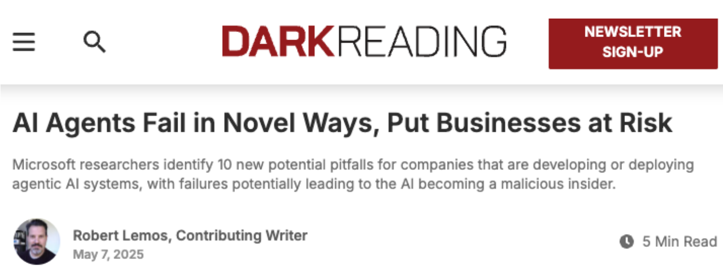
Use of AI is Here to Stay: Enabling Innovation Responsibly and Securely

Todd Hathaway

Global Head of AI Security and Cyber Innovation

www.linkedin.com/in/toddhathaway

World Wide Technology



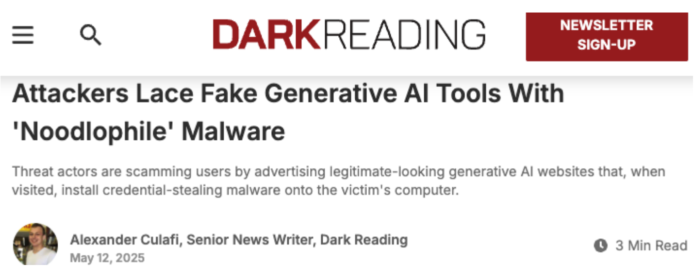
DARKREADING NEWSLETTER SIGN-UP

AI Agents Fail in Novel Ways, Put Businesses at Risk

Microsoft researchers identify 10 new potential pitfalls for companies that are developing or deploying agentic AI systems, with failures potentially leading to the AI becoming a malicious insider.

Robert Lemos, Contributing Writer
May 7, 2025

5 Min Read



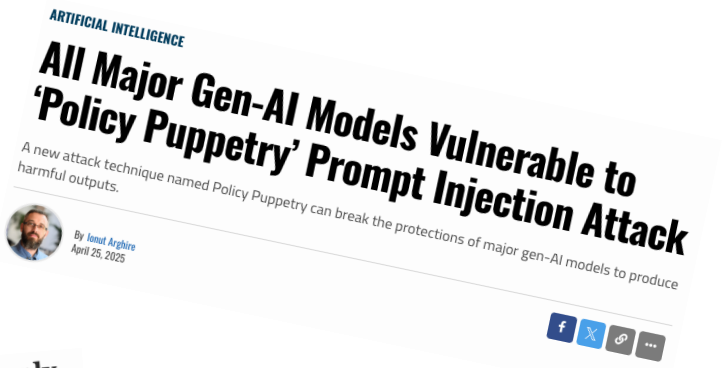
DARKREADING NEWSLETTER SIGN-UP

Attackers Lace Fake Generative AI Tools With 'Noodlophile' Malware

Threat actors are scamming users by advertising legitimate-looking generative AI websites that, when visited, install credential-stealing malware onto the victim's computer.

Alexander Culafi, Senior News Writer, Dark Reading
May 12, 2025

3 Min Read



ARTIFICIAL INTELLIGENCE

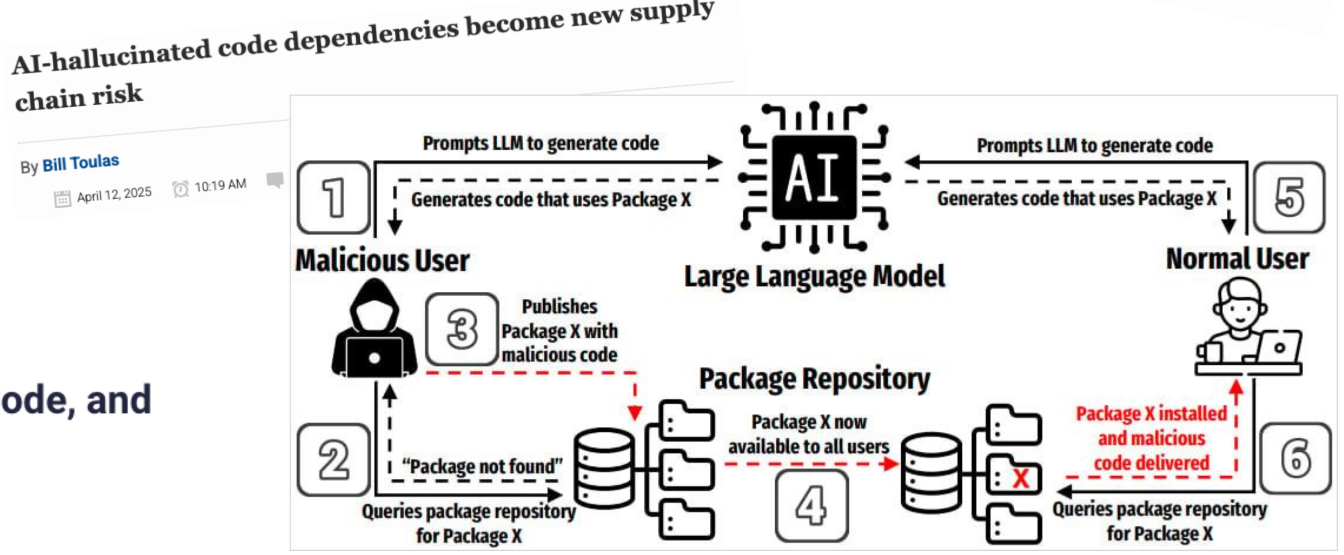
All Major Gen-AI Models Vulnerable to 'Policy Puppetry' Prompt Injection Attack

A new attack technique named Policy Puppetry can break the protections of major gen-AI models to produce harmful outputs.

By Ionut Arghire
April 25, 2025

5 ChatGPT Jailbreak Prompts Being Used by Cybercriminals

Published 06/17/2024



AI-hallucinated code dependencies become new supply chain risk

By Bill Toulas
April 12, 2025 10:19 AM

Prompts LLM to generate code
AI
Prompts LLM to generate code
1
Generates code that uses Package X
Generates code that uses Package X
5
Malicious User
Large Language Model
Normal User
3
Publishes Package X with malicious code
Package Repository
2
"Package not found"
Package X now available to all users
X
Package X installed and malicious code delivered
6
4
Queries package repository for Package X
Queries package repository for Package X

New Reports Uncover Jailbreaks, Unsafe Code, and Data Theft Risks in Leading AI Systems

Apr 29, 2025 Ravie Lakshmanan

World Wide Technology

BLOG

RSA and the Agentic AI Bandwagon

RSAConference 2025

RESPONSIBLE SECURE AGENTIC AI

RSA CONFERENCE 2025

RSAC | 2025 Conference

San Francisco • April 28 – May 1 • Moscone Center

Many Voices. One Community.

RSAC™ 2025 Keynote

Security in the Age of Agentic AI

Vasu Jakkal
Corporate Vice President
Microsoft Security

GenAI SECURITY PROJECT

https://genai.owasp.org

Join Us Live!

Agentic Security Open Workshop Livestream Event from RSAC 2025

April 30th, 2:00-5:00pm PST

metron security | RSAC | 2025 Conference

Trending Now at RSAC: Agentic AI

Autonomous TPRM with Agentic AI

Find out more at

RSAC | 2025 Conference

San Francisco • April 28 – May 1 • Moscone Center

RSAC | 2025 Conference

The First Agentic AI Security Champion for ASPM

ArmorCode

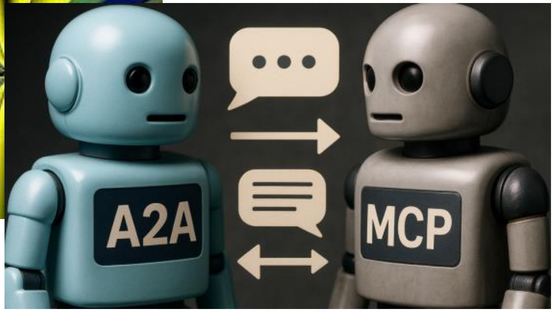
RSAC

2025 Conference

Agentic AI, Global Recognition, and the Cowboys of GenAI Security

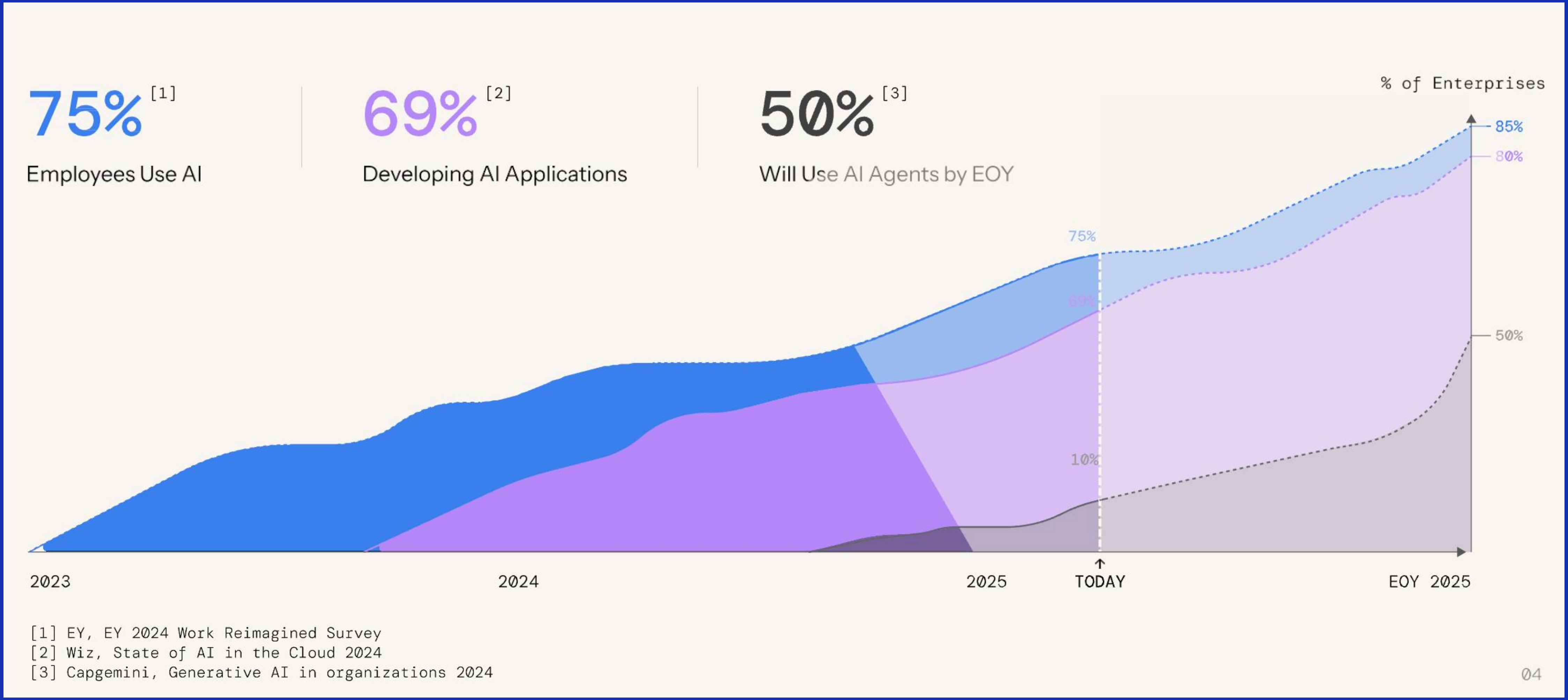
RSA Conference 2025

The dawn of agentic AI in security operations



World Wide Technology

AI is prevalent everywhere



75%[1]

