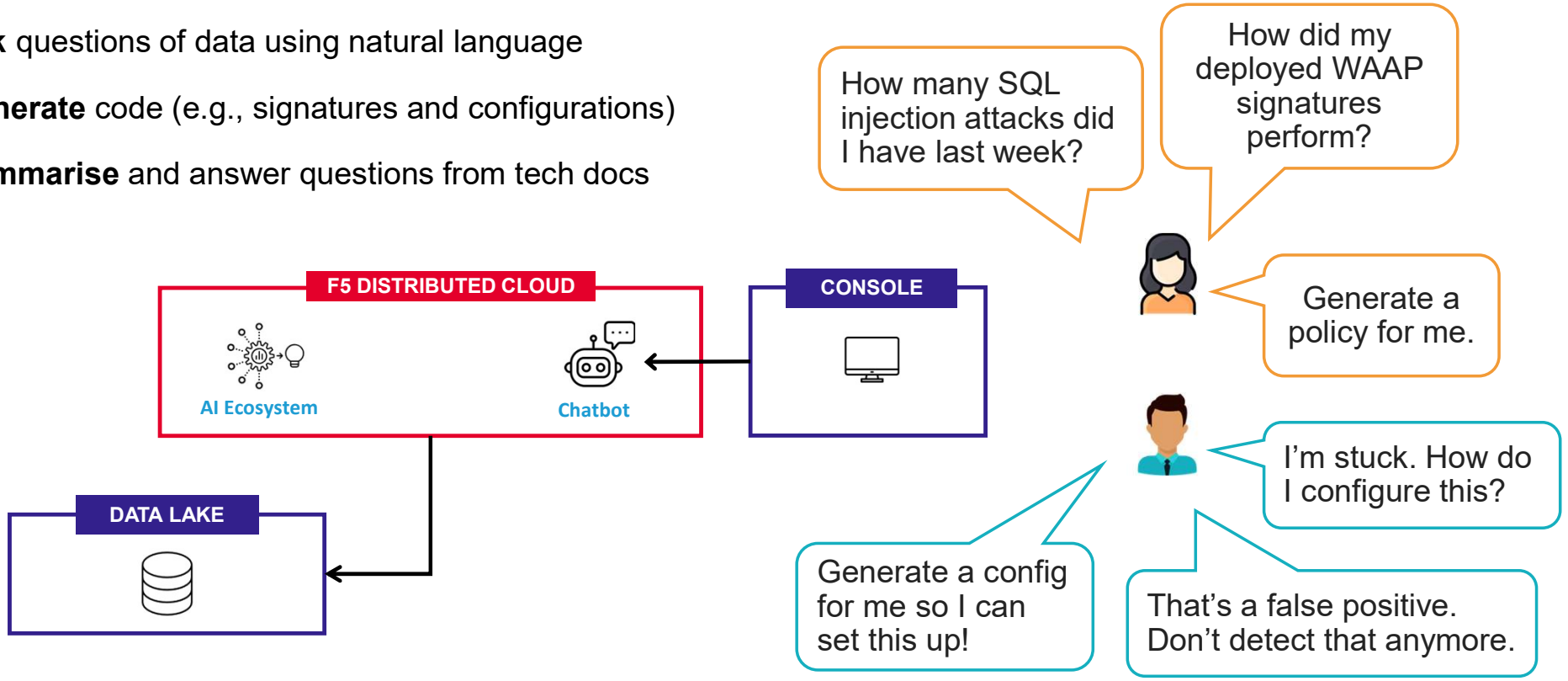
Behavior Based Security – Good and Bad Traffic Pattern recognition

