Agentic AI for Cybersecurity: Automating Risk Management with Intelligent Agents

ELNAZ EBADI

ISCISC 2025

TAPSI

انجمن رمزایران
Iranian Society of Cryptology

Meet the Presenter

- ✓+8 years of experience in Security Engineering
- ✓Master’s degree in telecommunication Engineering at UT
- ✓Master of Business administration at SUT
- ✓2 years of security research at ITRC
- ✓4 years of Pentesting Banking Solutions at TOSAN
- ✓+2 year of Risk and Compliance Team Leading at TAPSI

TAPSI

انجمن رمزایران
iranian Society of Cryptology

TAPSI

- ✓The first startup company in IT section entered the stock market
- ✓The first startup company in IT section with ISO 27001 since 1400
- ✓The first startup company in IT section with ISO 27001: 2022

TAPSI

انجمن رمزایران
iranian Society of Cryptology

Need of AI in Cyber Security

Why Cybersecurity Demands Agentic AI

The Data Deluge:

- Millions of security events per day.
- Terabytes of logs from endpoints, networks and security solutions.
- "Alert fatigue" is real – critical signals are lost in the noise.

The Skills Gap:

- Shortage of ~4 million cybersecurity professionals globally.
- Existing teams are overworked and focused on reactive firefighting.

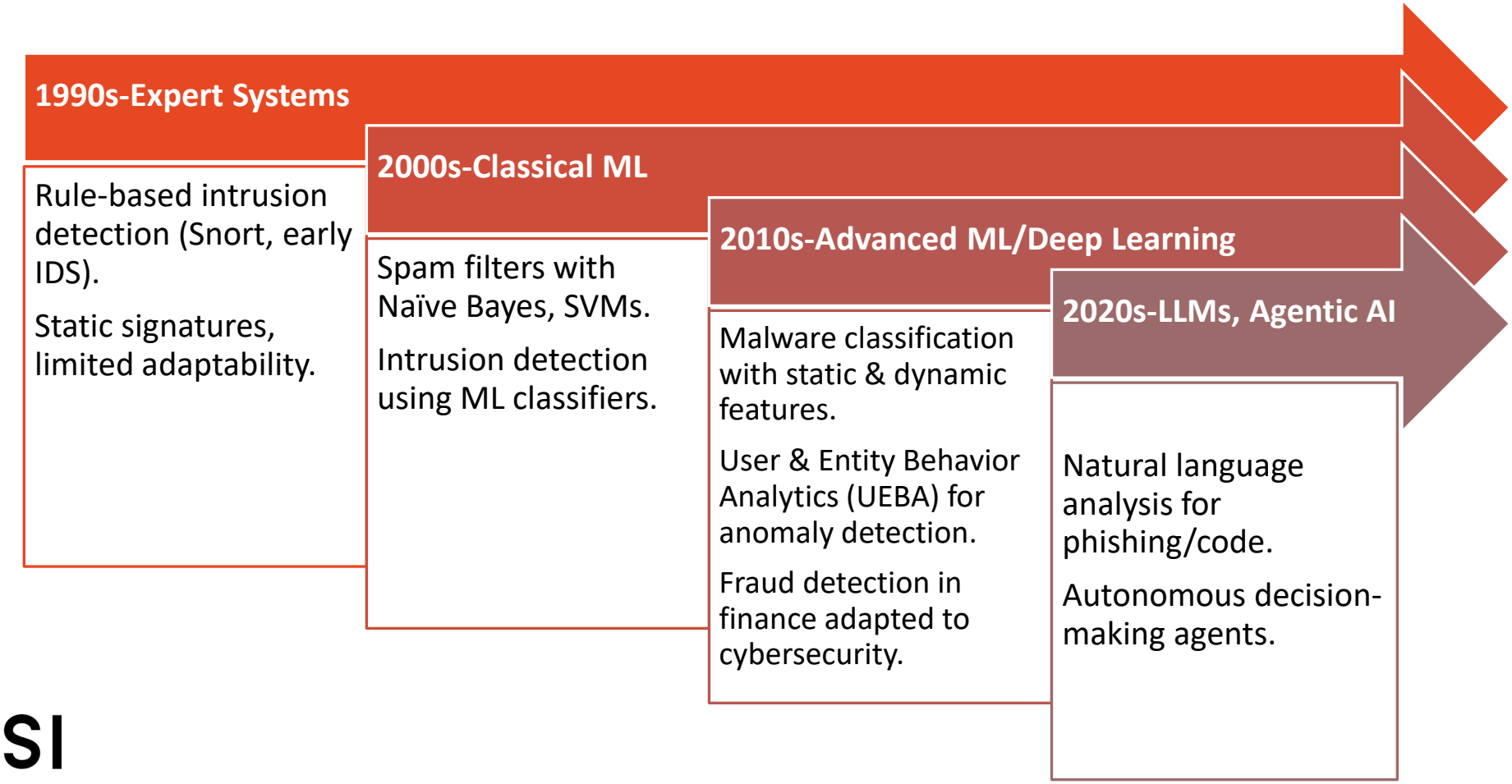
The Adversary's Advantage:

- AI-powered attacks are already here (e.g., automated phishing, malware generation).
- Attackers operate at machine speed; defenders are often on human time.

TAPSI

انجمن رمزایران
Iranian Society of Cryptology

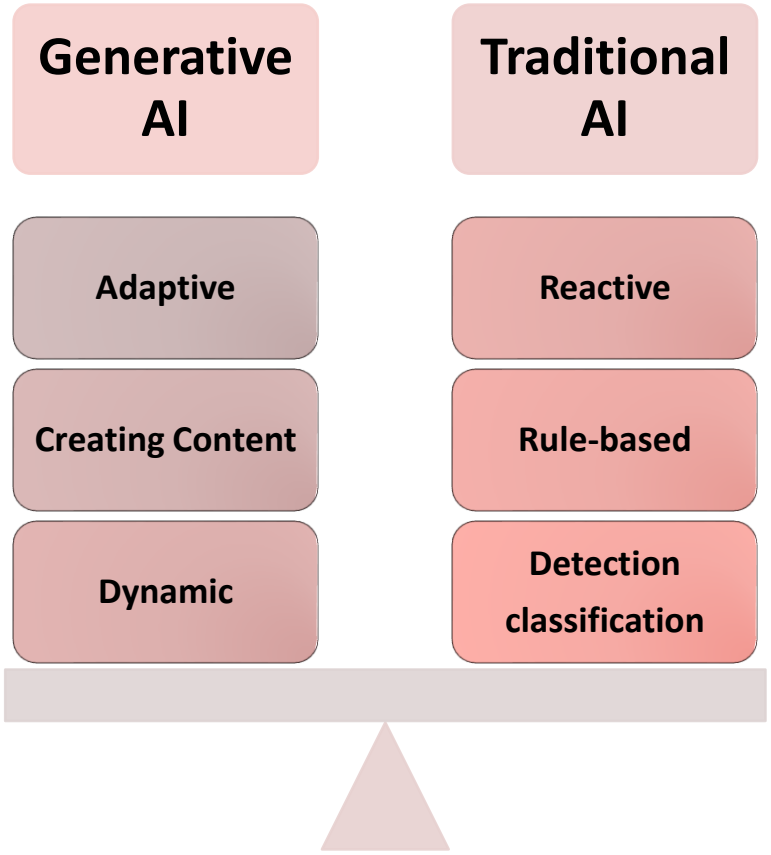
Not Just About LLMs



TAPSI

انجمن رمزایران
Iranian Society of Cryptology

Not Just About LLMs

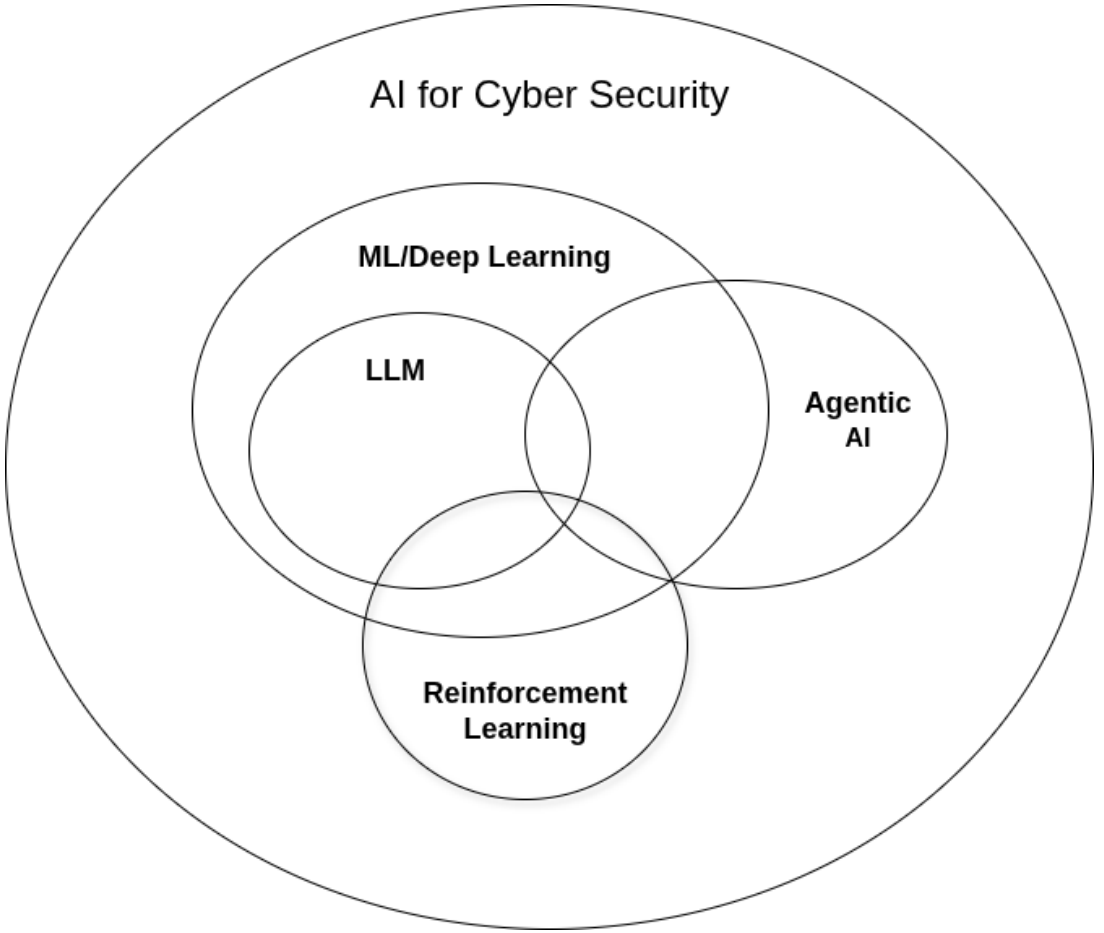


TAPSI

انجمن رمزایران
iranian Society of Cryptology

Not Just About LLMs

- Intrusion Detection & Anomaly Detection
- Malware classification
- Autonomous defense strategies
- Spam & Phishing Detection
- Fraud Detection
- User & Entity Behavior Analytics



TAPSI

انجمن رمزایران
Iranian Society of Cryptology

Use cases of AI in Cyber Security

Where is AI in Cybersecurity Today

Threat Detection & Prevention

Risk Management & Compliance

Incident Response & Automation

Identity & Access Security

Fraud & Financial Security

•**Key Message:** AI is no longer a single tool; it's a layer of intelligence across the entire security stack.