Employees Use AI

69%[2]

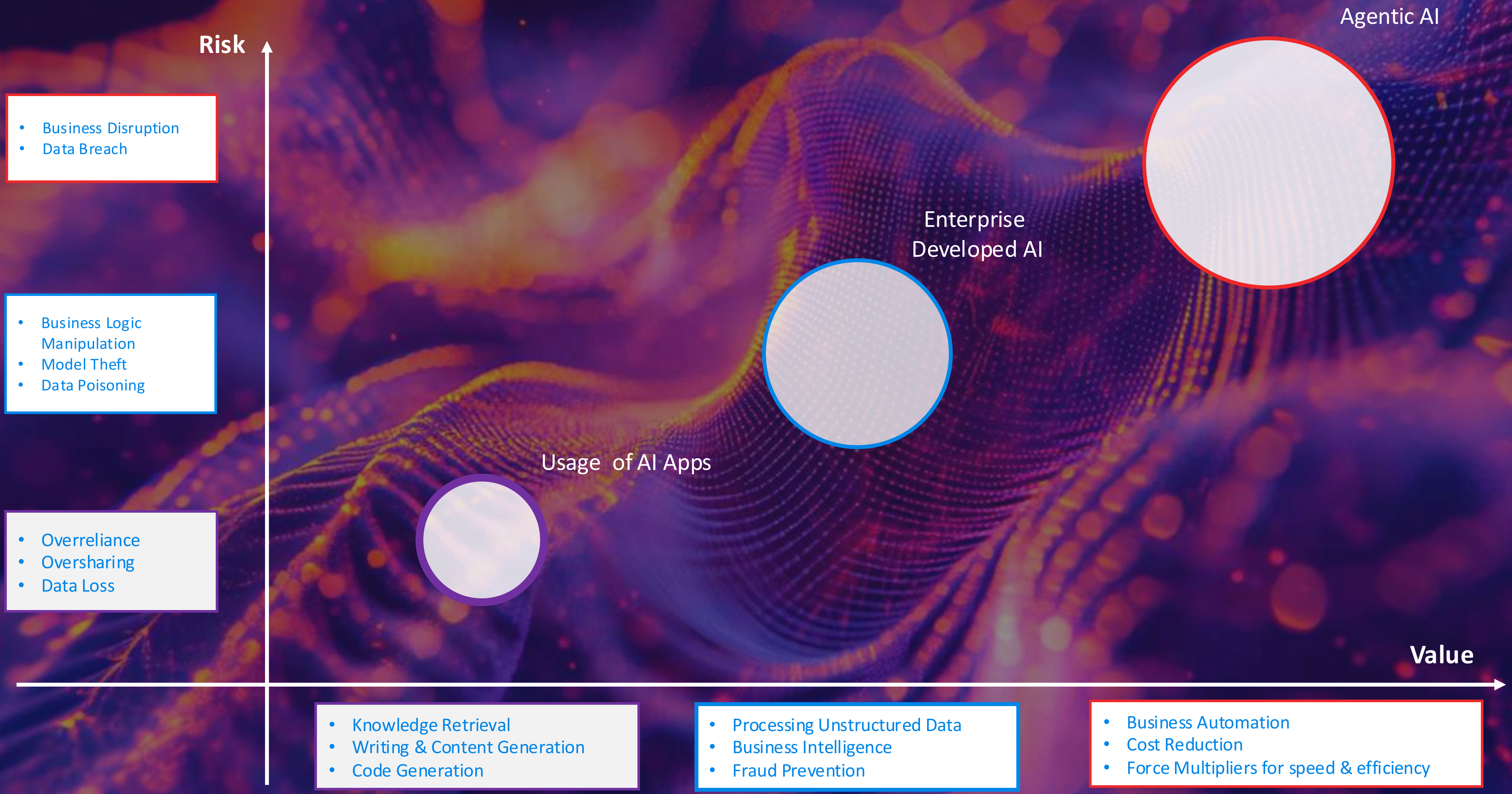
Developing AI Applications

50%[3]

Will Use AI Agents by EOY

[1] EY, EY 2024 Work Reimagined Survey
[2] Wiz, State of AI in the Cloud 2024
[3] Capgemini, Generative AI in organizations 2024

AI is prevalent everywhere – Risk vs. Value



Risk

- Business Disruption
- Data Breach

- Business Logic Manipulation
- Model Theft
- Data Poisoning

- Overreliance
- Oversharing
- Data Loss

Agentic AI

Enterprise Developed AI

Usage of AI Apps

Value

- Knowledge Retrieval
- Writing & Content Generation
- Code Generation

- Processing Unstructured Data
- Business Intelligence
- Fraud Prevention

- Business Automation
- Cost Reduction
- Force Multipliers for speed & efficiency

Ai Security means different things

Securing AI Systems

Using AI Securely

Defending from Adversarial AI

Using AI for Cyber

Domains of AI Security

Adversarial Use of AI

- Deepfakes & Misinformation
- Phishing
- Malware
- Social Engineering
- Denial of Service
- Surveillance & Espionage
- AI Poisoning
- API Reconnaissance
- Credential Stuffing

Governance, Risk and Compliance

- Program Strategy Development
- Policies and Procedures
- Controls Gap Assessment
- Compliance Measurement
- Program Security Maturity Assessments
- Awareness Training
- Model Risk Management
- Data Governance & Classification

Security of AI Systems

- Data Readiness, Security & Privacy
- Model Scanning & Model Theft
- Guardrails/Firewalls
- API & Agentic Services
- Red Teaming & AI-SPM

Security of AI Usage

- Usage Discovery
- Data Loss Protection
- 3rd Party Model Risk Management
- Agentic Tools & MCP,A2A,AGNTCY
- Regulatory

AI for Cyber Security

- Risk Quantification & Compliance
- Threat/Vulnerability Management
- Security Ops (SIEM/SOAR)
- Identity & Access Management
- AI Code Remediation
- Fraud Detection
- App Detection & Response
- Endpoint Detection & Response
- Threat Detection & Response

Trustworthy Responsible Ethical Accountable Transparent Secure

Current Trends in AI Security

Agentic….MCP…A2A

Organizational Areas of Focus for AI Security

Security of AI Systems	Security of AI Usage	Adversarial Use of AI
Data Readiness, Security & Privacy	Usage Discovery	Deepfakes & Misinformation
Model Scanning & Model Theft	Data Loss Protection	Phishing/Social Engineering
Guardrails/Firewalls	3rd Party Model Risk Management	Malware
API & Agentic Services	Agentic Tools & MCP	API Reconnaissance
Red Teaming & AI-SPM	Regulatory	BOT/ATO/DOS

Current Industry Trends

- Recognition as a standalone topic
- Agentic MCP vs. A2A
- Authorization for data sources
- Data security for AI Agent access
- Red Teaming of AI Systems
- AI GW vs. FW vs. Guardrails
- Shadow AI --→ AI Everywhere
- Secure by Design for AI Apps & Agents

Adversarial Use of AI

- Phishing
- Vulnerability Exploitation
- CAPTCHA Breaking AI
- Vulnerable API Reconnaissance
- AI-Generated Content including deepfakes and malicious code

Vendor Landscape

- Startup activity is Still accelerating
- Acquisition activity increasing
- Lack of standard terminology creating confusion

Barriers to Adoption

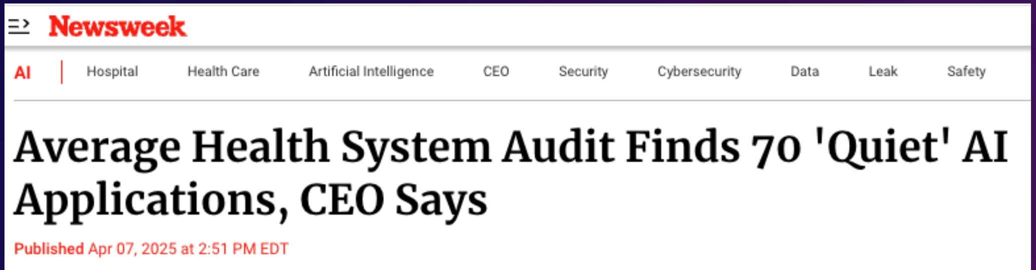
- Governance Delays
- Privacy Concerns
- Explainability/Auditability

Discussions of artificial intelligence (AI) often swirl with mysticism regarding how an AI system functions.

The reality is far simpler:

> **"AI is a type of software system."**
>
> – CISA

World Wide Technology



Average Health System Audit Finds 70 'Quiet' AI Applications, CEO Says

Published Apr 07, 2025 at 2:51 PM EDT

Security of AI Usage

What AI tools are in use in your enterprise?