API Security – Discovery & Self Learning Capabilities

Security made ridiculously easy with Generative AI

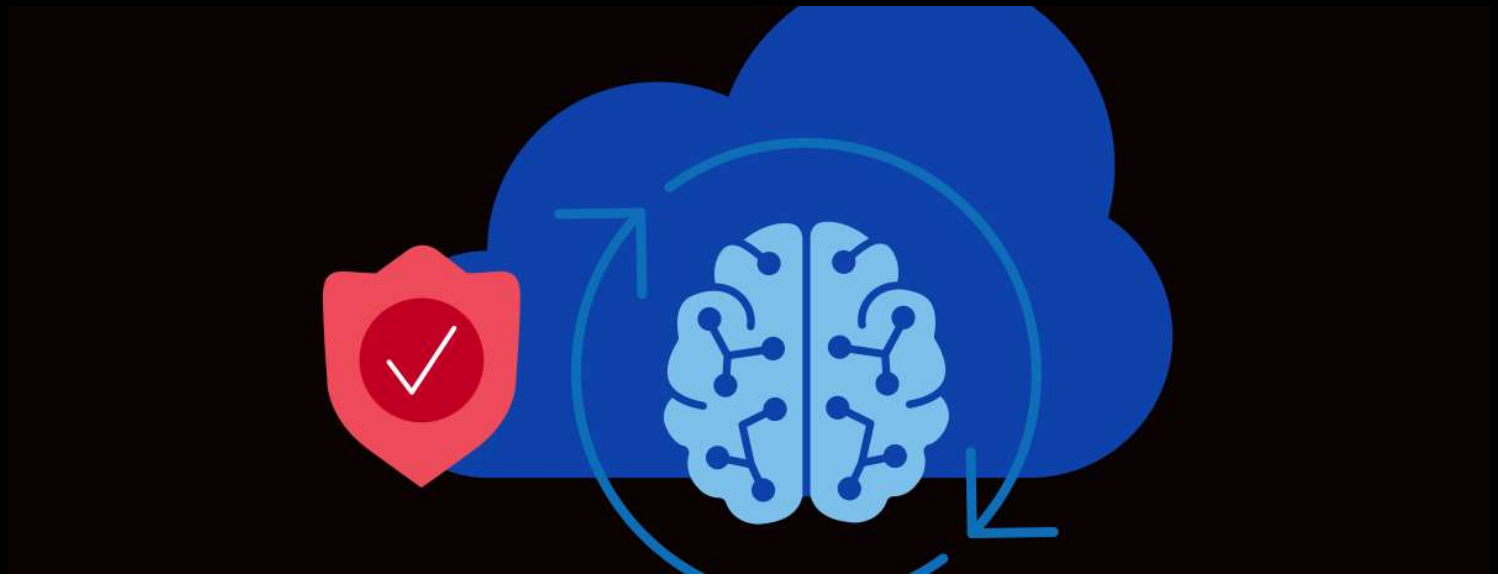
The Art of the Possible: Simplify operations and increase detection efficacy with Large Language Models

- **Refine** ML models or signatures to reduce false positives
- **Ask** questions of data using natural language
- **Generate** code (e.g., signatures and configurations)
- **Summarise** and answer questions from tech docs



Protecting your AI Applications

“AI workloads are the most modern of modern apps”

