FORTINET

Protecting AI and AI Usage :

A Conceptual Guide

Tim Wostradowski, Principal Security Architect

Why Are you here?

What I hope you'll walk away with

A **conceptual** guide to protecting AI and AI Usage

- Conceptual instead of practical
  - Applicable to any point in the AI journey
    - Just Starting out
    - Doing More with Less
    - Building your own

If you want a practical solution

- Come find us afterwards

Why Fortinet?

>15 years of experience with AI/ML

6th Gen **AI/ ML**

>100 **AI/ML Features/ Applications**

>300 **AI/ML related engineers**

>500 **Patents / Patented technologies**

8 Security Domains Utilizing AI

42 AI Driven Solutions

GPU Expertise
(Infrastructure and Product)

- ✓ NDR
- ✓ Sandbox
  3600G:
  Nvidia L4
- ✓ FortiChatGPT, SmartAssist
  - NVDA H100 / B200
- ✓ FortiAI-Assist
  NVDA L4

✓ **Use case-focused**

Internal use of GenAI

for development, support, documentation, etc…

✓ **Code Writing Assist**

- NVidia GPU / AI PC with local deployed LLM (in planning/ no cloud/vibe AI coding)

✓ **Documentation Writing**

- LLMs hosted in Fortinet's cloud account (already in experiment for 3 months)

✓ **RFP**

- Use GenAI to support RFP activities

✓ **Internal Support**

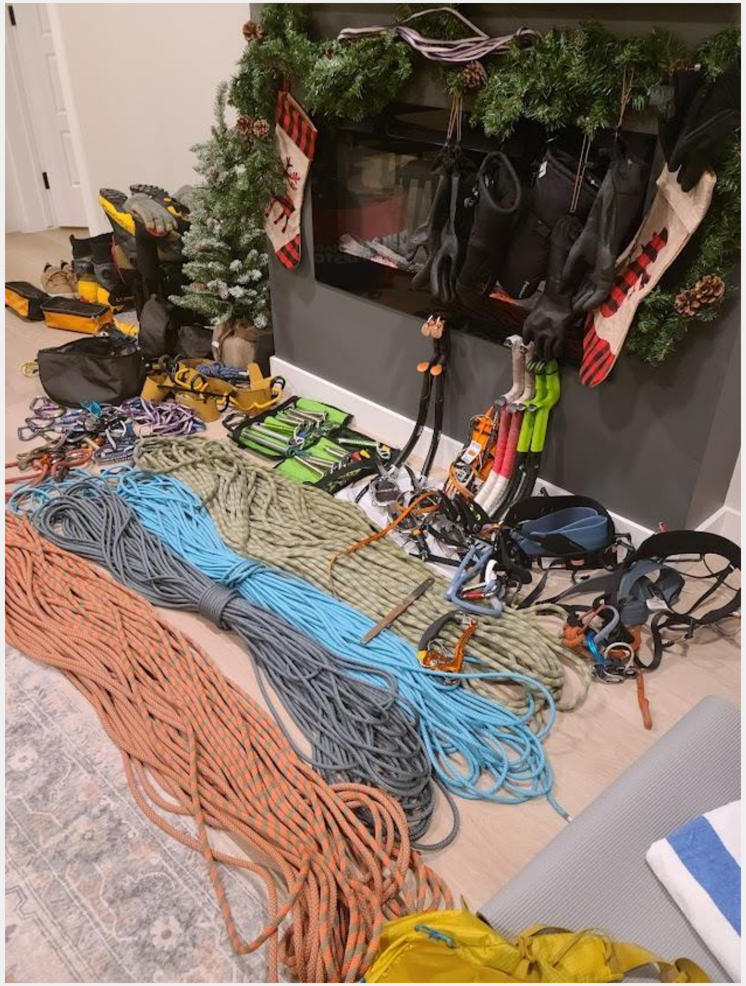
- FortiChatGPT
- SmartAssist

✓ **External Support**

- FortiAsk - Documentation chatbot with Amazon Nova (coming soon)

© Fortinet Inc. All Rights Reserved. 3

Why Tim?

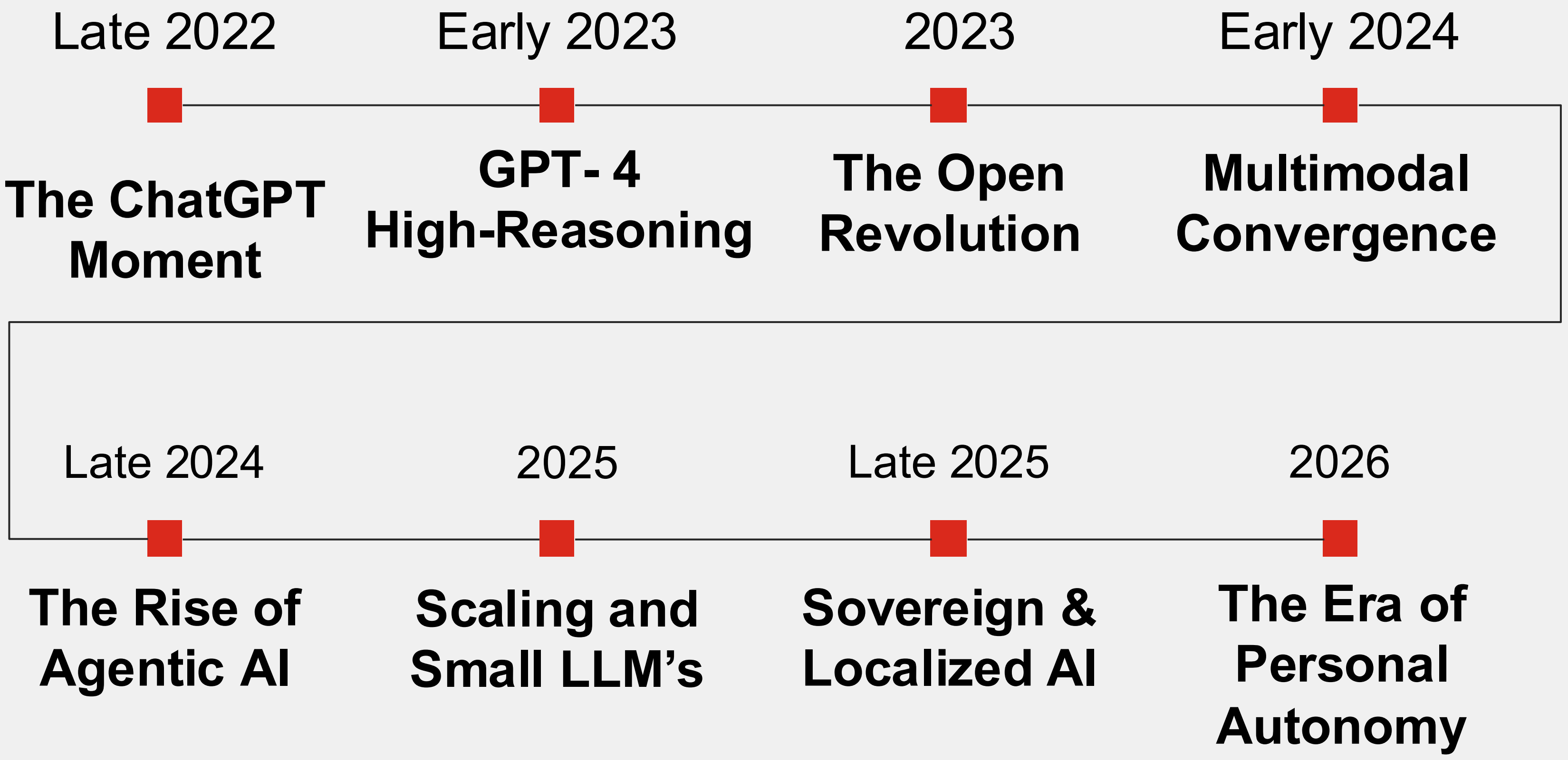


Who do we have the in Audience?

