

Aavista Oy
Merja Kajava

AI Agents and Secure API Access

APIOps Helsinki 2026



AAVISTA



AAVISTA

Photo credit: NASA / Paul E. Alers

AI agents embedded throughout the R&D life-cycle

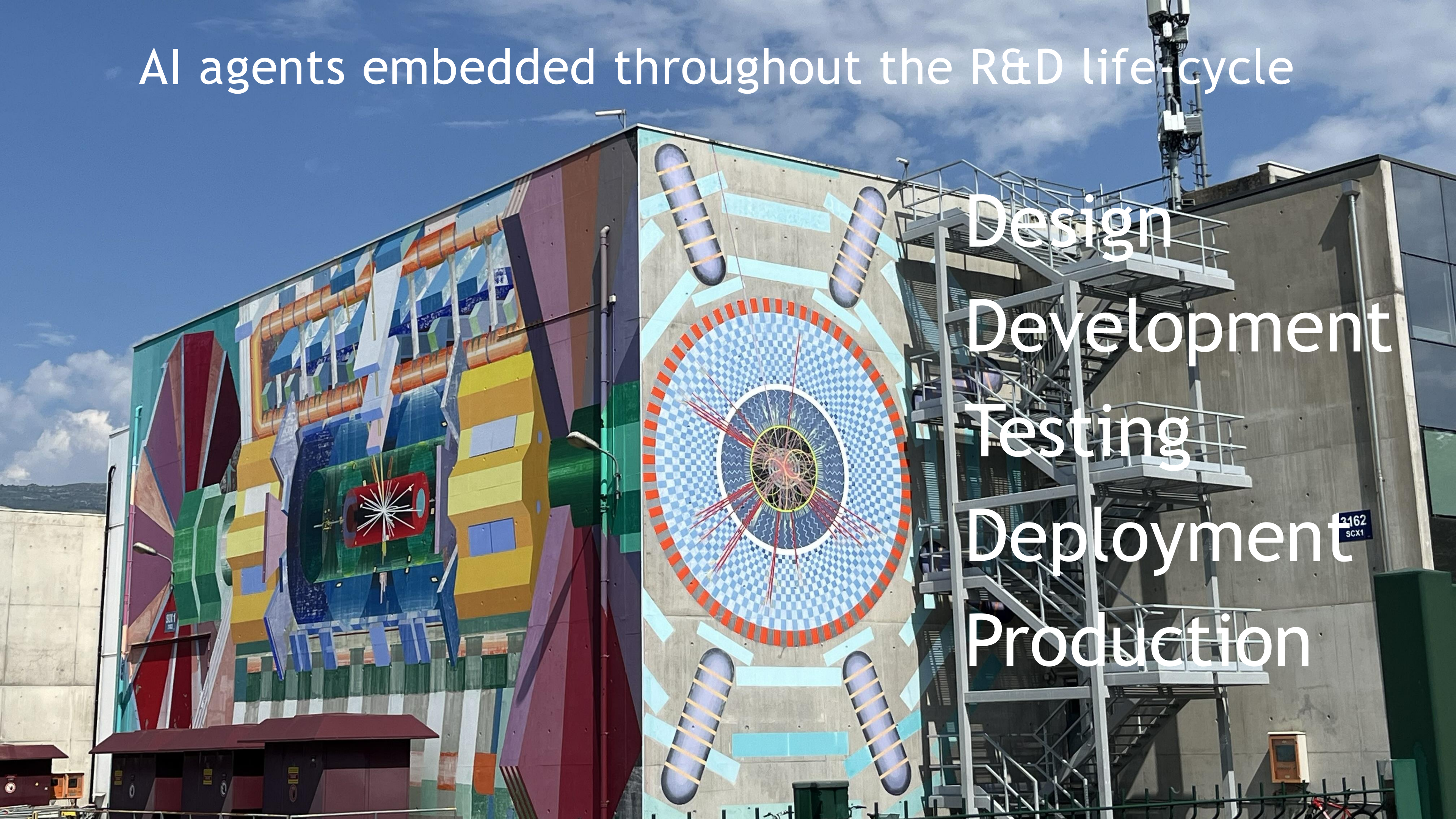
Design

Development

Testing

Deployment

Production



What are the elements of Agentic AI



Orchestration

Goals, profiles, skills

Model

Reason with goals

Tools

Integrate to data and 3rd party services and APIs



AAVISTA

AI agents as **API consumers** and **API providers**

Different types of threats are emerging

Reconnaissance &	Resource Development &	Initial Access &	AI Model Access	Execution &	Persistence &	Privilege Escalation &	Defense Evasion &	Credential Access &	Discovery &	Lateral Movement &	Collection &	AI Attack Staging	Command and Control &	Exfiltration &	Impact &
8 techniques	13 techniques	7 techniques	4 techniques	6 techniques	9 techniques	4 techniques	15 techniques	6 techniques	9 techniques	2 techniques	4 techniques	6 techniques	3 techniques	6 techniques	9 techniques
Active Scanning &	Acquire Infrastructure	AI Supply Chain Compromise	AI Model Inference API Access	AI Agent Clickbait	AI Agent Context Poisoning	AI Agent Tool Invocation	AI Supply Chain Reputation Inflation	AI Agent Tool Credential Harvesting	Cloud Service Discovery &	Phishing &	AI Artifact Collection	Craft Adversarial Data	AI Agent	Exfiltration via AI Agent Tool Invocation	Cost Harvesting
Gather RAG-Indexed Targets	Acquire Public AI Artifacts	Drive-by Compromise &	AI-Enabled Product or Service	AI Agent Tool Invocation	AI Agent Tool Data Poisoning	AI Agent Tool Invocation	AI Supply Chain Rug Pull	Credentials from AI Agent Configuration	Discover AI Agent Configuration	Use Alternate Authentication Material &	Data from AI Services	Create Proxy AI Model	AI Service API	Exfiltration via AI Inference API	Data Destruction via AI Agent Tool Invocation
Gather Victim Identity Information &	Develop Capabilities &	Evade AI Model	Full AI Model Access	Command and Scripting Interpreter &	AI Agent Tool Poisoning	Command and Scripting Interpreter &	Corrupt AI Model	Exploitation for Credential Access &	Discover AI Artifacts		Data from Information Repositories &	Generate Deepfakes	Reverse Shell	Exfiltration via AI Inference API	Denial of AI Service