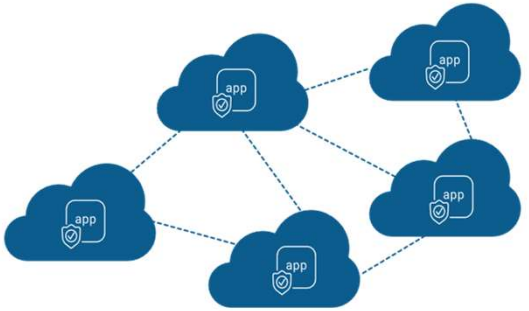


AI workloads

The most modern of modern apps

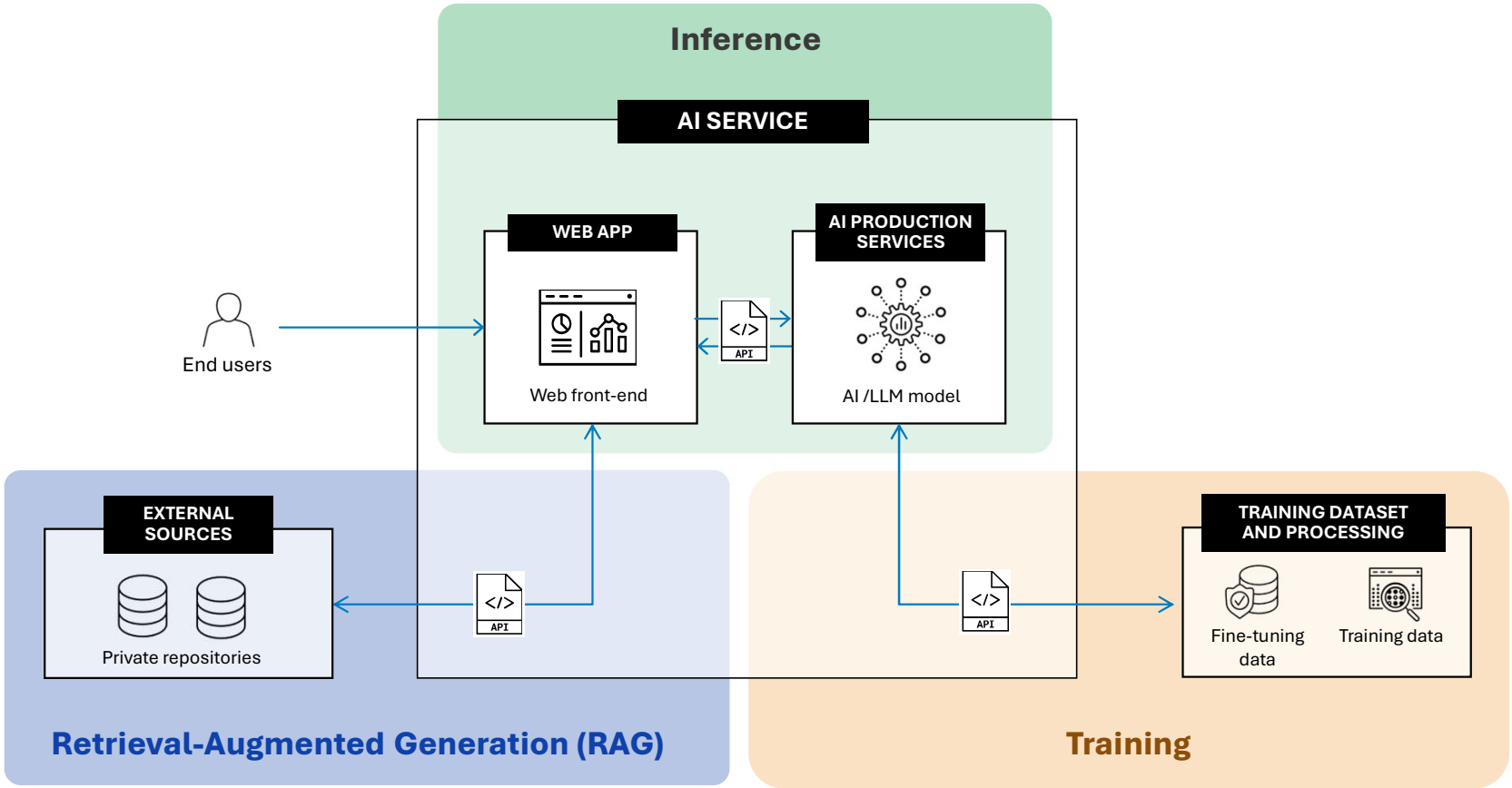
The Generative
AI Platform

The Connection for
AI Components

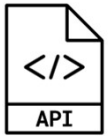


Workloads Across
Distributed Environments

Training, Inference, and Retrieval-Augmented Generation (RAG)



From web applications, to APIs, to AI



AI / LLM Apps

App Security

WEB APPLICATION FIREWALL

DDOS MITIGATION

BOT DEFENCE

API SECURITY

OWASP Top 10 API attacks
API discovery and scanning
API governance and compliance
API runtime protection

AI SECURITY

OWASP Top 10 LLM attacks
Prompt injection
Data poisoning
Model theft

App Delivery

APPLICATION DELIVERY CONTROLLER (ADC)

CONTENT DELIVERY NETWORK (CDN)