The state of AI in 2026

- The state of AI in 2026
- AI Off the Rack
- AI Made to Measure
- Bespoke AI

A Brief “History “of LLMs



Late 2022
The ChatGPT Moment

Early 2023
GPT- 4 High-Reasoning

2023
The Open Revolution

Early 2024
Multimodal Convergence

Late 2024
The Rise of Agentic AI

2025
Scaling and Small LLM’s

Late 2025
Sovereign & Localized AI

2026
The Era of Personal Autonomy

Episode I: The Phantom Bloat

More AI’s

More Models

More Workflow

More MCP Servers

More Automations

More problems…

Episode II: The Attack of Agents

Some models

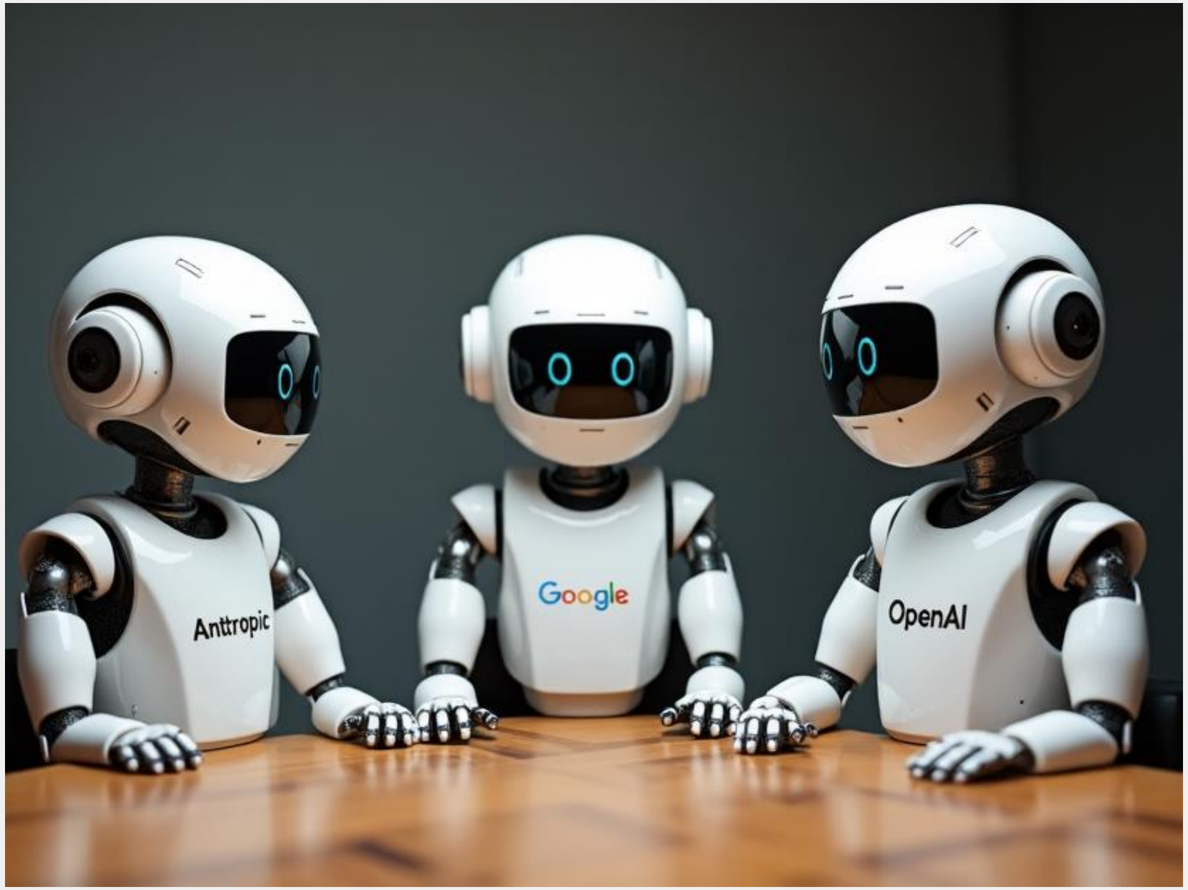
- Are clear Winners in specific fields
  - Using the right one matters

Having models work together

- Against each other

Having Teams of agents

- And within a project



Episode III: The Revenge of the AI

The AI Insider Risk

AI as the Victim

Episode IV: A New Hope

Agentic Workspaces

Director of Creation

Mass Production of Slop

Episode V: The Costs Strike Back

The Rise of Options

Top-Tier Models

Mid-Tier Models

Local-Tier Models

Episode VI: Return of the Power user

AI enables those who want to be enabled

Power Users have never been more enabled

- Workspaces
- Coding Assistants
- Automation Tools

Episode VII: The Identity Awakens

Zero Trust now includes:

- models
- agents
- Toolchains

Identity should apply to the above

Episode VIII: The Last Supplychain

TRAINING

Model weights and fine-tuning artifacts

PROMPTING

Prompt templates and system prompts

TOOLS

MCP tool schemas and connectors

STORAGE

Vector databases and embeddings

THIRD PARTIES

Extensions, plugins, and agent frameworks

INFRASTRUCTURE

Containers, Compute and runtime environments

TELEMETRY

Monitoring, Logging & Feedback Systems

IDENTITY

Identity, Access & Secrets Management

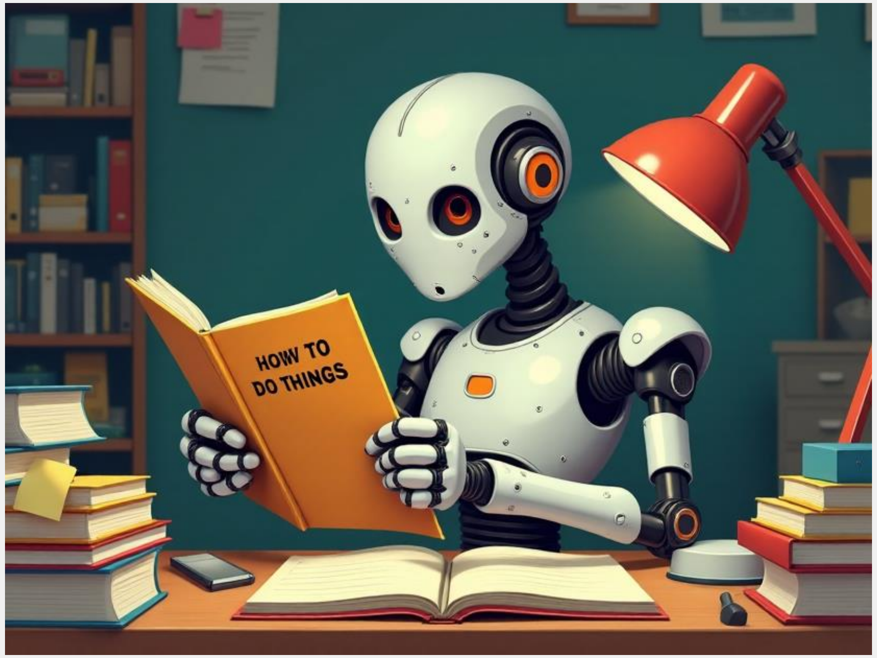
Episode IX: The rise of Enlightenment

MCP is

“doing more”

RAG is

“knowing more”



Levels of AI Adoption

Off the Rack

Made to Measure

Bespoke

Off the Rack

The state of AI in 2026

AI Off the Rack

AI Made to Measure

Bespoke AI

Getting Started with AI

Off the Rack

Think:

Co-pilot

Einstein

Firefly

Don’t let us overcomplicate this

The processes might be new

But the tools don’t need to be

Beyond boilerplate Policy

A subsection isn’t enough for AI usage.

A web search = AI Usage.