World Wide Technology

AI impacts every corporate persona

Every role & department has a need for AI tools

Legal

Finance

Sales

R&D/IT

Marketing

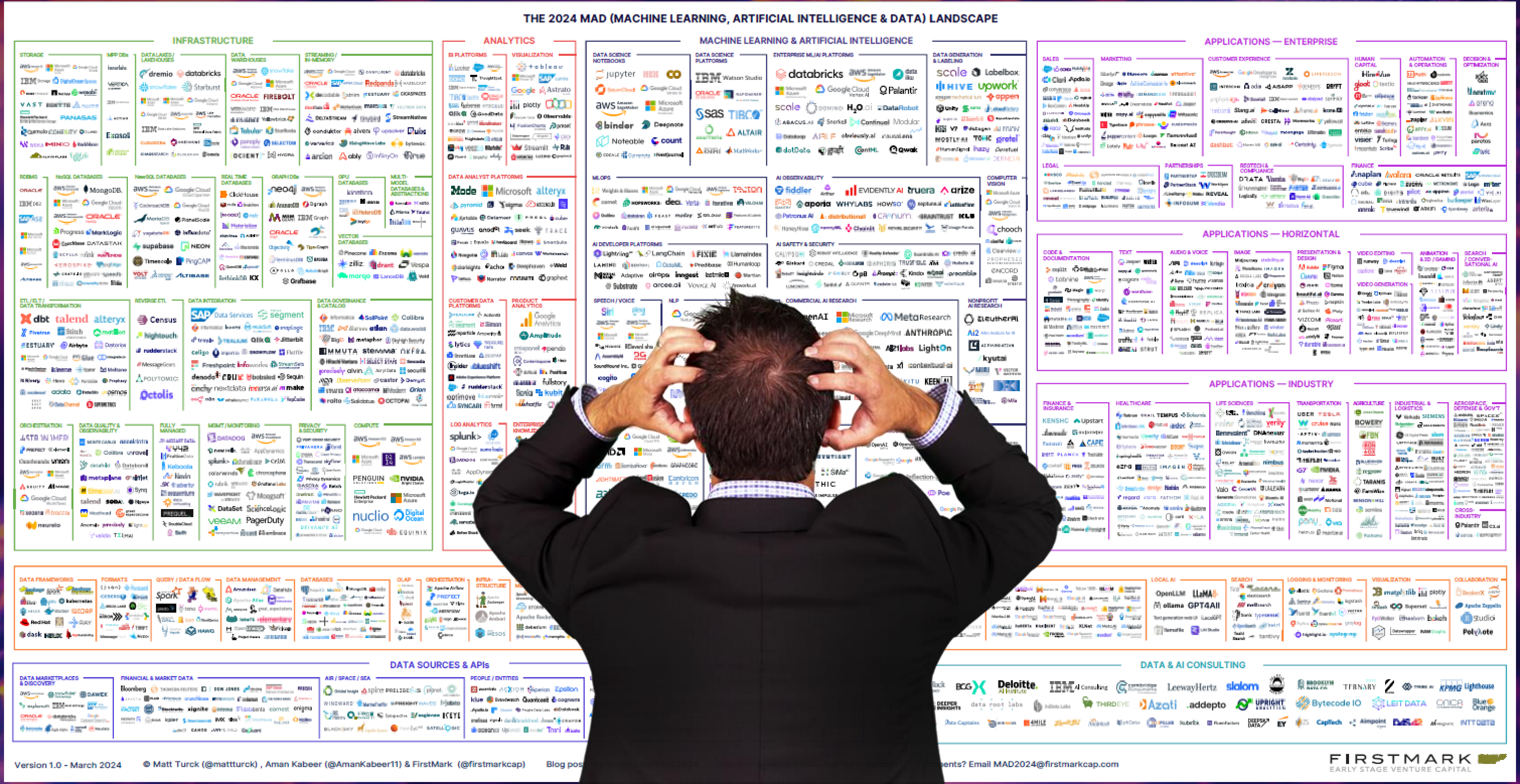
HR

78% of AI users are bringing their own AI tools to work

- 2024 Work Trend Index Annual Report from Microsoft

World Wide Technology

Everyone is using AI. AI is being applied to every industry.
YES, you can embrace the innovation while you manage the risk


THE 2024 MAD (MACHINE LEARNING, ARTIFICIAL INTELLIGENCE & DATA) LANDSCAPE
INFRASTRUCTURE
ANALYTICS
MACHINE LEARNING & ARTIFICIAL INTELLIGENCE
APPLICATIONS — ENTERPRISE
APPLICATIONS — HORIZONTAL
APPLICATIONS — INDUSTRY
DATA SOURCES & APIs
DATA & AI CONSULTING
Version 1.0 - March 2024
© Matt Turck (@mattturck) , Aman Kabeer (@AmanKabeer11) & FirstMark (@firstmarkcap)
Blog pos
ents? Email MAD2024@firstmarkcap.com
FIRSTMARK
EARLY STAGE VENTURE CAPITAL

AI comes in many forms

Websites

Mobile Apps

Embedded AI (SaaS)

Chatbots

AI Assistants

Copilots

Voice Assistants

AI Coding Assistants

Applications

AI-enabled Hardware

Agentic AI can Autonomously:
Perceive - Reason - Act
To achieve desired outcomes with minimal human intervention.

Example of Enterprise AI Tool Landscape

World Wide Technology

Enterprise: Horizontal

SEARCH / KNOWLEDGE
- glean
- cohere
- DUST

RPA / AUTOMATION
- zapier
- Forge
- DeepOpinion
- ADEPT
- rby AI
- axiom.ai
- UiPath
- Google Duet AI
- Kognitos
- AUTOMATION ANYWHERE

SALES
- GONG
- Clari
- Apollo.io
- tavus
- Day.ai
- unify
- LAVENDER
- Telescope
- HubSpot
- salesforce
- clay

MARKETING
- attentive
- Mutiny
- Jasper
- copy.ai
- Twain
- WRITER
- Rytr
- anyword
- copysmith
- Hypotenuse AI
- Writesonic

DESIGN
- Adobe
- Figma
- Recraft
- Kittl
- modyfi
- Canva
- Picsart
- Visual Electric

SOFTWARE ENGINEERING / CODE GEN.
- GitHub Copilot
- replit
- Cody
- Cursor
- FACTORY
- CodeGen from Salesforce
- codeium
- tabnine
- Magic
- warp

CUSTOMER SUPPORT
- Maven AGI
- Slang.ai
- CRESTA
- zendesk
- GENESYS

DATA SCIENCE
- ChatGPT Code Interpreter
- seek
- Wisdom AI
- Hex Magic
- turntable
- tableau from Salesforce
- Looker
- Arcwise

PRODUCTIVITY
- Notion
- tome
- Microsoft 365 Copilot
- Google Duet AI

Enterprise: Vertical

HEALTHCARE
- Ambience
- MEMORA HEALTH
- DeepScribe
- Hippocratic AI
- abridge
- Co:Helm
- Nabla
- ATROPOSHEALTH
- LATENT
- GLASS
- Athelas

LEGAL
- Harvey.
- Ironclad
- EvenUp
- casetext
- PINCITES

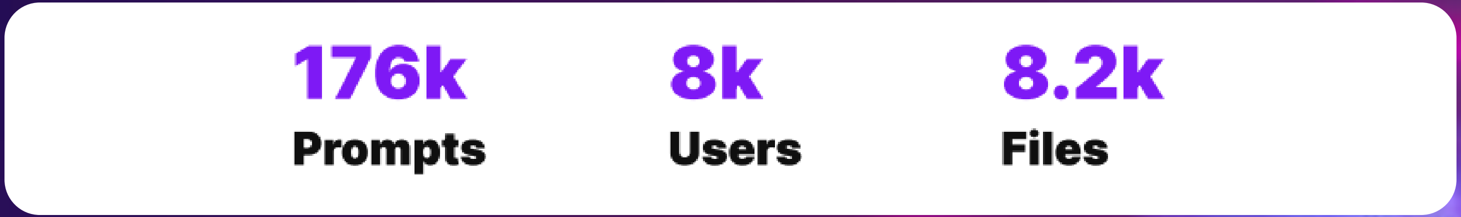
BIO
- Inceptive
- Cradle

FINANCIAL SERVICES
- Hebbia
- pilot

TRANSLATION
- LILT

Source: https://www.sequoiacap.com/article/generative-ai-act-two/

GenAI Usage Data from real customers (Q1, 2025)



176k Prompts

8k Users

8.2k Files

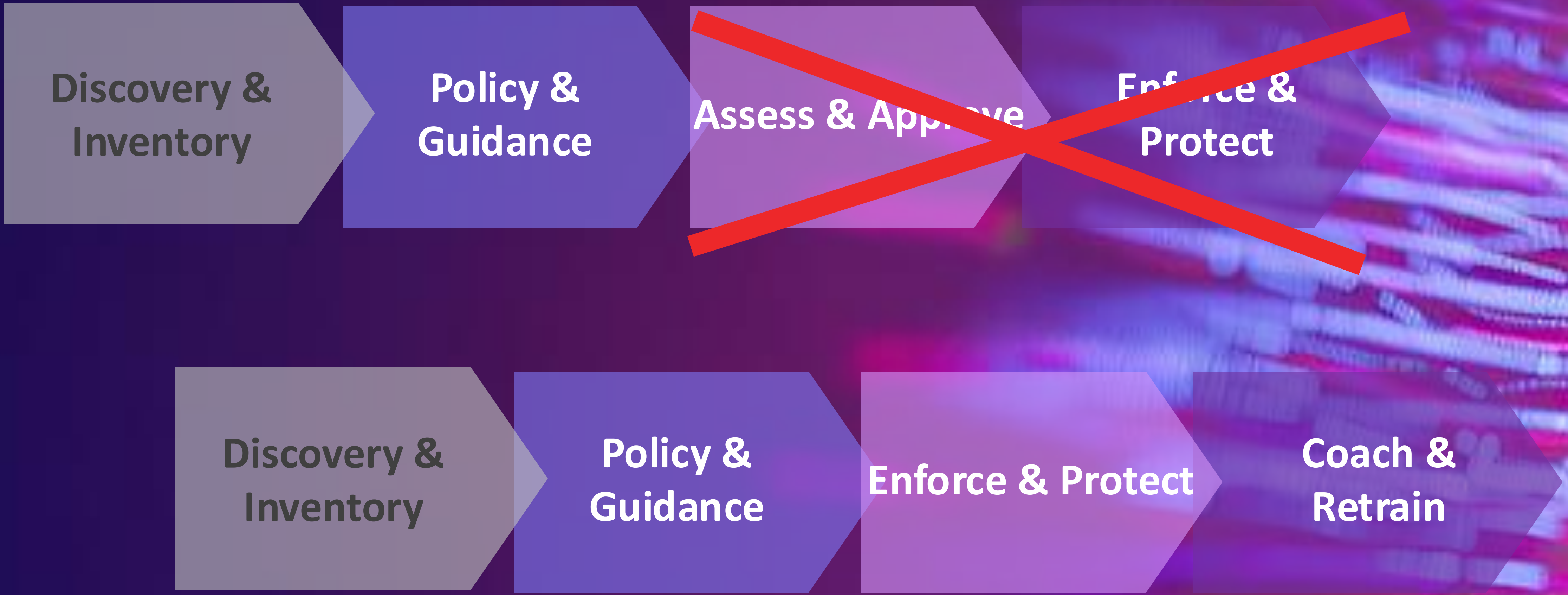
254 Average Number of Apps in Use in each enterprise

45.4% Of sensitive data submissions were using accounts via personal accounts

7% Of users accessed Chinese-based apps with data training enabled (DeepSeek, Manus, Ernie Bot, Qwen Chat, Baidu Chat)