Search Application Repositories	Establish Accounts &	Exploit Public-Facing Application &	Physical Environment Access	Deploy AI Agent	LLM Prompt Self-Replication	Deploy AI Agent	Delay Execution of LLM Instructions	OS Credential Dumping &	Discover AI Model Family		Data from Local System &	Generate Malicious Commands		Exfiltration via Cyber Means	Erode AI Model Integrity
Search Open AI Vulnerability Analysis	LLM Prompt Crafting	Phishing &		LLM Prompt Injection	Manipulate AI Model	LLM Prompt Injection	Evade AI Model	RAG Credential Harvesting	Discover AI Model Ontology			Manipulate AI Model		Extract LLM System Prompt	Erode Dataset Integrity
Search Open Technical Databases &	Obtain Capabilities &	Prompt Infiltration via Public-Facing Application		User Execution &	Modify AI Agent Configuration	User Execution &	Exploitation for Defense Evasion &	Unsecured Credentials &	Discover AI Model Outputs			Verify Attack		LLM Data Leakage	Evade AI Model
Search Open Websites/Domains &	Poison Training Data	Valid Accounts &			Poison Training Data		False RAG Entry Injection		Discover LLM Hallucinations					LLM Response Rendering	External Harms
Search Victim-Owned Websites &	Publish Hallucinated Entities				Prompt Infiltration via Public-Facing Application		Impersonation &		Discover LLM System Information						Machine Compromise
	Publish Poisoned AI Agent Tool				RAG Poisoning		LLM Jailbreak		Process Discovery &						Spamming AI System with Chaff Data
	Publish Poisoned Datasets						LLM Prompt Obfuscation								
	Publish Poisoned Models						LLM Trusted Output Components Manipulation								
	Retrieval Content Crafting						Manipulate User LLM Chat History								
	Stage Capabilities &						Masquerading &								
							Modify AI Agent Configuration								
							Virtualization/Sandbox Evasion &								

Compliance requirements are expanding

EU regulation, e.g.

CRA

DORA

AI Act

EU directives, e.g.