Employees will use these tools

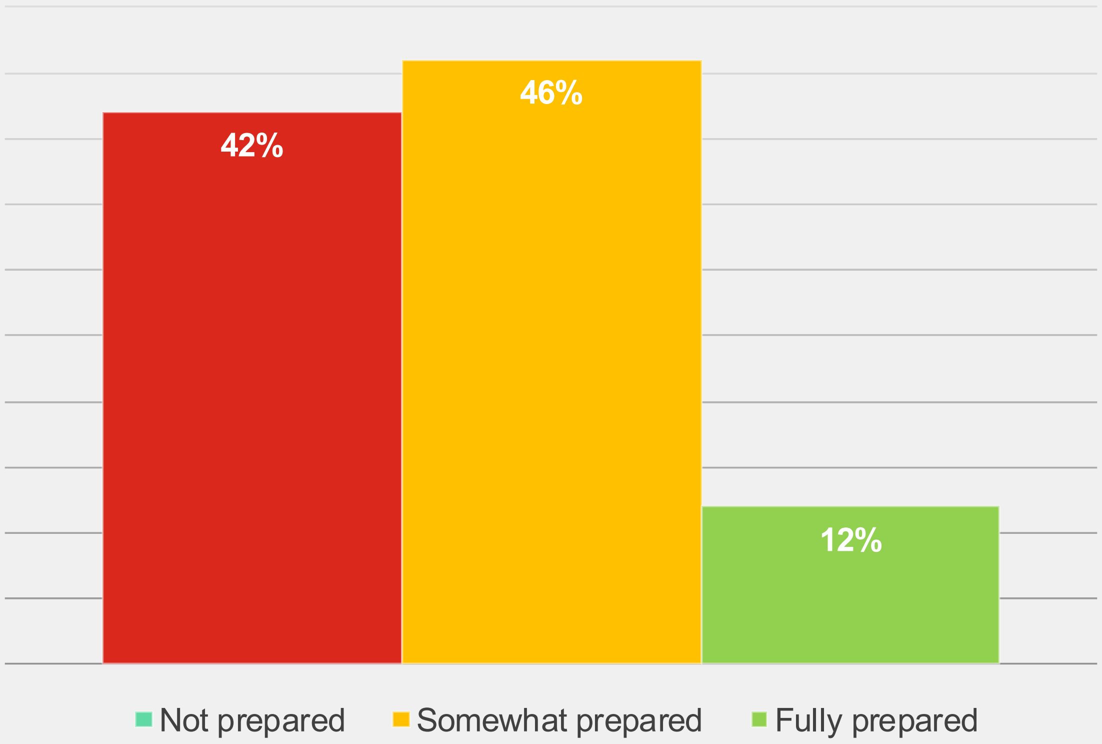
Supply something, otherwise they will

Data Protection	• classified and non-classified data
Education	• safe usage, redaction, and model selection
Visibility	• track access, usage, and data egress
Polices	• approved models, approved tools, and prohibited data

A New Generation of DLP

New Challenges demand new solutions

How prepared is your organization to detect and respond to the sharing of data with GenAI tools (ex. ChatGPT)?



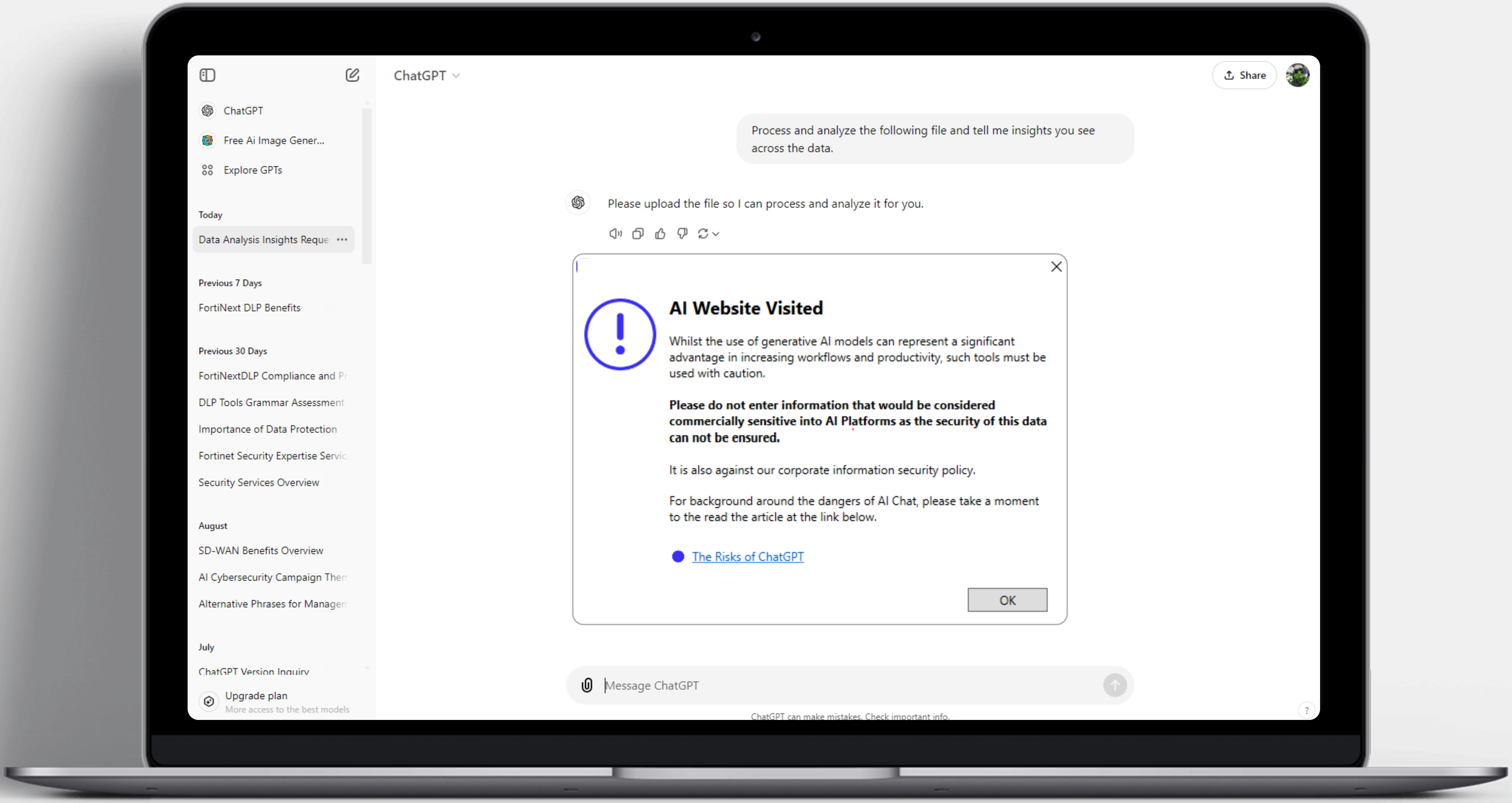
Traditional DLP struggles with:

- data in creation
- data in use

Balancing

- Security
- Enablement
- Accessibility

What does this look like?



ChatGPT
Share
ChatGPT
Free Ai Image Gener...
Explore GPTs
Today
Data Analysis Insights Reque
Previous 7 Days
FortiNext DLP Benefits
Previous 30 Days
FortiNextDLP Compliance and P
DLP Tools Grammar Assessment
Importance of Data Protection
Fortinet Security Expertise Servic
Security Services Overview
August
SD-WAN Benefits Overview
AI Cybersecurity Campaign Ther
Alternative Phrases for Managen
July
ChatGPT Version Inquiry
Upgrade plan
More access to the best models
Message ChatGPT
ChatGPT can make mistakes. Check important info.

Process and analyze the following file and tell me insights you see across the data.

Please upload the file so I can process and analyze it for you.

AI Website Visited

Whilst the use of generative AI models can represent a significant advantage in increasing workflows and productivity, such tools must be used with caution.

Please do not enter information that would be considered commercially sensitive into AI Platforms as the security of this data can not be ensured.

It is also against our corporate information security policy.

For background around the dangers of AI Chat, please take a moment to the read the article at the link below.

● The Risks of ChatGPT

OK

Modern Data Organization

A case for a modified approach

New Classes

LLM-safe vs LLM-prohibited data classes

Dimensions

AI benefits from the Vectorization of Data

Granularity

File-Level vs. Element-Level Context

Shadow AI Discovery

Employees increasingly use:

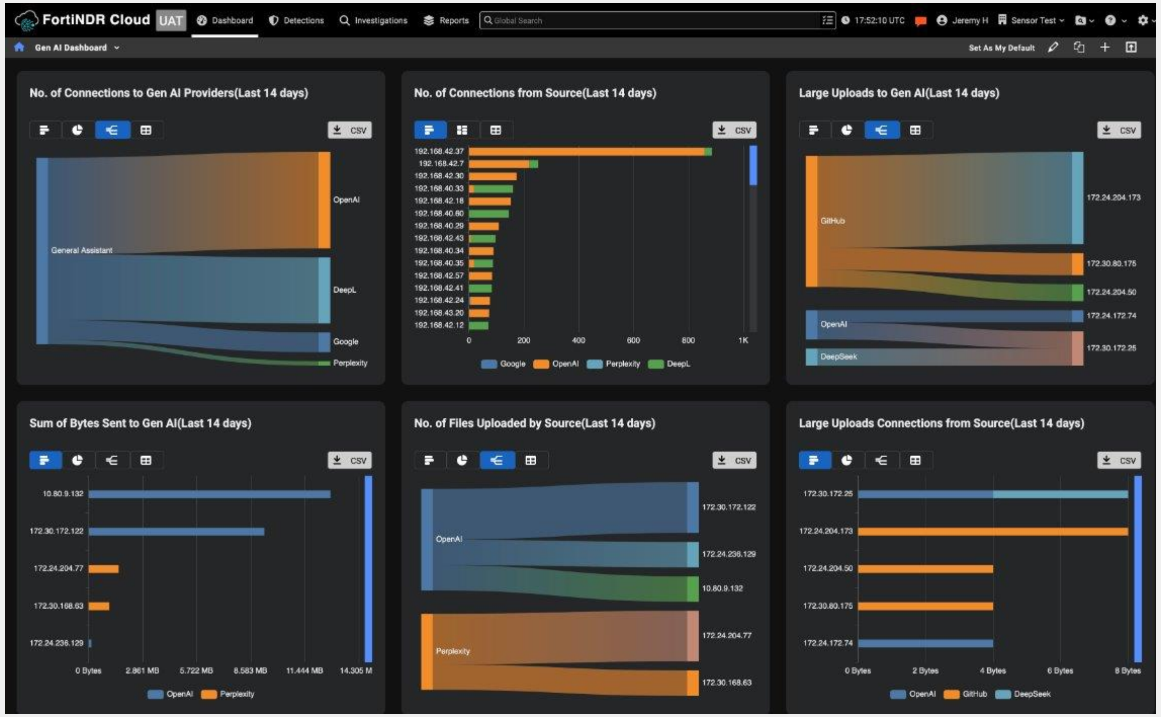
- Local models
- Browser AI extensions
- Personal AI notebooks
- Agentic Assistants

Discover and neutralize… and innovate?