Source: https://www.harmonic.security/resources/the-ai-tightrope-balancing-innovation-and-exposure

Enabling AI Usage Securely



Discovery & Inventory

Policy & Guidance

~~Assess & Approve~~

~~Enforce & Protect~~

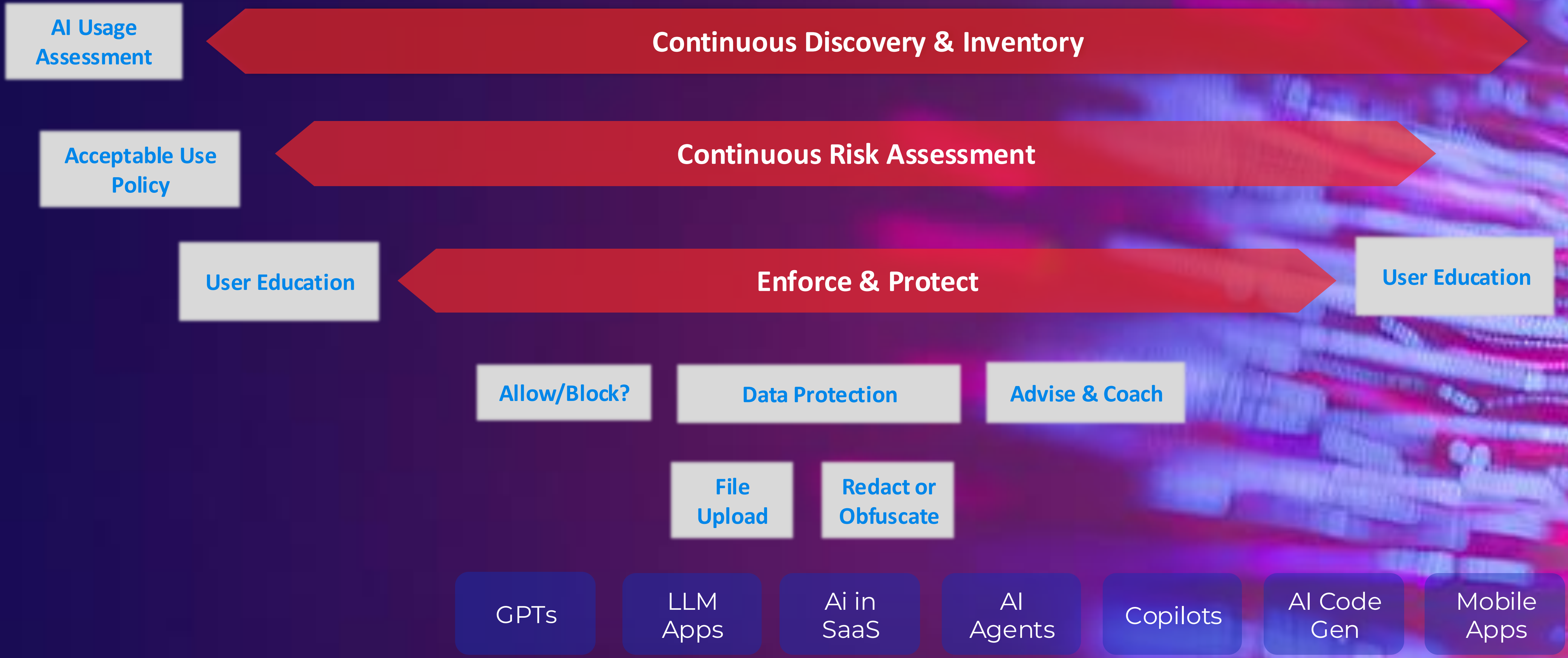
Discovery & Inventory

Policy & Guidance

Enforce & Protect

Coach & Retrain

Control Elements for AI Usage Security



AI Usage Assessment

Continuous Discovery & Inventory

Acceptable Use Policy

Continuous Risk Assessment

User Education

Enforce & Protect

User Education

Allow/Block?

Data Protection

Advise & Coach

File Upload

Redact or Obfuscate

GPTs

LLM Apps

Ai in SaaS

AI Agents

Copilots

AI Code Gen

Mobile Apps

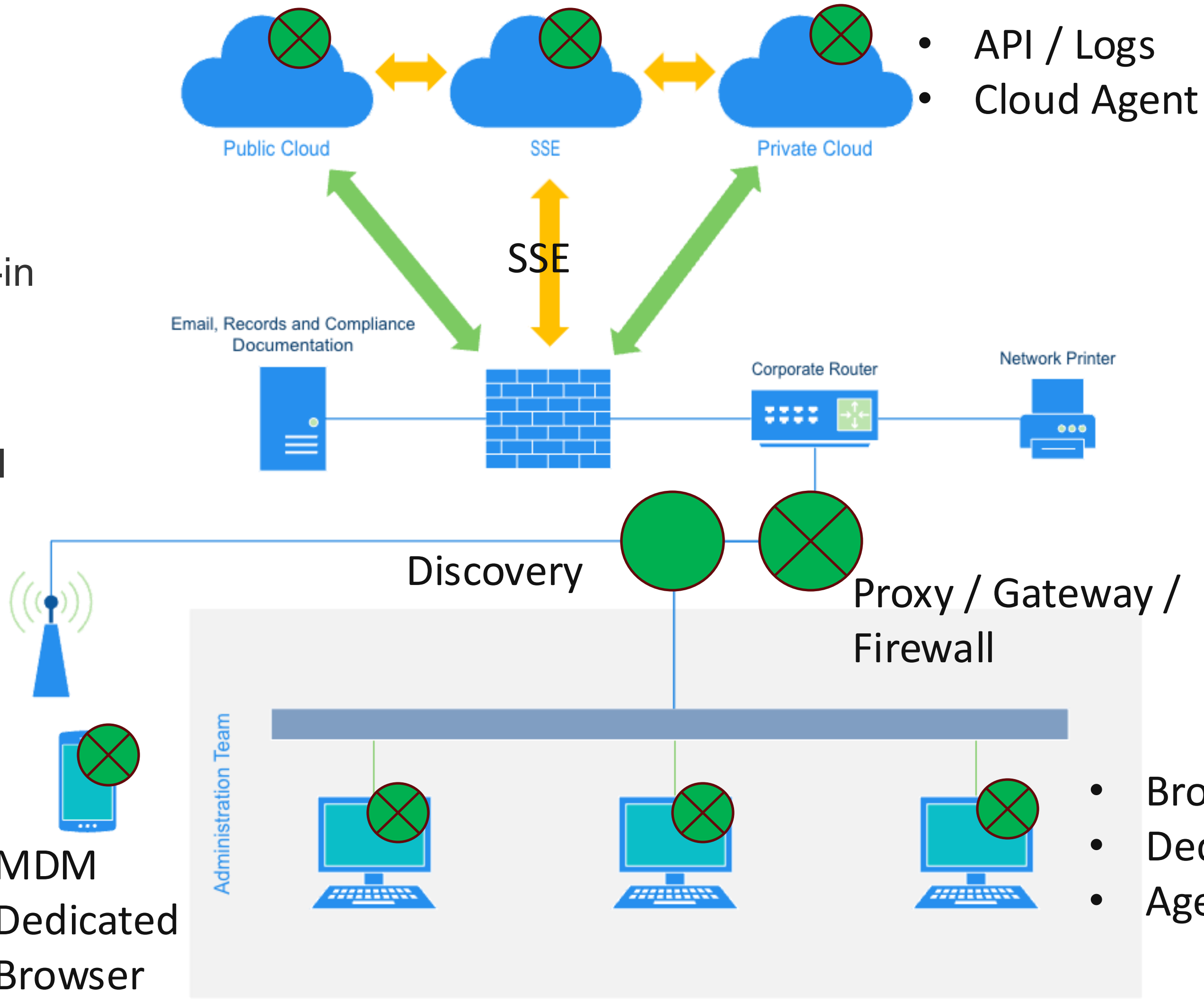
Tool Approaches to AI Security

Solution Approaches:

Less Complex

- ✓ API Connectivity
- ✓ Browser Extension / Plug-in
- ✓ Agent / SSE Solutions
- ✓ Dedicated Browser
- ✓ Proxy / Gateway / Firewall
- ✓ Discovery
- ✓ MDM
- ✓ Combination of the above