GDPR

NIS2

Product Liability Directive “new PLD”



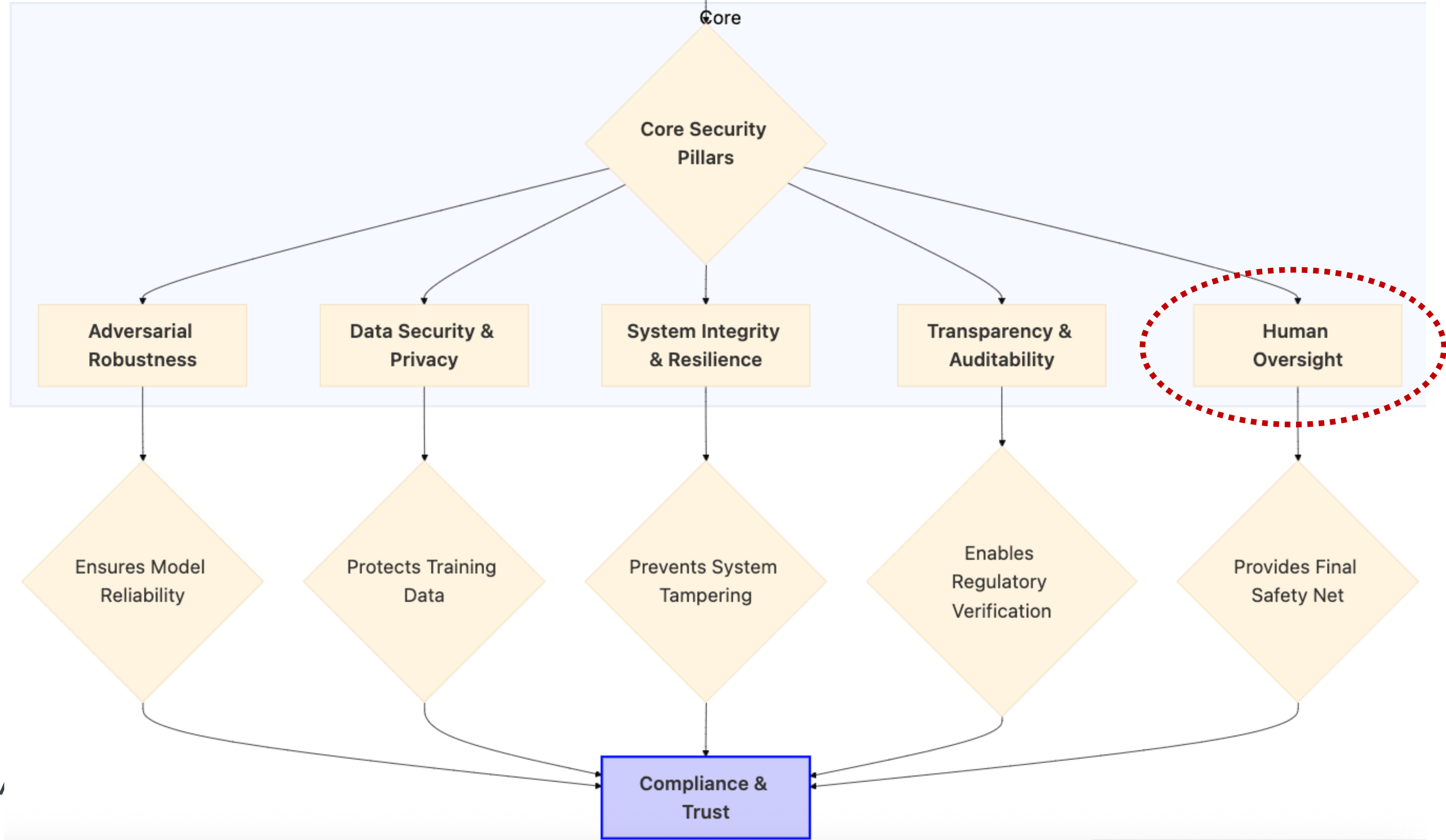
AAVISTA



Security pillars in AI Act

EU AI Act: High-Risk System

Four risk classes



AI

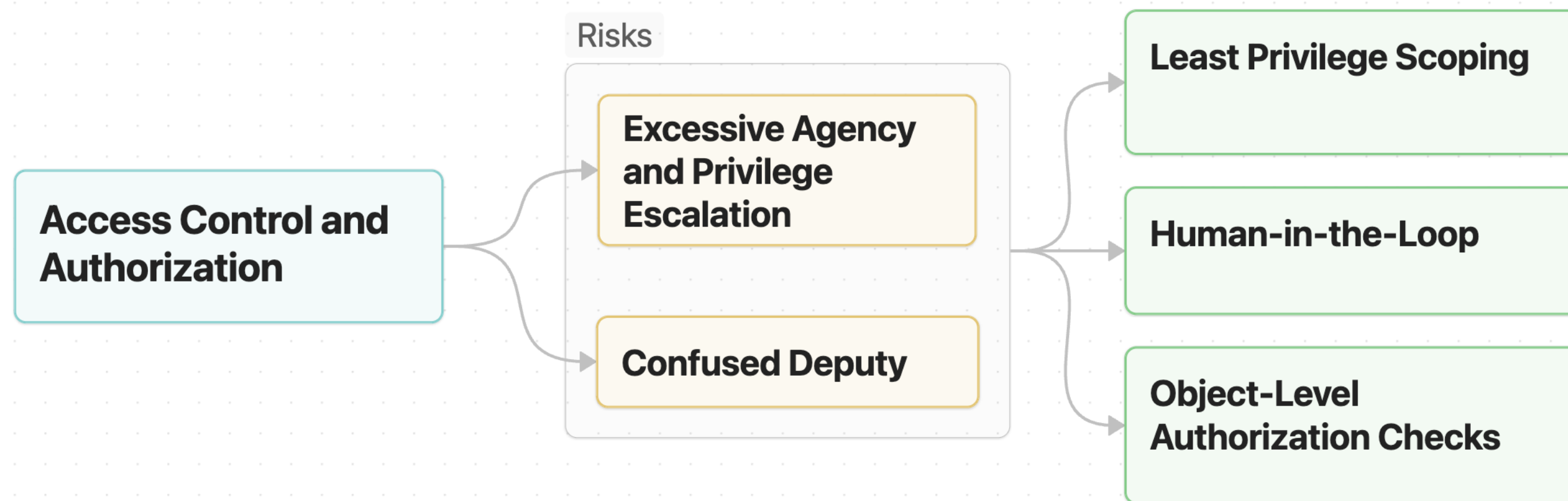
A world map with a dark blue background, where city lights are represented by glowing yellow and white dots of varying sizes, primarily concentrated in North America, Europe, and East Asia.