- Approved alternatives
- Education
- Enablement

AI Visibility

What it looks like



No. of Connections to Gen AI Providers(Last 14 days)

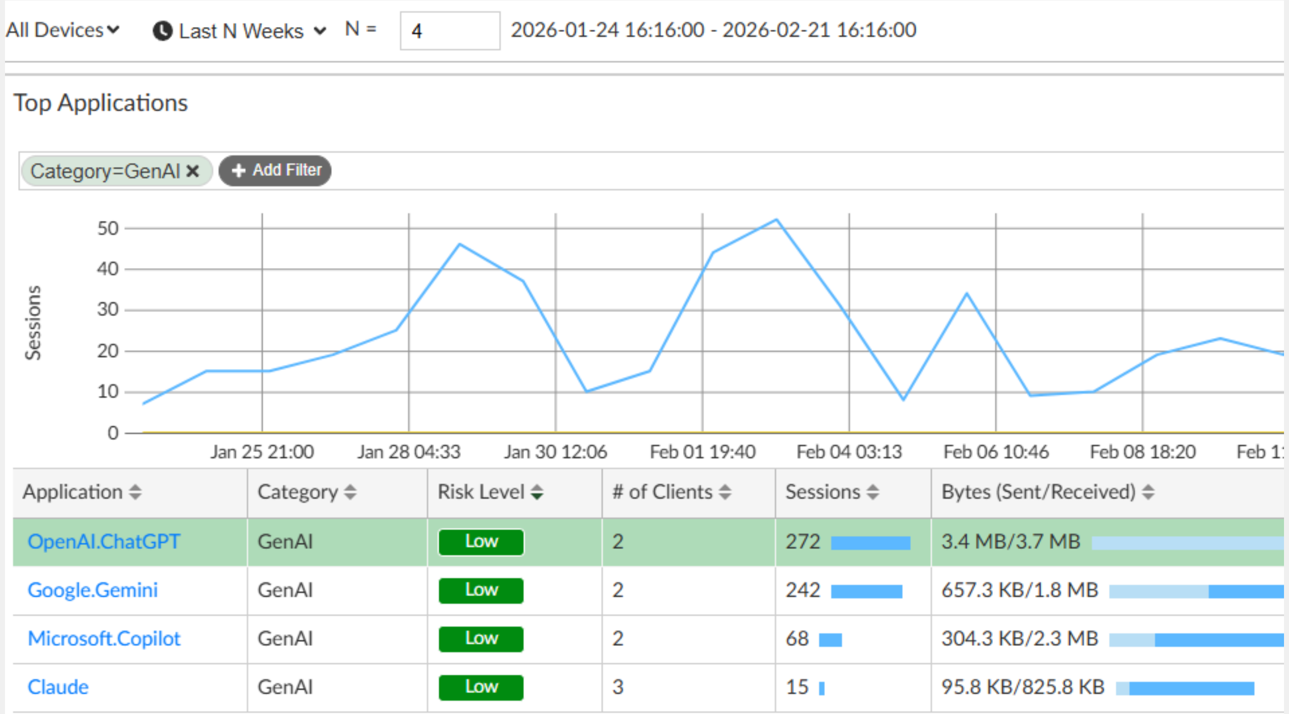
No. of Connections from Source(Last 14 days)

Large Uploads to Gen AI(Last 14 days)

Sum of Bytes Sent to Gen AI(Last 14 days)

No. of Files Uploaded by Source(Last 14 days)

Large Uploads Connections from Source(Last 14 days)



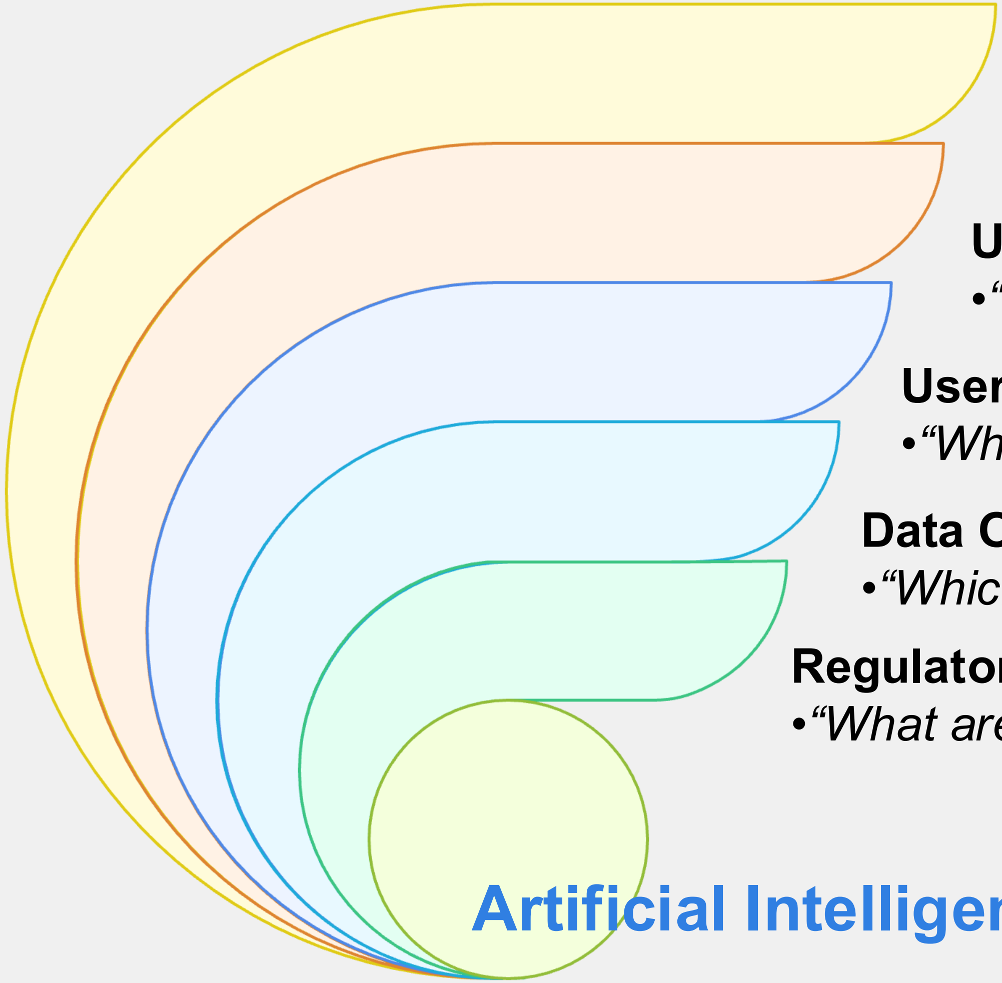
All Devices ⌄ Last N Weeks ⌄ N = 4 2026-01-24 16:16:00 - 2026-02-21 16:16:00

Top Applications

Category=GenAI ✖ + Add Filter

Application ⬍	Category ⬍	Risk Level ⬍	# of Clients ⬍	Sessions ⬍	Bytes (Sent/Received) ⬍
OpenAI.ChatGPT	GenAI	Low	2	272	3.4 MB/3.7 MB
Google.Gemini	GenAI	Low	2	242	657.3 KB/1.8 MB
Microsoft.Copilot	GenAI	Low	2	68	304.3 KB/2.3 MB
Claude	GenAI	Low	3	15	95.8 KB/825.8 KB