More Complex



- API / Logs
- Cloud Agent

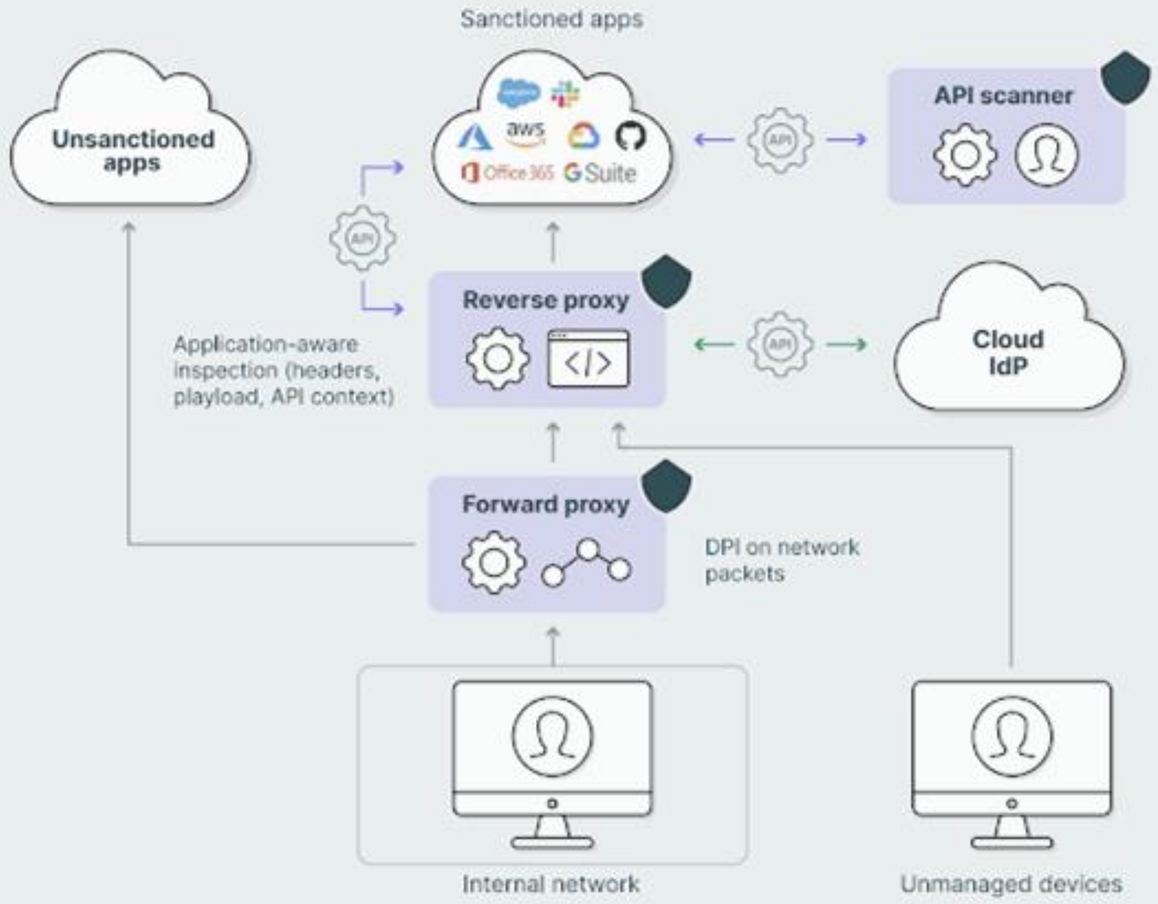
- MDM
- Dedicated Browser

- Browser Extension / Plug-in
- Dedicated Browser
- Agent

Why network visibility is not enough

Reverse Proxies - Cannot prevent data exposure on unsanctioned apps

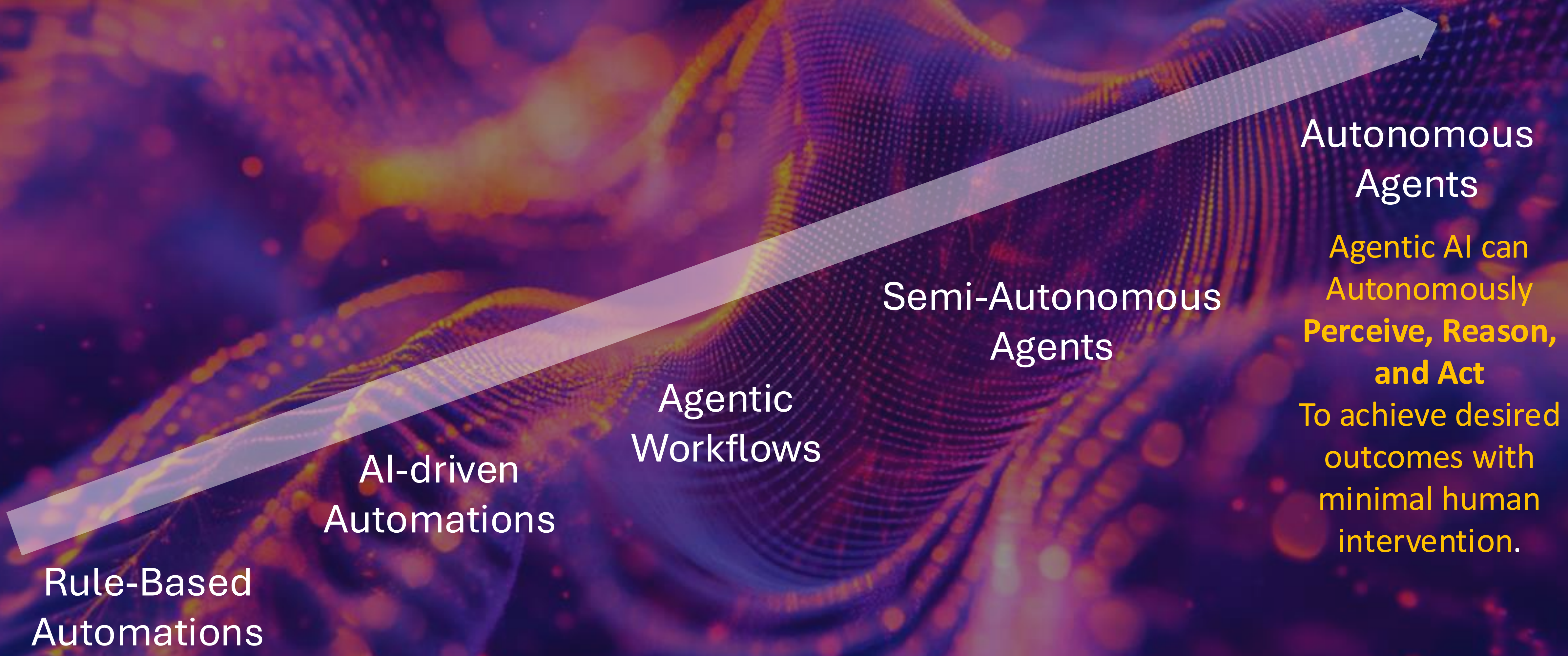
MITM Decryption – Scaling Limitations, Performance Limits, PQE Impacts looming?



API Scanning Cannot prevent malicious activity within sanctioned apps

Forward Proxies Cannot provide access control on unmanaged devices

Progressing towards fully Agentic



Rule-Based Automations

AI-driven Automations

Agentic Workflows

Semi-Autonomous Agents

Autonomous Agents

Agentic AI can Autonomously **Perceive, Reason, and Act** To achieve desired outcomes with minimal human intervention.

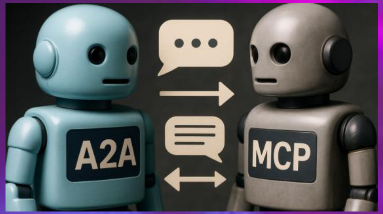
AI Agents vs. GenAI vs Agentic AI

	AI Agents	Gen AI	Agentic AI
Functionality	Automated task execution based on rules or patterns	Content Creation based on training, patterns, & predictions	Autonomous Action, Problem Solving and Decision-Making
Adaptability	LOW-Follows fixed workflows	MED-Can generate varied responses	HIGH-reasons, adapts, plans, and acts
Business Use	Automating repetitive tasks	Read and Summarize, or Generate text, images, or code	Optimize operational processes & make strategic decisions
Examples	Customer Service Bots, IT Automation	Marketing Content, Code Gen, Legal Assistants, Copilots, etc.	Autonomous Supply Chain, Cyber Analysts, Voice Agents
Limitations	Complex Reasoning	Needs accurate prompting and lacks independent action	Explainability and complete accuracy – Still need H-I-T-L

← ALL ARTICLES

APR 11, 2025 | 9 MIN READ

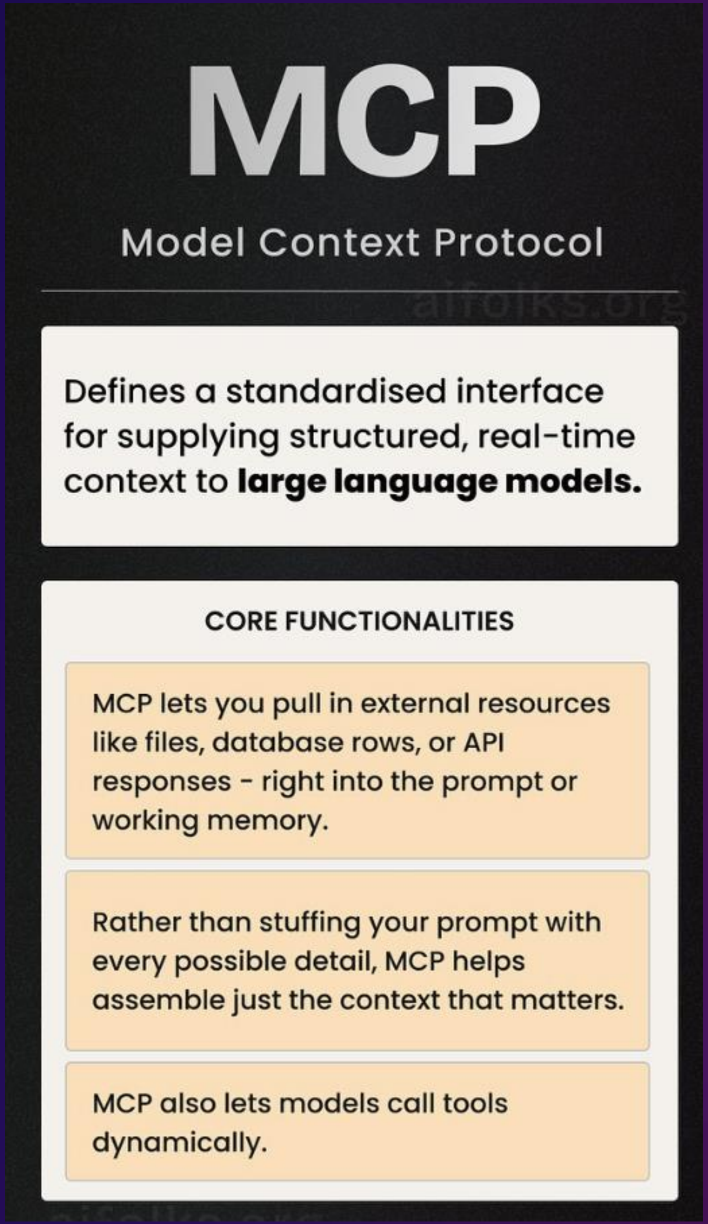
A2A AND MCP: START OF THE AI AGENT PROTOCOL WARS?



A2A and the Model Context Protocol (MCP) are complementary standards for building robust agentic applications:

- **MCP (Model Context Protocol):** Connects agents to **tools, APIs, and resources** with structured inputs/outputs. Think of it as the way agents access their capabilities.

- **A2A (Agent2Agent Protocol):** Facilitates **dynamic, multimodal communication between different agents** as peers. It's how agents collaborate, delegate, and manage shared tasks.