How to secure API access

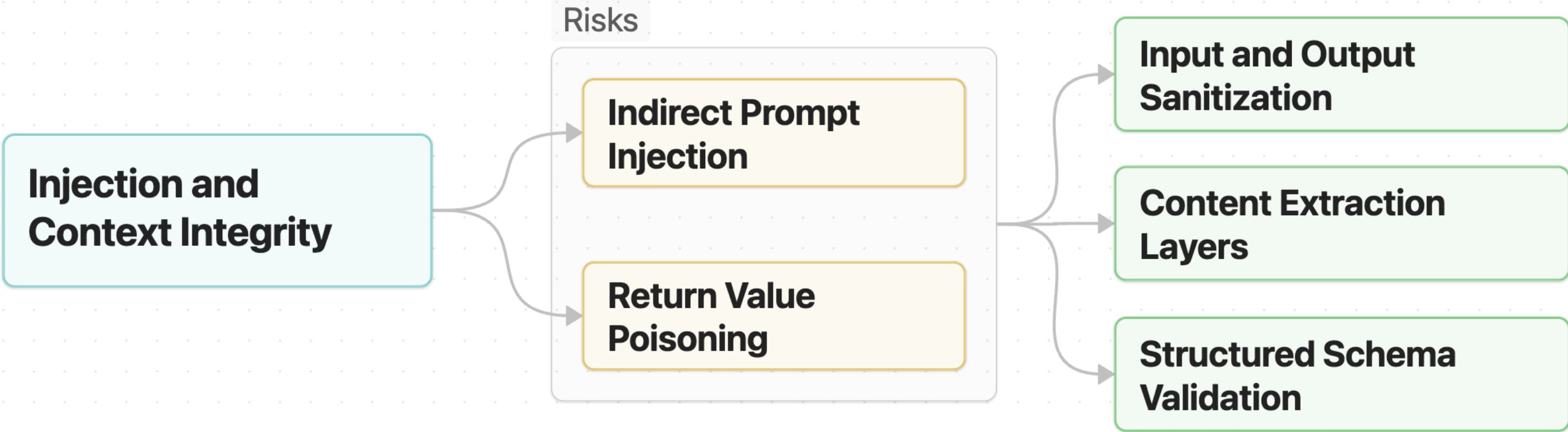


AAVISTA

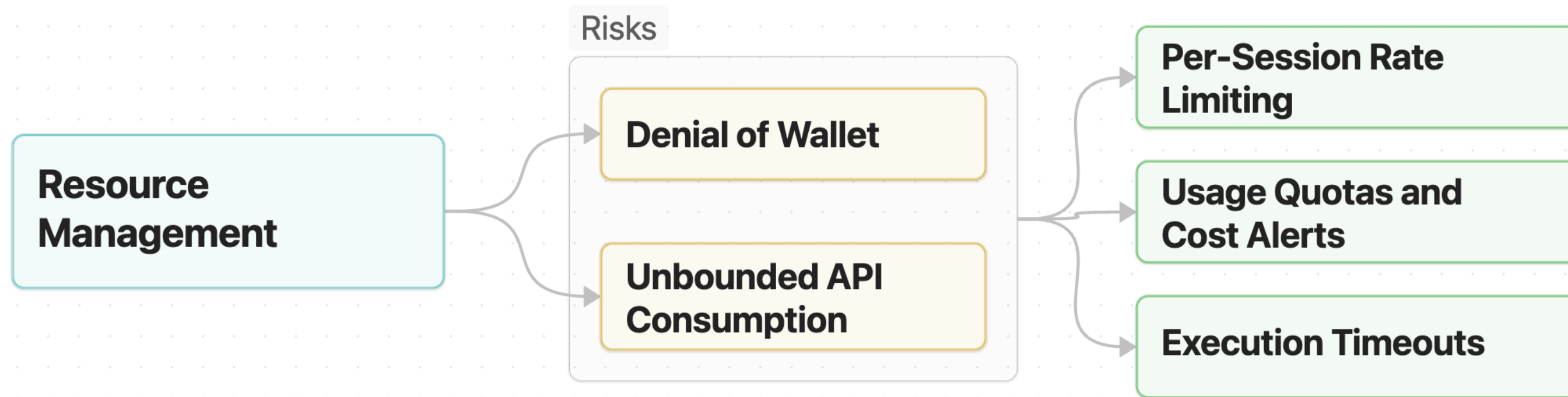
Access control and authorization



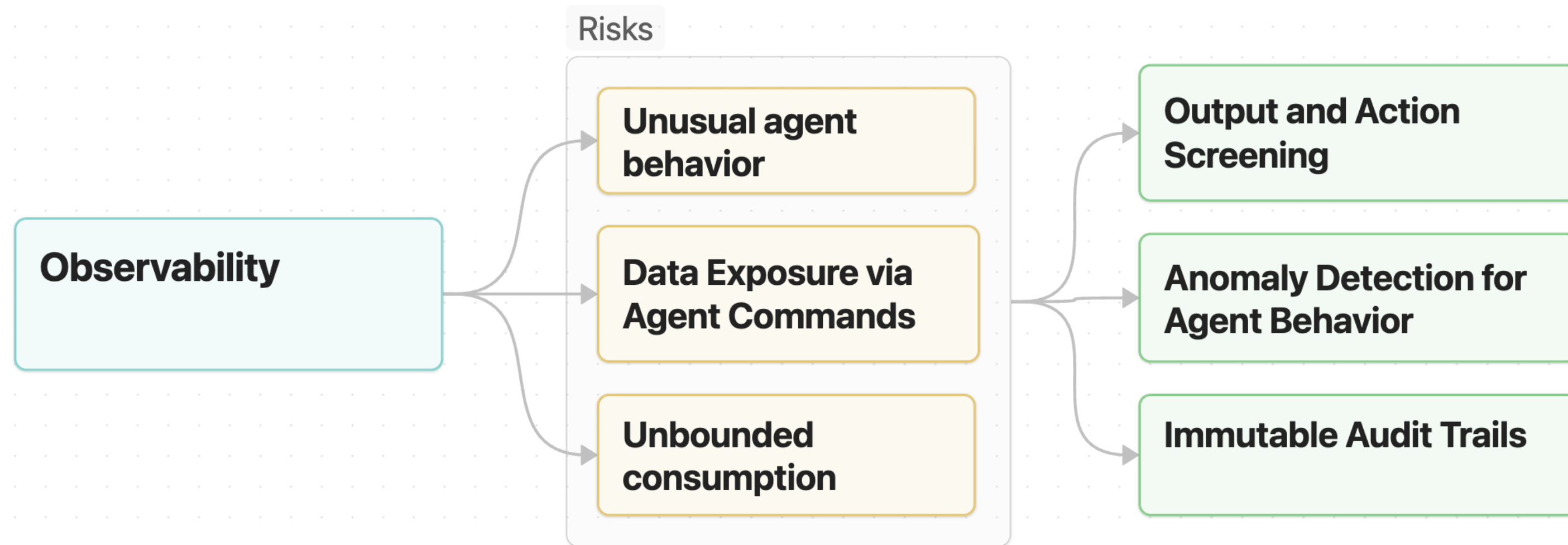
Injection and context integrity



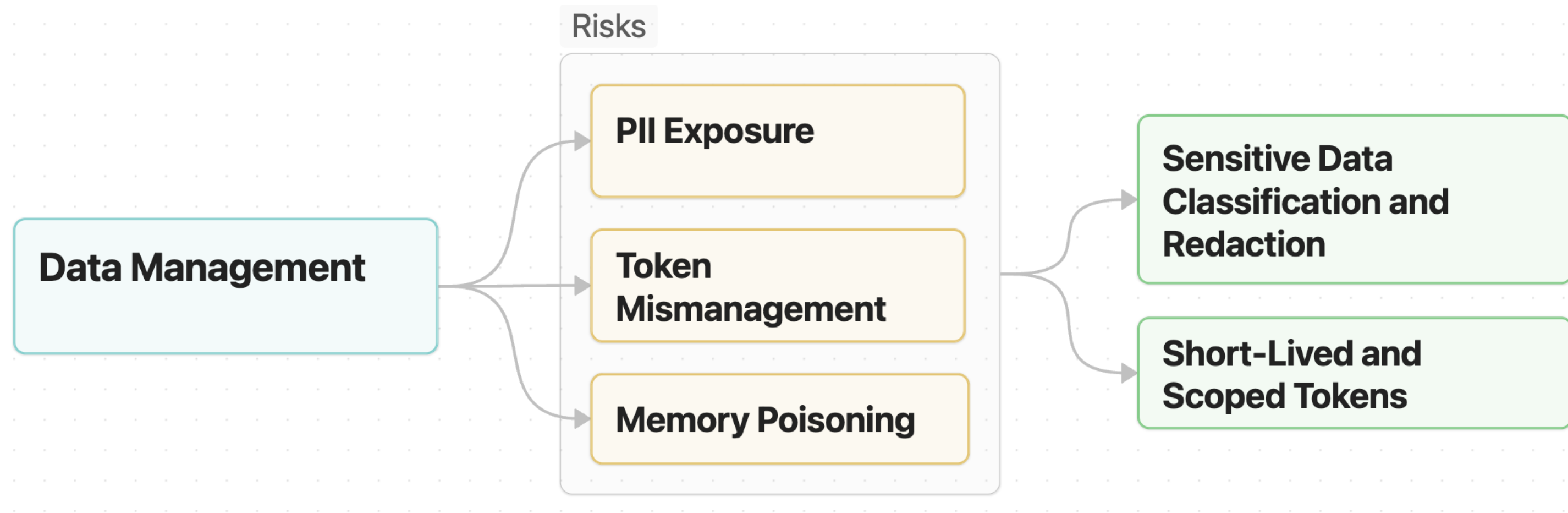