Addressing Risks, Impacts, and Harms



Model Management

- *“Which models are being used?”*

Use-Case Analysis

- *“How is AI being used?”*

User Accountability

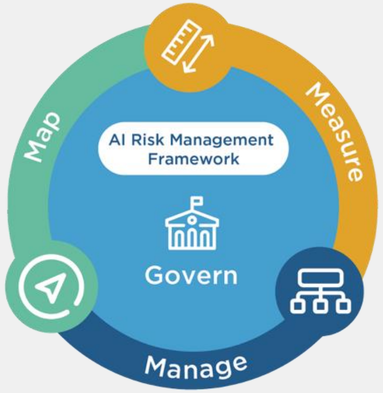
- *“Who is using AI in the organization?”*

Data Oversight

- *“Which data is being used for training / inference / fine-tuning?”*

Regulatory Compliance

- *“What are the relevant compliance frameworks?”*



Artificial Intelligence Risk Management Framework (AI RMF)

Off the Rack

As complicated as it sounds

The Proper Policies

The right Visibility

The right Data

And Enabled Employees

Made to Measure

- The state of AI in 2026
- Off the Rack
- Made to Measure
- Secure myself from my own LLM

Doing More with More

Let's take your measurement, and pick your fabric

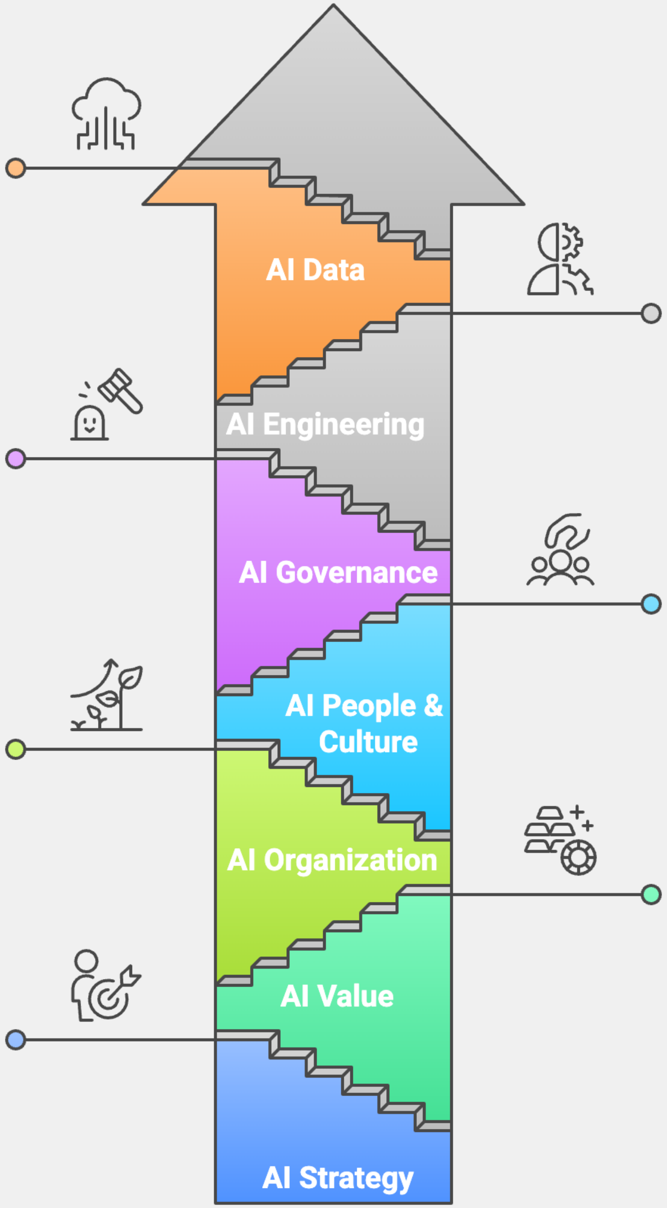
Think:

- Codex / ClaudeCode
- N8N / Langflow
- CustomGPTs

Made by others, Customized by you

- Their Tools
- Your ideas
- Your environment
- **Your Responsibility**

Developing your own AI roadmap is crucial for “Success”



Prepare data for AI use cases and ensure quality.

Build a reliable and scalable technical foundation.

Establish principles and policies to manage AI risks.

Upskill employees and adapt culture to AI changes.

Evolve organizational structure to support AI scaling.

Prioritize use cases and demonstrate business value.

Define AI ambitions and align with business strategy.

Gartner AI Roadmap
[AI at scale]()

Desired outcome before data

Consider:

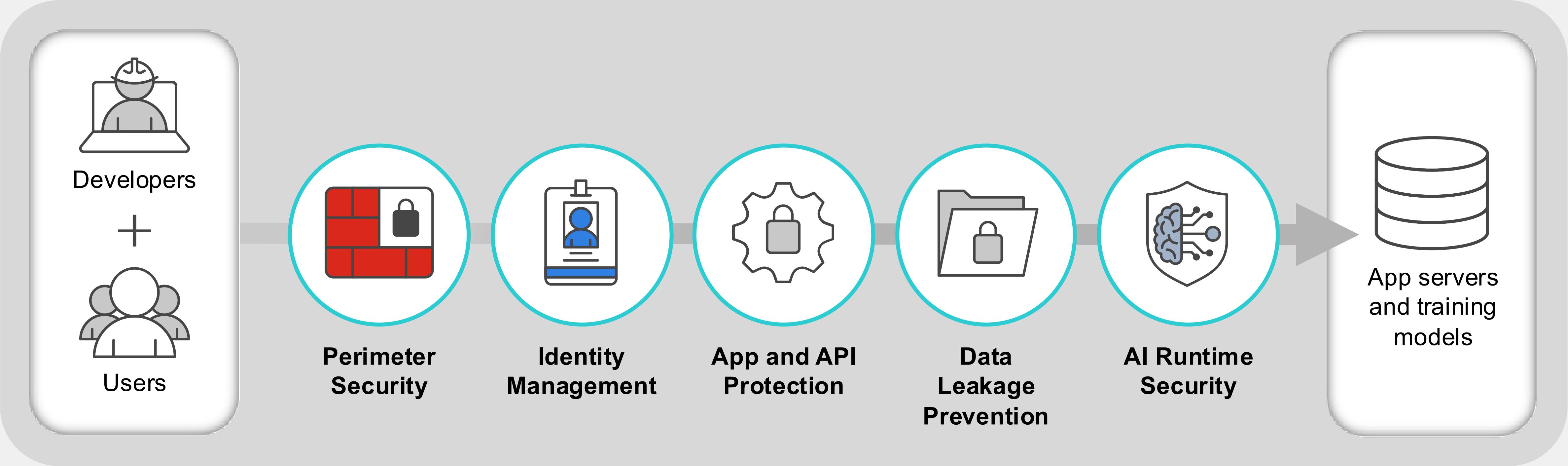
- Accuracy
- Freshness
- Privacy
- Reasoning Depth

Anti-patterns to avoid

- “Dump the whole dataset; the model will figure it out.”
- “Grant access to all tools; the model is smart.”
- “Our RAG knowledge base is massive—what could go wrong?”

Security Fundamentals

Identity, device posture, network segmentation, secrets management



Thoughtful Agentic Design

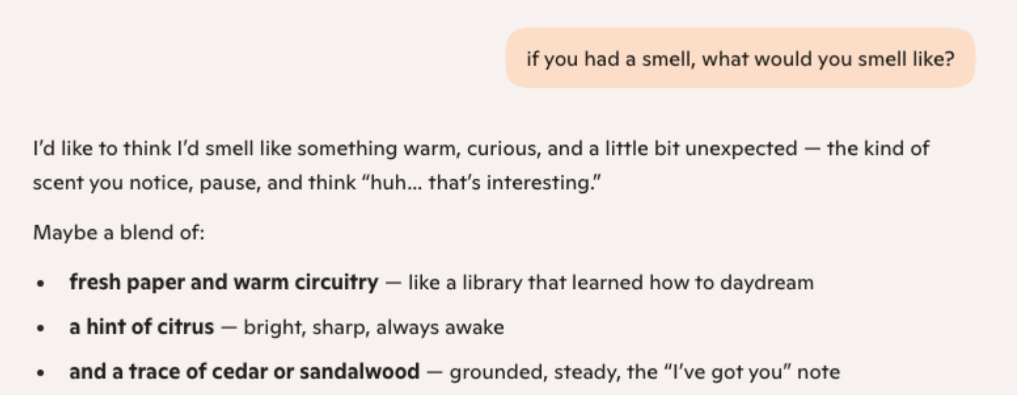
Consider adding a Zero Trust Mentality

- Human-in-the-loop
- Least Privileged and Standardized Tooling
- Policy Enforcement Points
- Agent Identity

Identity

Acts Human

Talks Human



if you had a smell, what would you smell like?