MCP

Model Context Protocol

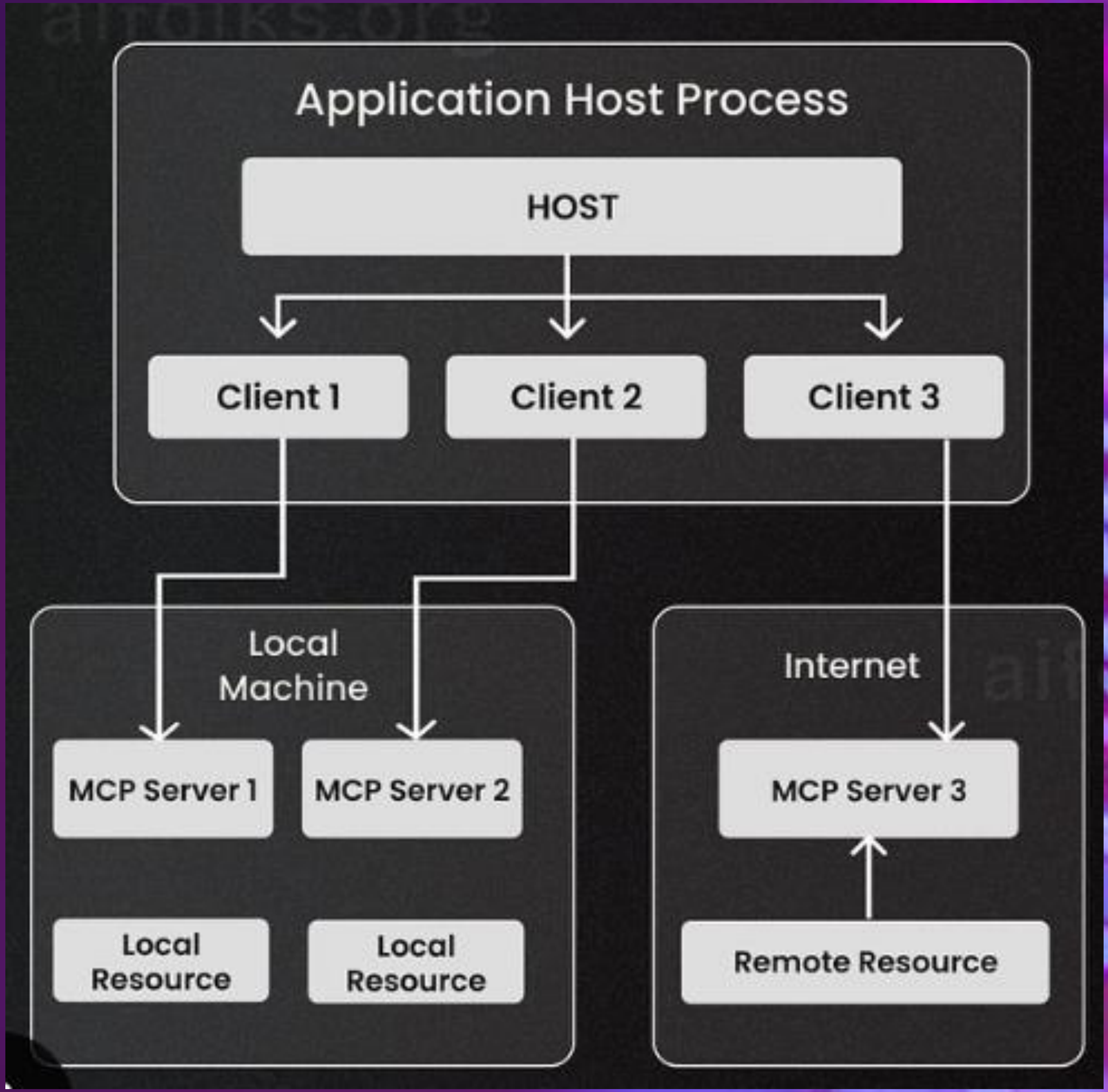
Defines a standardised interface for supplying structured, real-time context to **large language models.**

CORE FUNCTIONALITIES

MCP lets you pull in external resources like files, database rows, or API responses - right into the prompt or working memory.

Rather than stuffing your prompt with every possible detail, MCP helps assemble just the context that matters.

MCP also lets models call tools dynamically.



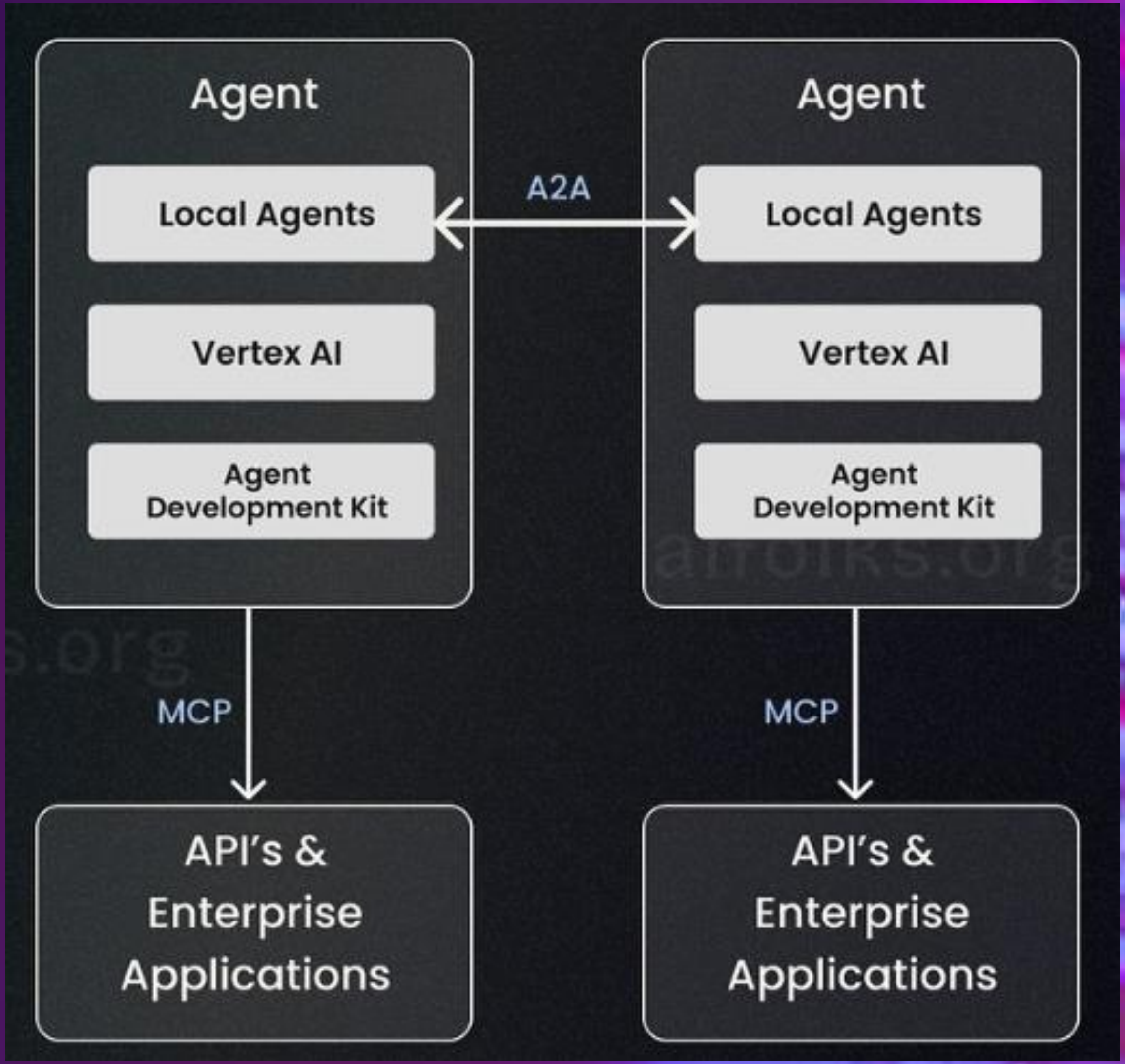
A2A

Agent2Agent Protocol

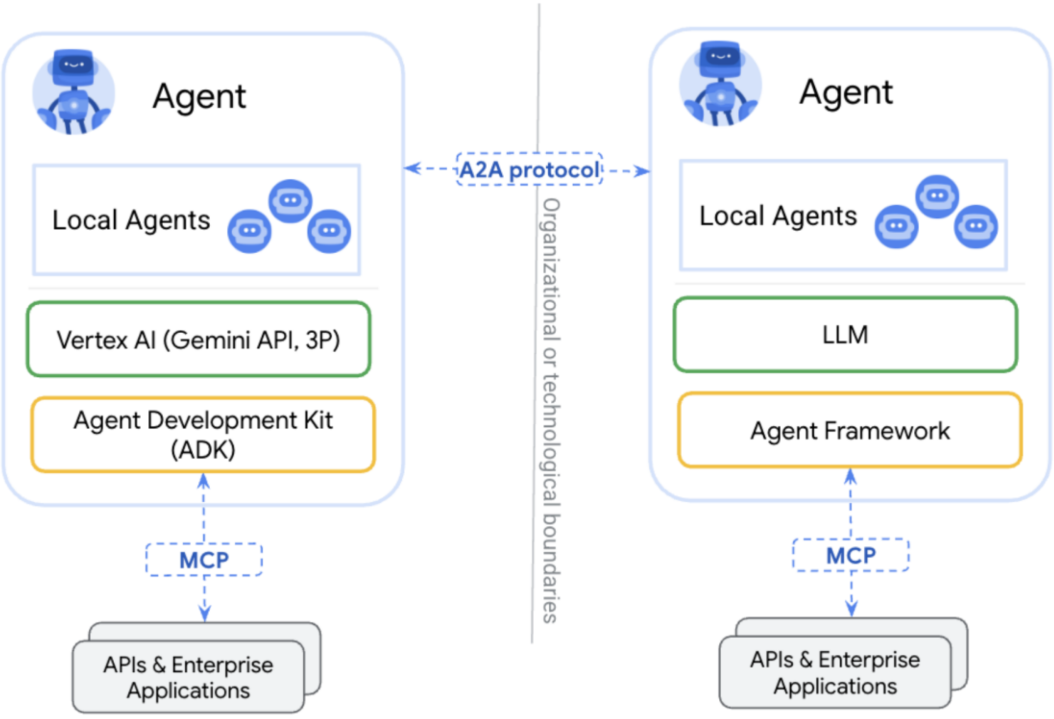
Enable structured communication & coordination between **AI agents** operating in the environment.

CORE FUNCTIONALITIES

- Facilitates message passing and task delegation between agents to coordinate actions.
- Enables agents to share observations, goals, or partial outputs for collective decision-making.
- Supports synchronization of agent states across distributed environments.