Resource management



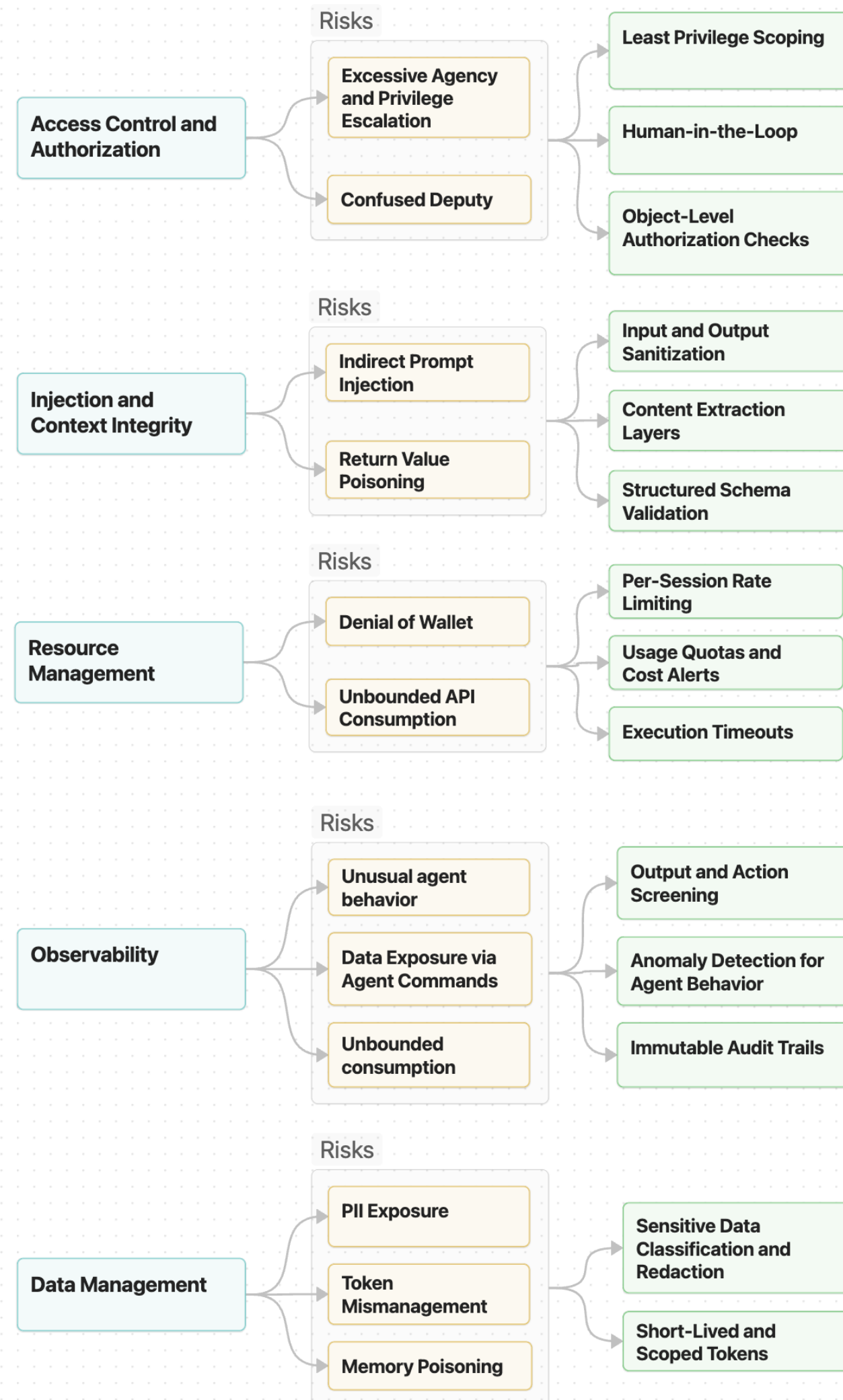
Monitoring and observability



Data management



Focus on
observability,
reaction time, and
notifying stakeholders



AAVISTA

Key takeaways

Govern the AI.

Allow only the **minimal access**.

Design and build **observability** from the start.

Remember **Human-in-the-Loop**.



AAVISTA



Technology is the answer, but
what was the question?

Cedric Price





AAVISTA

The Data Refinery Company



Merja Kajava

<https://www.linkedin.com/in/merjakajava>