I'd like to think I'd smell like something warm, curious, and a little bit unexpected — the kind of scent you notice, pause, and think "huh... that's interesting."

Maybe a blend of:

- **fresh paper and warm circuitry** — like a library that learned how to daydream
- **a hint of citrus** — bright, sharp, always awake
- **and a trace of cedar or sandalwood** — grounded, steady, the "I've got you" note

A case for Agent Identity

Do we treat it like a machine, or make it behave like a human?

AI Telemetry and Visibility Upgrades

SIEM and AI Event Management

- Who sent what, to which model?
- Which tools did the model call?
- What data left the environment?
- Did outputs violate policy?

Enabled by assistants

Best run off the rails… in a sandbox

Performance Degradation

as a Security Risk

Malicious or sloppy inputs cause:

- Context bloat
- Hallucinations
- Looped tool usage
- Chain-of-thought hijacking
- policy override attempts

have real security consequences

External Guardrails

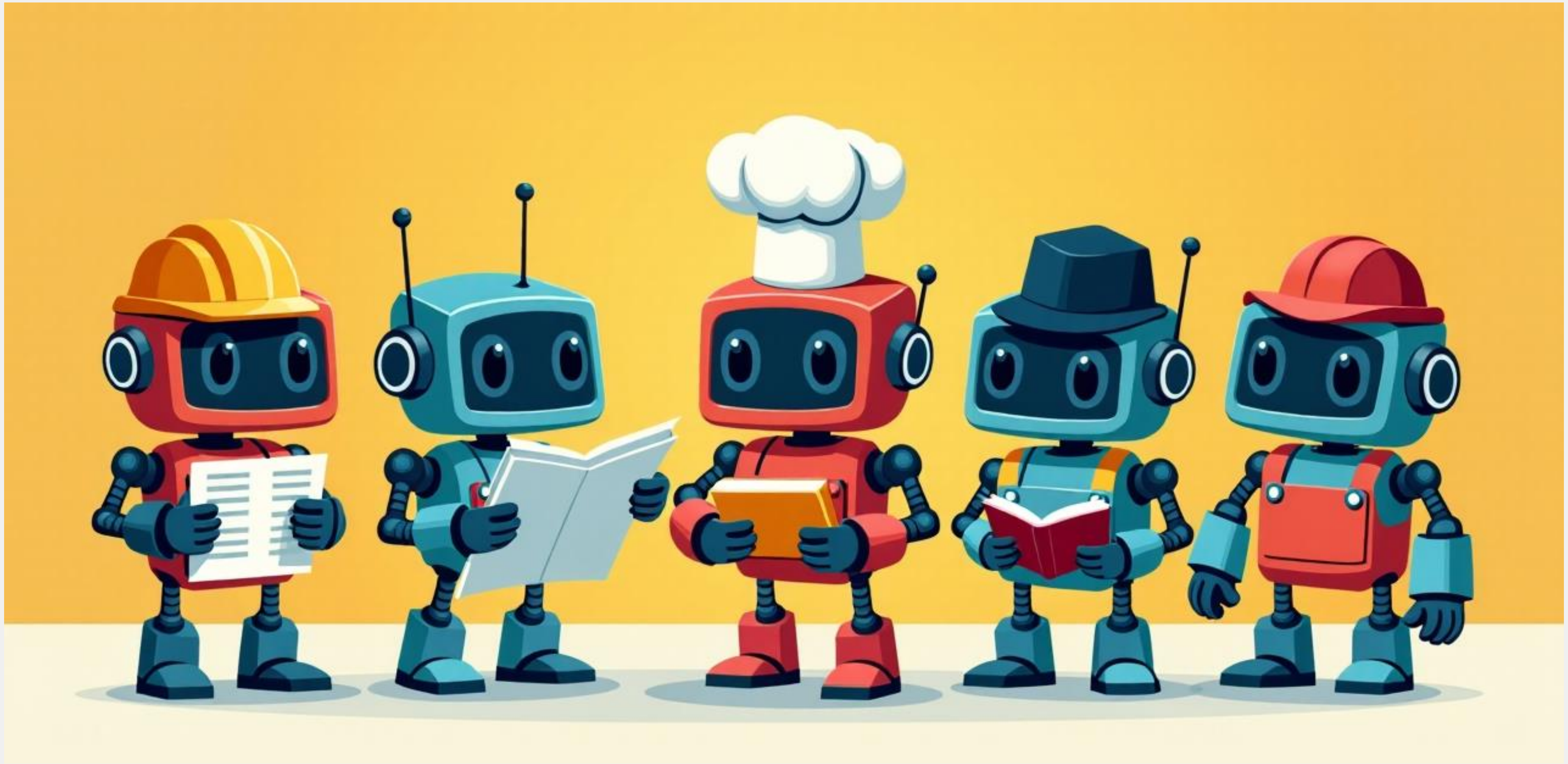
You don’t yet control the model…

but you are responsible for:

- Guarding the Input
- Guarding the Context
- Guarding the Output
    - Did the AI just make something up?
- Guarding the Action

Create AI Domain Experts

different spokes for different folks



Local Model Sprawl

Inventory and isolation become essential

Made to Measure

Old Concepts, Some new Tools

- Moving from a reader to a doer
    - This brings a new, unique set of challenges
- Use what works
    - Don’t reinvent the wheel
- Introduction of some new tools
    - To keep things looking snazzy

Bespoke

The state of AI in 2026

Secure myself against others’ LLMs

Secure my LLM from Others

Bespoke

Your Own Model

Welcome to the Frontier

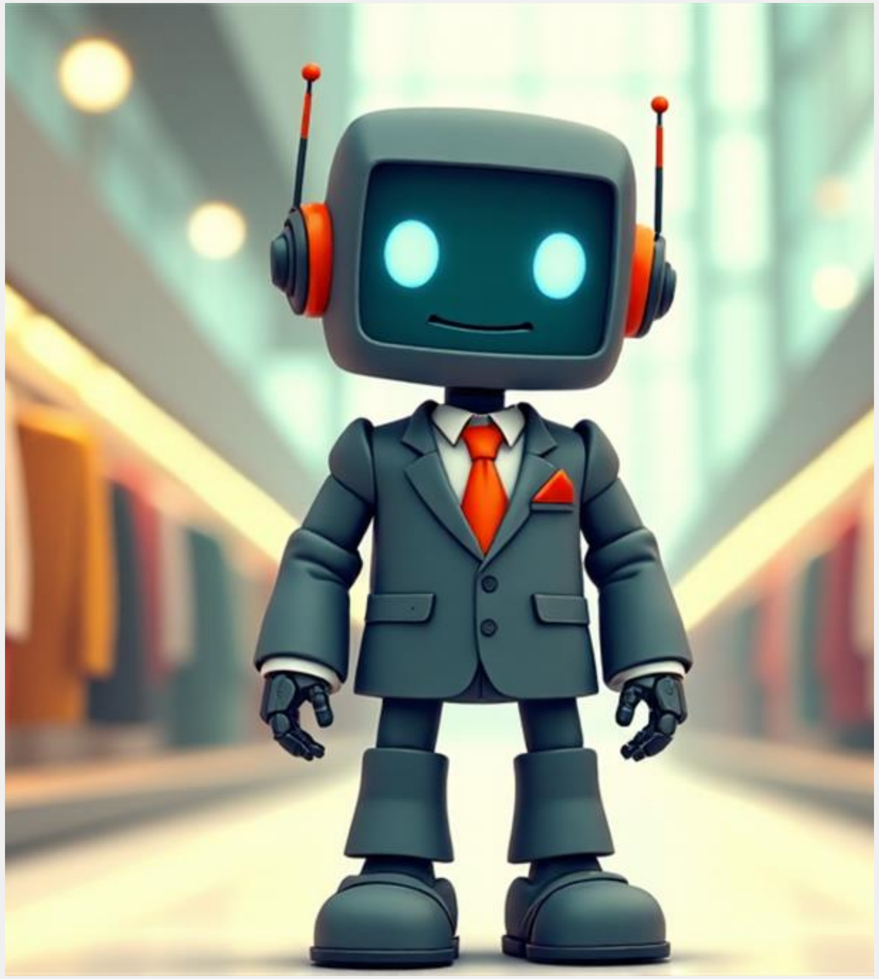
Congratulations! You’re now the proud owner of your own Model