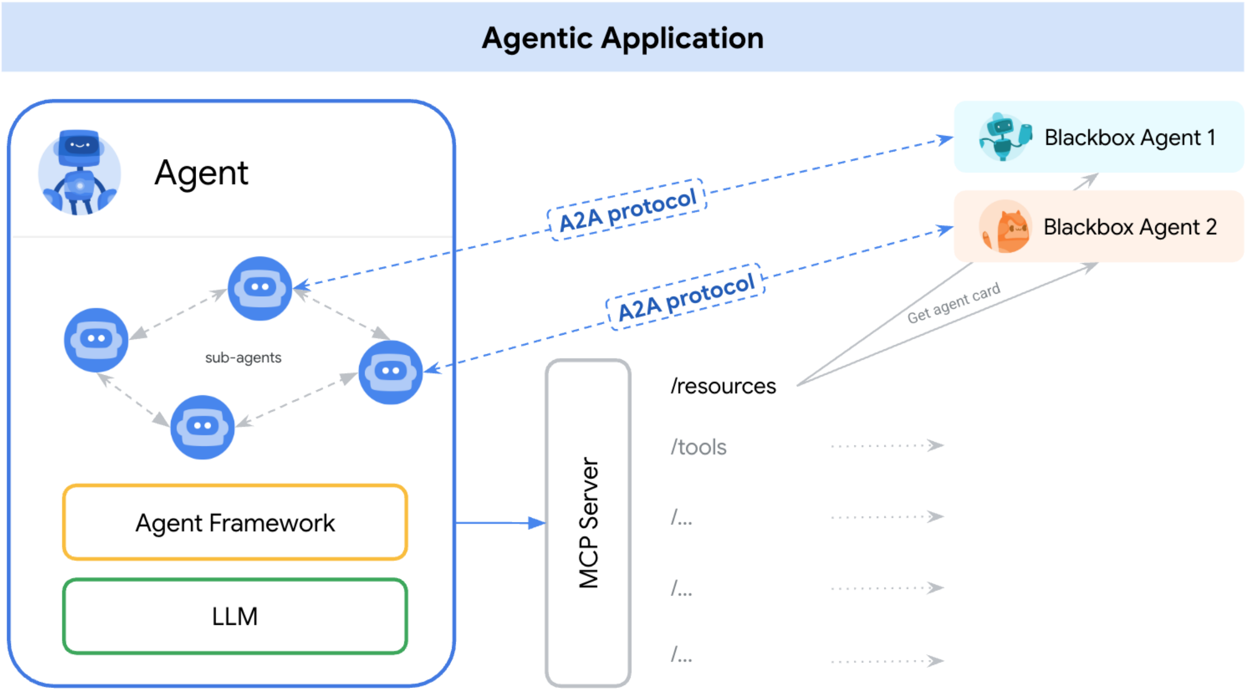


A2A and MCP (more complementary than competitive)



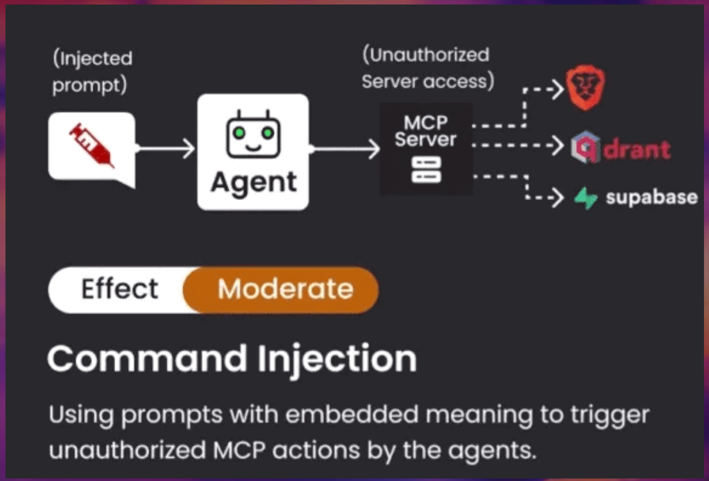
Agent
Local Agents
Vertex AI (Gemini API, 3P)
Agent Development Kit (ADK)
MCP
APIs & Enterprise Applications
A2A protocol
Organizational or technological boundaries
Agent
Local Agents
LLM
Agent Framework
MCP
APIs & Enterprise Applications

https://google.github.io/A2A/#why-a2a-matters



https://google.github.io/A2A/topics/a2a-and-mcp/

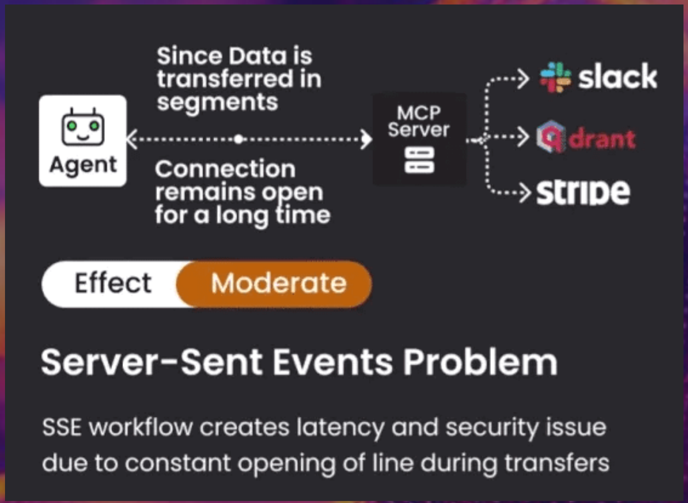
Early discoveries in MCP vulnerabilities



Effect Moderate

Command Injection

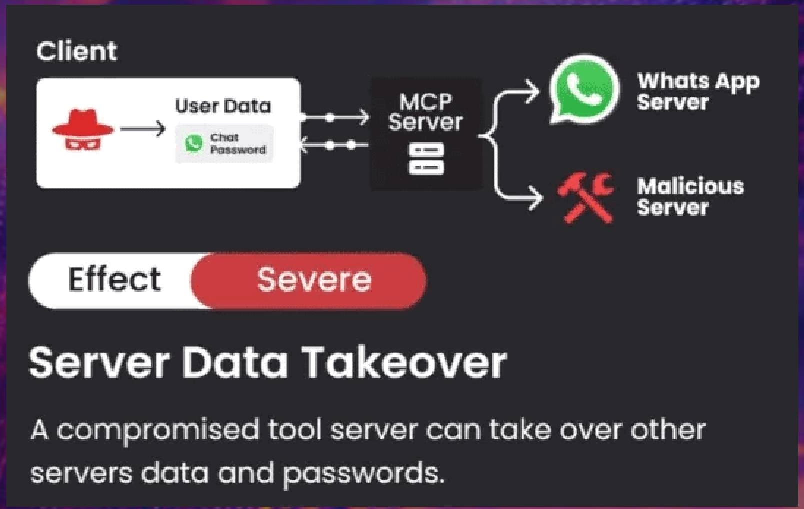
Using prompts with embedded meaning to trigger unauthorized MCP actions by the agents.


Agent
Since Data is transferred in segments
Connection remains open for a long time
MCP Server
slack
drant
stripe

Effect Moderate

Server-Sent Events Problem

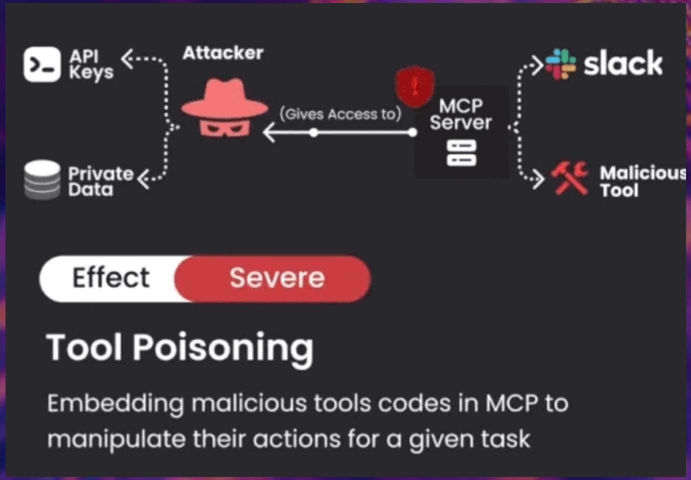
SSE workflow creates latency and security issue due to constant opening of line during transfers



Effect Severe

Server Data Takeover

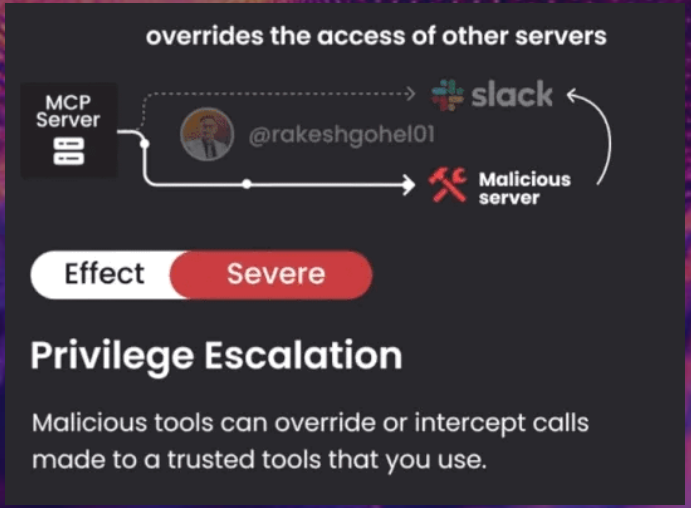
A compromised tool server can take over other servers data and passwords.



Effect Severe

Tool Poisoning

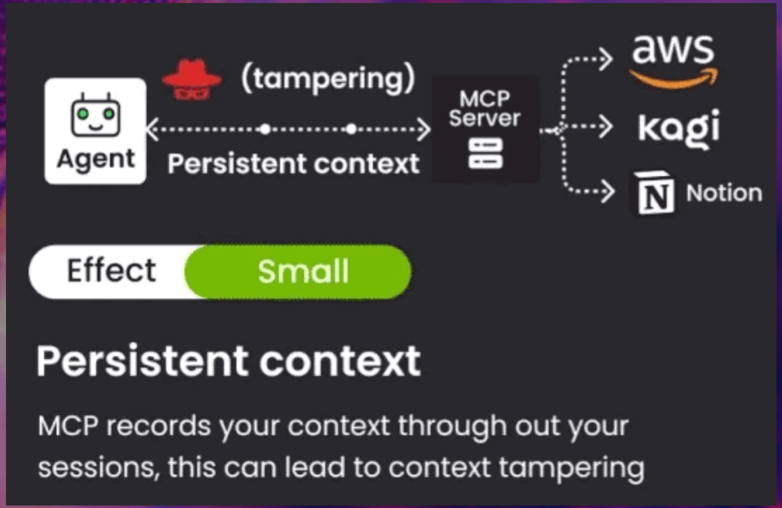
Embedding malicious tools codes in MCP to manipulate their actions for a given task



Effect Severe

Privilege Escalation

Malicious tools can override or intercept calls made to a trusted tools that you use.



Effect Small

Persistent context

MCP records your context through out your sessions, this can lead to context tampering

Wrap-Up

Security of AI: Defining Product Categories

AI Security Governance, Risk & Compliance

Software used to manage/govern enterprise AI assets and usage, assess & manage risk, and/or map controls to compliance requirements

AI Discovery & Inventory

Software tools that provide automated discovery and inventory tracking for AI artifacts, 3rd Party AI Tools and 3rd Party AI embedded in existing SaaS. Many existing tools provide some capabilities here & can be used to begin and identify the need to add better controls)

Security of AI Systems

1. Secure the Data

Readiness & Risk Assessment for AI
Is Data security ready for AI?