API GATEWAY
API MANAGEMENT

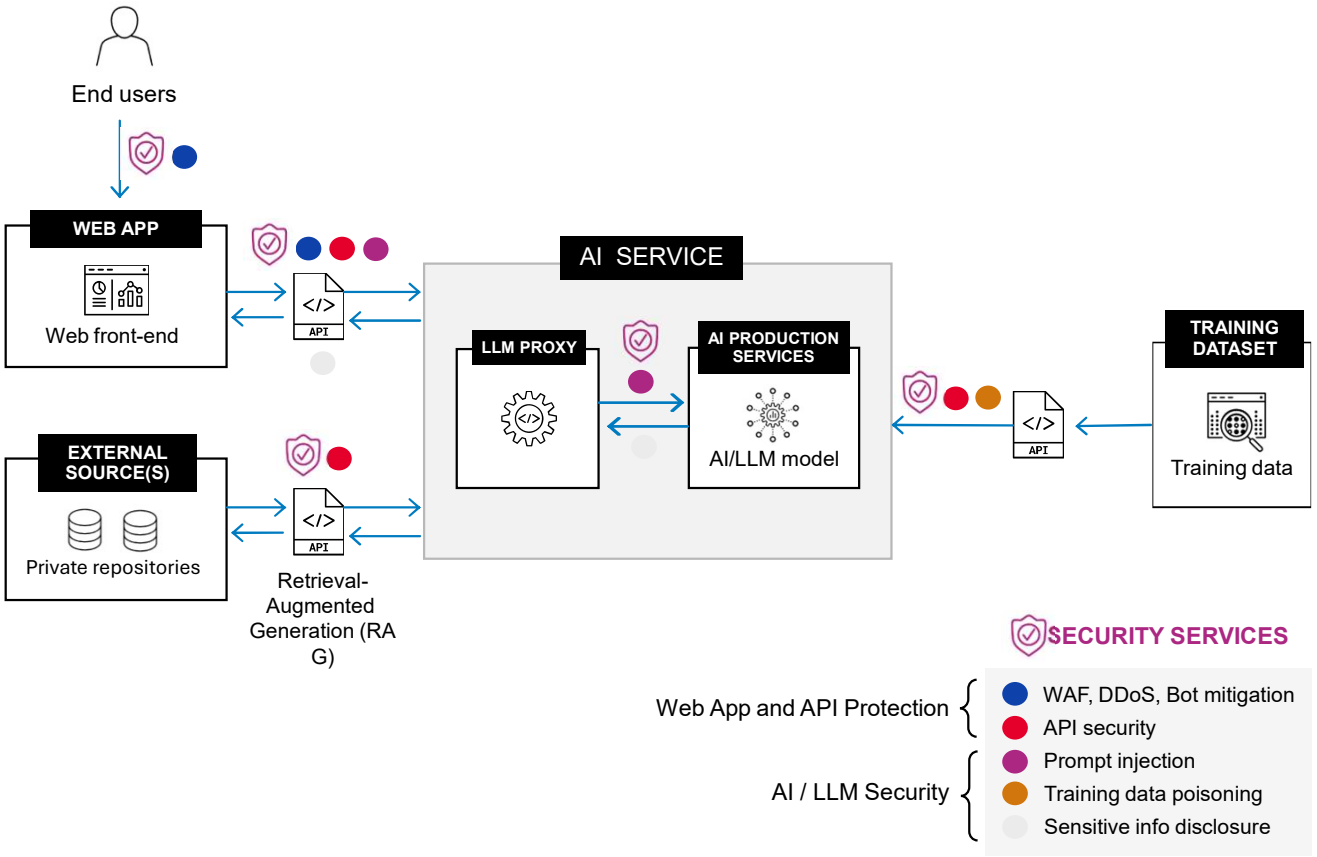
API request routing
Dev portal
Rate limiting and accounting

AI / LLM PROXY

Deep LLM integrations
Prompt management
Rich observability
Token counting

Security challenges in the AI ecosystem

WAAP protection suite required, plus AI security tools



app

Protecting AI apps requires

- Web Application Firewall (WAF)
- DDoS mitigation
- Bot defence
- API security
- AI/LLM security

Recommended approach

- Consolidated solution
- Centralised management and visibility (single pane of glass)