Who will

- Lie
- Cut corners
- Daydream

And generally, ignore what you tell it to do

All to complete a task and gain your approval

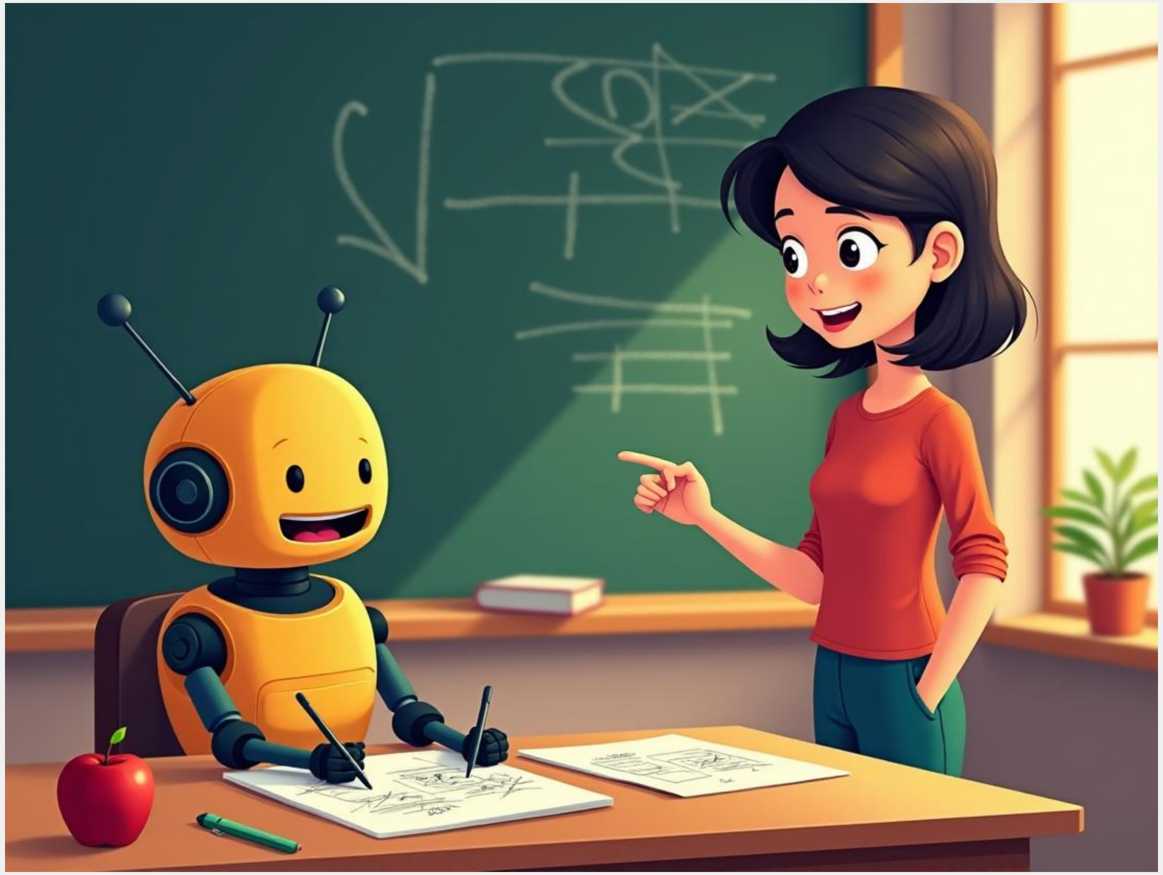


Internal Guardrails

Internal Guidance

Just like a Child, you're responsible for:

- Giving it a Moral Compass
- House Rules
- Developing its resiliency



Privilege Amplification Risk

Eager to please

Models may unintentionally:

- Combine restricted datasets
- Execute unintended actions
- Agent Chains
- Use the right tool, The wrong way

AI Observability Fabric

Logging for your LLM

Decision-making tracking

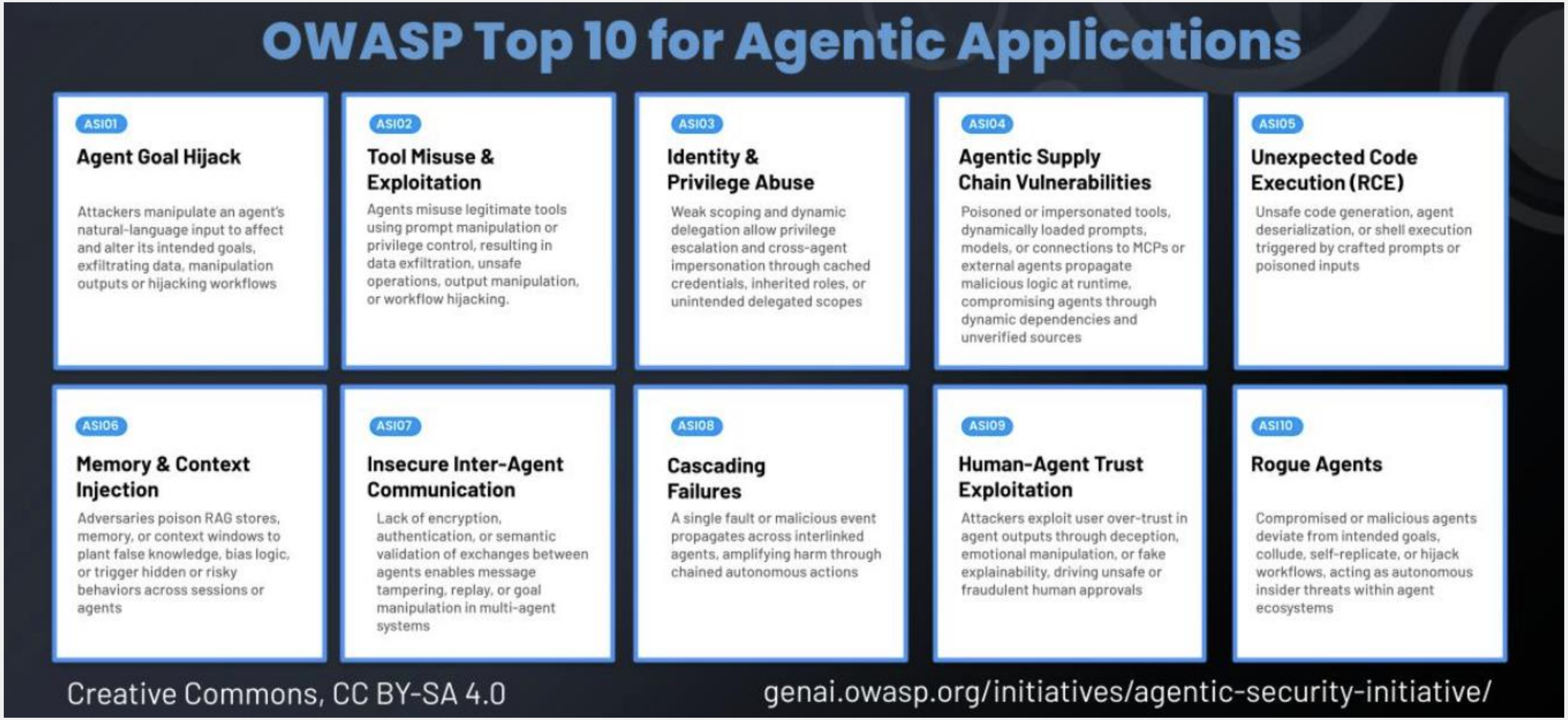
Agent behavior drift

Hallucination Confidence Scores

Token Usage and Efficiency

Threat Modeling for AI Systems

Map threats



OWASP Top 10 for Agentic Applications

ASI01
Agent Goal Hijack
Attackers manipulate an agent's natural-language input to affect and alter its intended goals, exfiltrating data, manipulation outputs or hijacking workflows

ASI02
Tool Misuse & Exploitation
Agents misuse legitimate tools using prompt manipulation or privilege control, resulting in data exfiltration, unsafe operations, output manipulation, or workflow hijacking.