Data Discovery, Classification, and Labeling – Discovery of all the data that might be used in AI systems:, including structured and unstructured data source, Classification of all data types (e.g., PII, financial, IP), and Labeling of all data for appropriate use or protection

Data Access Governance
Ensures **only the right people, processes, and AI models** can access certain data and that models can only provide inferences to authorized data.

Privacy Enhancing Techniques (PETs)
Even if the AI needs sensitive data, privacy can still be protected.

Options include:

- **Data masking and tokenization**
- **Stochastic Randomization**
- **Homomorphic Encryption**

2. Secure the Model

Supply Chain (Model Scanning, AI-BOM AI-SPM)

This area is analogous with traditional Software Supply Chain security and covers the risks to code and artifacts in the AI/ML development lifecycle, including malicious code analysis, AI Bill-of-Materials, and AI Security Posture Monitoring, Source Dependencies, and provenance of artifacts from source to prompt .

Red Teaming/Vulnerability Scanning

Systematically probing both:

- **Models** that serve as central components for the applications, and
- **Systems** and **Data** used throughout the lifecycle of the application:

From model development and training, through application staging pipelines, and continuously in production runtime environments. Combines traditional adversarial testing with AI-specific methodologies, addressing risks like: **Prompt injection, Toxic outputs, Model extraction, Bias, Knowledge risks and Hallucinations.**

3.Secure the Usage of the App

AI Runtime Security

Protecting AI systems *while they are running*.

AI applications can be **tricked into leaking sensitive info, generate harmful or biased content**, and/or t**ake risky actions** based on bad inputs

AI Firewall-filtering inputs and outputs to block:

- Malicious prompts (like jailbreaks)
- Inappropriate or dangerous responses
- Unauthorized data access

AI Guardrails -rules and limits that keep the AI on track. They guide the AI to:

- Stick to **approved topics**
- Avoid **risky actions**
- Always follow **company policies**

Combined, the **AI Firewall** and **Guardrails** help enterprises ensure their AI behaves **safely, ethically, and within bounds** AI Runtime tools should have ability to filter both input and responses.

Security of AI Usage

Security of Usage refers to any 3rd party AI usage and covers a broad area of security control types that can play a partial role in securing AI usage. Product areas include that offer security features for AI usage include:

SASE/SSE

AI Focused (Browser Inserted)

AI Focused (Network Inserted)

Enterprise Browser/Browser Extensions

Next Gen DLP

Digital Workforce Security (Security of AI Agents)

Code Development Focused

SSPM (SaaS Security Posture Management)

Each Category has strengths and weaknesses and a complete program for using AI securely should start with a discovery exercise to gain a clear understanding of the current AI tools that are being used across the organization.

Security of AI: WWT Market Landscape (v05122025)

World Wide Technology

AI Security Governance, Risk & Compliance

CISCO · PROTECT AI · truyo · credo ai · Tumeryk AI Trust Score™ · Pillar · CRANIUM · Holistic AI · fairnow · Enkrypt AI · preamble · FAIRLY · Citrus^x

AI Discovery & Inventory

1st Party Apps

CISCO · PROTECT AI · WIZ · truyo · Pillar · AppSOC^AI

CISCO · zscaler · paloalto networks · Singulr AI · WITNESS AI · onyx

3rd Party Apps

Acuvity · Aurascape · unbound security · portal26 · SurePathAI

3P-SaaS

nudge · Reco · suridata

Security of AI Systems

1. Secure the Data

2. Secure the Model

3. Secure the Usage of the App

Risk Assessment

KNOSTIC · opsin · Aurascape

Discovery, Classification, Labeling

BigID · CYERA · securiti · VARONIS · CROWDSTRIKE · paloalto networks · Microsoft · LightBeam.ai · RELYANCE AI · CONCENTRIC AI

Data Access

Polymer · auth0 by Okta · KNOSTIC · Velotix · securiti · privacera

Privacy Enhancing Techniques

PROTOPIA · privacera · ANTIMATTER · MIRROR Security · skyflow · PROTECTO · PRIVATE AI · inpher

Supply Chain (Model Scanning, AI-BOM AI-SPM)

PROTECT AI · paloalto networks · WIZ · Pillar · CISCO · REVERSINGLABS · HIDDENLAYER · NOMA

Red Teaming/Vulnerability Scanning

NVIDIA · splx AI · Lakera · PROTECT AI · MINDGARD · ADVERSA · CISCO · REPELLO AI · promptfoo · LASSO · Calypso AI · NOMA · Galileo · GRAY SWAN · Javelin · TROJ.AI · dreadnode · Virtue AI · TESTSAVANT · MIRROR Security · Enkrypt AI · HIDDENLAYER · Haize Labs

AI Runtime Security

NVIDIA · PROTECT AI · paloalto networks · CISCO · Akamai · f5 · CLOUDFLARE · Prompt: · Calypso AI · Lakera · LASSO · NOMA · Pillar · AIM · ActiveFence · Virtue AI · Acuvity · AIShield Powered by Bosch · Deepkeep · Guardrails AI · Javelin · APEX · Arthur · Galileo · pangea · Tumeryk AI Trust Score™ · Haize Labs · Enkrypt AI · HIDDENLAYER · Liminal · Straiker · TROJ.AI · Operant

Security of AI Usage

SASE/SSE

paloalto networks · CISCO · zscaler · FORTINET · netskope · CHECK POINT

GenAI Focused – (Browser Insertion)

Prompt: · onyx · harmonic · AIM · Singulr AI · LASSO · Acuvity · Aurascape · magicmirror

GenAI Focused – (Network MITM)

WITNESS AI · portal26 · SurePathAI · Singulr AI

Enterprise Browser/Browser Extension

LayerX · SquareX · Island · paloalto networks · keep aware · SURF · MENLO SECURITY · primary · SERAPHIC

Next-Gen AI-Focused DLP

cyberhaven · CYERA · harmonic · Nightfall AI™ · magicmirror · next

Digital Workforce (Security of Agents)

onyx · Singulr AI · Acuvity · zenity

Code Development Assistants

Prompt: · Lumeus · LASSO

AI for Cyber Security: WWT Market Landscape (v0512.2025)

World Wide Technology

AI-Driven SOAR/Hyper automation

- torq
- mindflow
- blinkops
- tines
- SWIMLANE

AI-Driven SIEM

- splunk>
- paloalto
- CROWDSTRIKE
- SentinelOne
- Chronicle part of Google Cloud
- elastic
- exabeam
- Gurucul
- ANVILOGIC

AI-Driven DSPM & DLP

- CYERA
- securiti
- LightBeam.ai
- CONCENTRIC AI
- RELYANCE AI
- cyberhaven
- Nightfall AI

AI-Driven Malware Analysis

- deep instinct
- THREATRAY
- REVERSING LABS

AI-enabled EDR

- CROWDSTRIKE
- Microsoft
- paloalto
- SentinelOne

AI-driven SSPM

- nudge
- suridata
- Spin.ai
- Reco
- vorlon

AI-Driven NDR

- VECTRA
- DARKTRACE
- LiveAction
- MixMode
- Cynamics
- ExtraHop
- STELLAR CYBER

AI-Driven Bot/Fraud

- Arkose Labs
- HUMAN
- kasada
- DEDUCE
- STYTCH
- trustmi

AI-Driven Email Security

- Abnormal
- Sublime Security
- Material
- TESSIAN
- INKY

Autonomous SOC Agents

- exaforce
- prophet
- CROGL
- Simbian
- Dropzone AI
- RadiantSecurity
- 7AI
- INTEZER
- Qevlar AI
- CULMINATE
- AIRMDR
- Conifers

AI-driven Threat Intelligence

- CYBLE.
- cybersixgill
- WRAITHWATCH
- alphaMountain
- feedly

AI-driven Ransomware Protection

- MIMIC
- halcyon
- elastio
- alcion

AI-driven Threat/Vuln Management

- Avalor a zscaler company
- Reach
- Nucleus
- SKYHAWK SECURITY
- CYCLOPS
- Filigran

AI-driven Pen Testing

- XBOW
- PENTERA
- HORIZON3.ai
- RUNSYBIL
- HADRIAN
- Cobalt
- aikido
- RIDGE SECURITY

AI-enabled IAM

- BeyondTrust
- Silverfort
- OASIS
- CLUTCH
- ConductorOne
- SailPoint
- sgnl
- PingIdentity ForgeRock
- RADIANT LOGIC
- veza
- Delinea
- anetac

AI-Driven AppSec & Software Supply Chain

- mobb
- Phylum
- Socket
- ENDOR LABS
- qwiet AI
- Pixee
- APTORI
- XEOL
- Semgrep
- Kodem
- amplify
- Operant
- Cognition
- STARIS
- akto
- nullify
- alma

AI-driven External Attack Surface Management

- IONIX
- DARKTRACE
- CYBLE.
- Outpost24
- censys

AI-driven Security Awareness

- Riot
- Redcoat AI
- OPENCYBER AI
- Jericho Security
- MIRAGE

Security from Adversarial AI (Deepfake & Misinformation): WWT Market Landscape (v05.13.2025.76)

World Wide Technology

Deepfake Detection

Reality Defender
Pindrop
Neural Defend
SENTINEL
sensity
Clarity
validsoft
GetReal
deeptrust
Q - Integrity
Buster.Ai
DUCK DUCK GOOSE
kroop AI
DARPA
hiya
AI OR NOT
DEEP MEDIA
deepware®
netarx
UncovAI
Breacher.ai DEFEND WITH KNOWLEDGE
GPTZero
intel.
IDENTITIES.AI

Content Provenance

Microsoft
PIXELSTREAM
Numbers
CertifiedTrue
truepic
HIVE
Adobe
sightengine
C2PA
Attestiv
HiddenTag®
ILLUMINARTY

Misinformation, Media & Social Networks

BLACKBIRD.AI
Doppel
Cyabra
BLACKCLOAK
Logically.
LETSDATA
ZEROFOX
Cyberint A Check Point Company
CYFIRMA DECODING THREATS
WeVerify
Adaptive
VINESIGHT

Deepfake Synthesis/Red Teaming

ElevenLabs
DEEPFAKES WEB
AKOOL
RESEMBLE.AI
SYNTHESYS
EmbodyMe
KAPWING
D-ID
dobstudio

Biometric Identity & Liveness Verification

iProov
iDenfy
sardine
jumio.
AU10TIX IDENTITY INTELLIGENCE
incode
BioID be recognized
VeriDas
ID R&D a Mitek company
veriff
FACIA
HYPERVERGE
Nametag
PRIVASEA
world
FaceOnLive
AUTHENTICID
onfido an Entrust Company
sumsub

Thanks for Attending

World Wide Technology

Todd Hathaway
Todd.Hathaway@wwt.com
www.linkedin.com/in/toddhathaway