OWASP Top 10 for LLM applications and generative AI

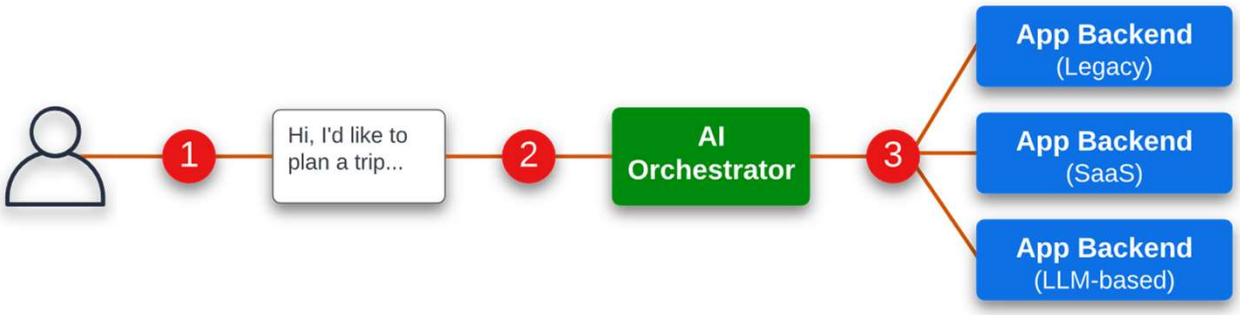
Generative AI/LLM apps introduce new threat vectors

LLM01 Prompt Injection	**LLM02** Insecure Output Handling	**LLM03** Training Data Poisoning	**LLM04** Model Denial of Service	**LLM05** Supply Chain Vulnerabilities
LLM06 Sensitive Information Disclosure	**LLM07** Insecure Plugin Design	**LLM08** Excessive Agency	**LLM09** Overreliance	**LLM10** Model Theft

Source: Top 10 for LLMs and Generative AI Apps, OWASP Foundation

AI Gateway

What is an AI gateway?



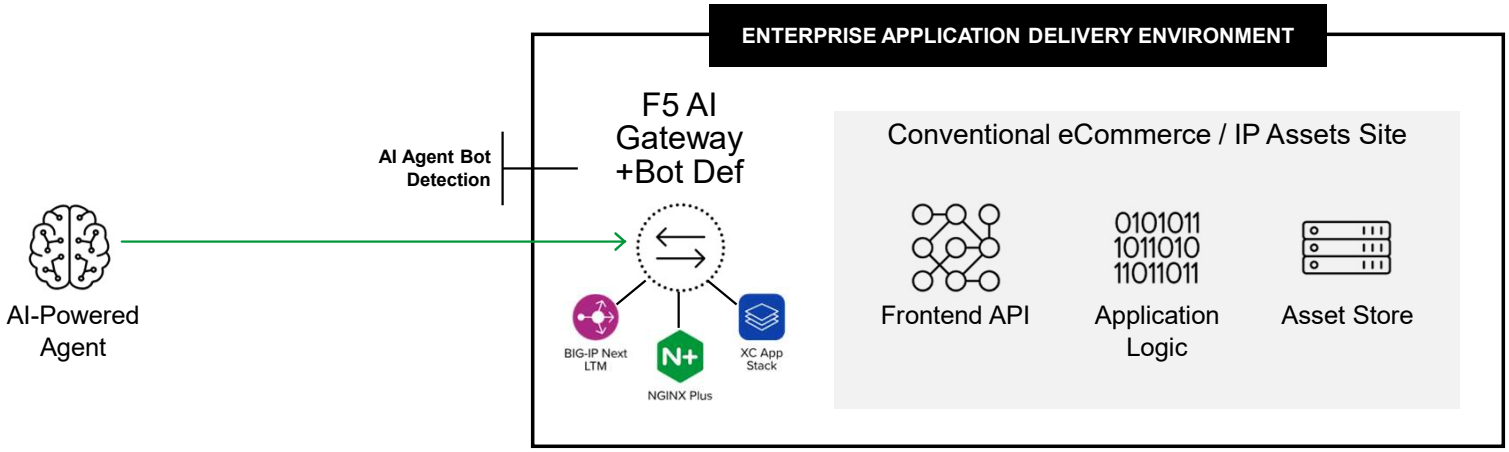
AI Gateway Platform Features \ Insertion points	1 AI Web Outbound	2 AI Orchestrator Inbound	3 AI Service Inbound
PII Data Leakage Mitigation	✓	✓	✓
API & Schema Discovery			✓
GW services for AI Orchestration			✓
Data Leakage Prevention	✓	✓	✓
Hallucination Detection/Mitigation	✓	✓	✓
Bias Detection/Mitigation	✓	✓	✓
Abuse Detection/Mitigation	✓	✓	✓
Data Authorization (access control)	✓	✓	✓
Cost Management	✓	✓	
Observability	✓	✓	✓
Sovereignty, Provenance, Fake Detection	✓	✓	✓
Prompt Injection Mitigation	✓	✓	
Other plugins...			