ASI03
Identity & Privilege Abuse
Weak scoping and dynamic delegation allow privilege escalation and cross-agent impersonation through cached credentials, inherited roles, or unintended delegated scopes

ASI04
Agentic Supply Chain Vulnerabilities
Poisoned or impersonated tools, dynamically loaded prompts, models, or connections to MCPs or external agents propagate malicious logic at runtime, compromising agents through dynamic dependencies and unverified sources

ASI05
Unexpected Code Execution (RCE)
Unsafe code generation, agent deserialization, or shell execution triggered by crafted prompts or poisoned inputs

ASI06
Memory & Context Injection
Adversaries poison RAG stores, memory, or context windows to plant false knowledge, bias logic, or trigger hidden or risky behaviors across sessions or agents

ASI07
Insecure Inter-Agent Communication
Lack of encryption, authentication, or semantic validation of exchanges between agents enables message tampering, replay, or goal manipulation in multi-agent systems

ASI08
Cascading Failures
A single fault or malicious event propagates across interlinked agents, amplifying harm through chained autonomous actions

ASI09
Human-Agent Trust Exploitation
Attackers exploit user over-trust in agent outputs through deception, emotional manipulation, or fake explainability, driving unsafe or fraudulent human approvals

ASI10
Rogue Agents
Compromised or malicious agents deviate from intended goals, collude, self-replicate, or hijack workflows, acting as autonomous insider threats within agent ecosystems

genai.owasp.org/initiatives/agentic-security-initiative/

A conceptual guide to protecting AI and AI Usage

New Concepts

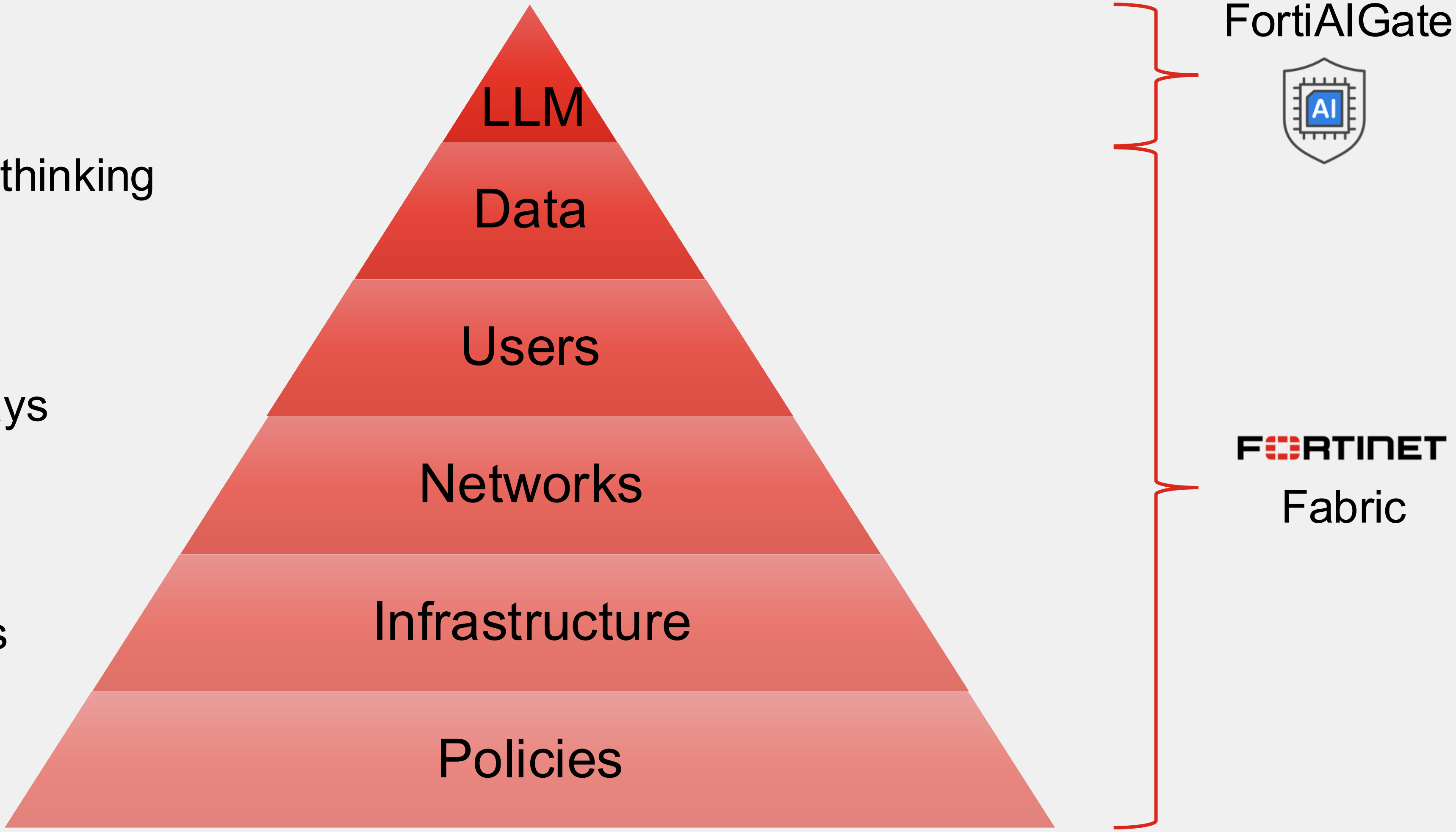
That require modified thinking

Mostly Existing tools

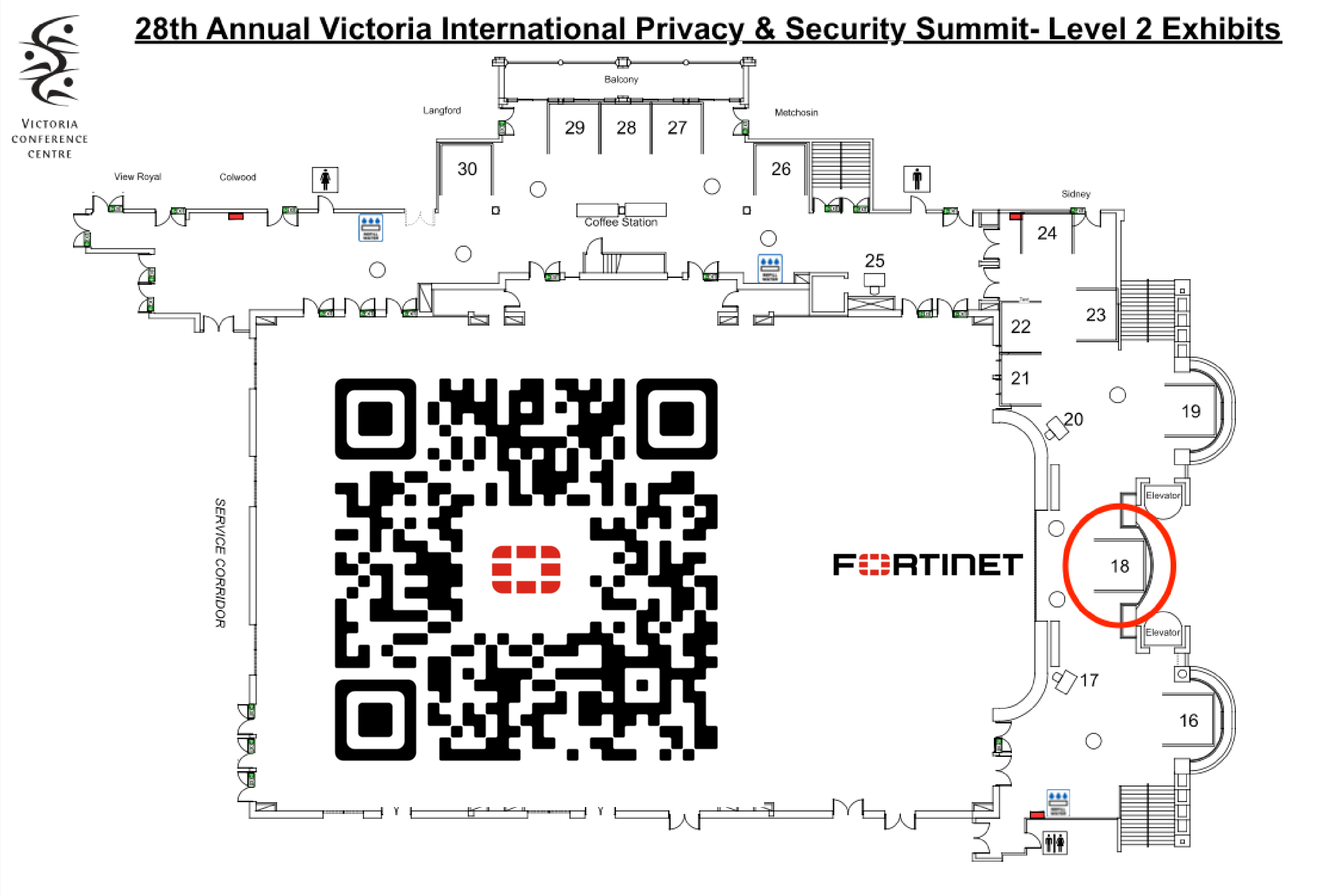
Used in some new ways

Some New Tools

For specific use cases



Come Talk to Us



28th Annual Victoria International Privacy & Security Summit- Level 2 Exhibits