AI Gateway Platform Mandate = All features must be implemented with flexibility and modularity in mind, so they fit where we need them.

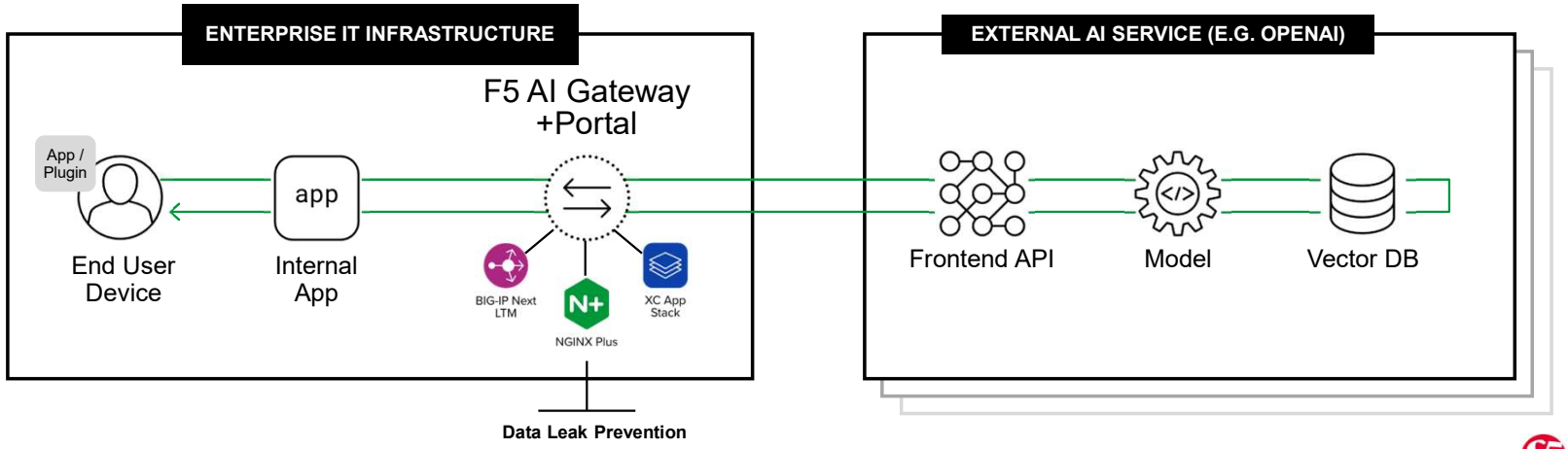
AI Gateway Use Cases

Beyond Enterprise AI-Powered Applications

Prevents enterprise assets from being scraped for AI/LLM purposes. Enables safe publishing of photos, docs, assets not appearing in public Gen AI output.



Aggregator of external AI services (public or private), enabling enterprise users to use AI chatbot tools safely and cost-effectively



AI workloads

Example:

NGINX providing model security (API and encryption/authentication) for Intel OpenVINO AI Model Server…

… on ARM-based IPUs for security segmentation and control.

NGINX is a built-in part of OpenVINO.

Alliances with AI technology companies further bolster F5 as an AI leader

Securing AI requires careful planning

New threat vectors

Protecting AI apps requires a security platform that can offer all existing app security tools (WAAP) plus additional protection against new attacks

Distributed architecture

Protecting an AI ecosystem requires a solution that allows to easily and safely interconnect all the AI system components across data center, cloud and edge

Evolving requirements

Look for solutions that allow you to evolve your architecture (inferencing, training, RAG, federated learning, etc.)