TAPSI

انجمن رمزایران
Iranian Society of Cryptology

AI in Threat Detection

Advanced Threat Detection

- ✓**The Problem:** Legacy SIEMs and signature-based tools miss novel, low-and-slow, and fileless attacks.
- ✓**How AI is Used:**
  - ✓**User and Entity Behavior Analytics (UEBA):** Baselines normal behavior for every user and device. Flags significant deviations (e.g., a user accessing data at 3 AM from a foreign country).
  - ✓**Network Traffic Analysis (NTA):** Uses ML to model normal network traffic patterns and detect C2 communication, data exfiltration, and lateral movement.
- ✓**Business Value:** Reduces false positives, detects insider threats, and identifies unknown attacks.

TAPSI

انجمن رمزایران
iranian Society of Cryptology

AI in Threat Detection

UEBA in Action

Core Technique: Unsupervised and Supervised Machine Learning.

Data Sources:

- Authentication Logs (Active Directory, VPN)
- Network Flow Data (NetFlow)
- Endpoint Activity (Process execution, file access)

The ML Process:

- **Feature Engineering:** Creating metrics like "logon frequency," "data volume accessed," "geographic velocity."
- **Model Training:** Clustering users by role to establish peer-group baselines.
- **Scoring & Alerting:** Using algorithms like Isolation Forests or Autoencoders to score behavior anomalies and generate a risk score.

TAPSI

انجمن رمزایران
iranian Society of Cryptology

AI in Phishing Detection

Phishing and Email Security

- ✓**The Problem:** Phishing emails are increasingly sophisticated and personalized, bypassing traditional URL blocklists and signature-based filters.
- ✓**How AI is Used:**
  - ✓**Natural Language Processing (NLP):** Analyzes email content for social engineering tactics, urgency, and sentiment.
  - ✓**Computer Vision:** Analyzes logos and images in emails to detect brand impersonation.
  - ✓**Link Analysis:** Executes URLs in sandboxes to analyze behavior in real-time, rather than just checking a blocklist.
- ✓**Business Value:** Catches zero-day phishing campaigns and Business Email Compromise (BEC) attempts that lack malicious payloads.

TAPSI

انجمن رمزایران
iranian Society of Cryptology

AI in Phishing Detection

Phishing and Email Security

- ✓**Core Technique:** Natural Language Processing (NLP) with Transformer Models (like BERT).
- ✓**Feature Extraction:**
  - ✓**Semantic Analysis:** Does the email's intent match a known phishing template (e.g., "Verify your account," "Urgent payment required")?
  - ✓**Syntax & Grammar:** Poor grammar and odd formatting can be signals.
  - ✓**Entity Recognition:** Identifying and validating senders, company names, and phone numbers against known good lists.
- ✓**The Technical Process:**
  - ✓**Tokenization:** Breaking the email text into words/sub-words.
  - ✓**Embedding:** Converting tokens into numerical vectors.
  - ✓**Classification:** Feeding the vectors into a classification model (e.g., a Neural Network) to output a probability: Phishing: 97%.

TAPSI

انجمن رمزایران
iranian Society of Cryptology

AI in Phishing Detection

Phishing and Email Security

- Pure Machine Learning
- API Call
- Local LLM

TAPSI

انجمن رمزایران
Iranian Society of Cryptology

AI in Phishing Detection

Phishing and Email Security

Pure Machine Learning

- ✓Finding a Good Dataset
- ✓Analyzing and Cleaning Data
- ✓Train and Test

TAPSI

انجمن رمزایران
Iranian Society of Cryptology

AI in Phishing Detection

Phishing and Email Security

API Call

OpenRouter

OpenAI

TAPSI

انجمن رمزایران
iranian Society of Cryptology

AI in Phishing Detection

Phishing and Email Security

Local LLM

Ollama is an open-source tool that allows you to run large language models (LLMs) like GPT and LLaMA directly on your personal computer

TAPSI

انجمن رمزایران
Iranian Society of Cryptology

AI in EDR

Endpoint Detection & Response (EDR)

- ✓**The Problem:** File-based antivirus is obsolete. Modern attacks live in memory, scripts, and legitimate system tools (Living-off-the-Land).
- ✓**How AI is Used:**
  - ✓**Behavioral AI:** Monitors process execution chains, API calls, and network connections in real-time to detect malicious sequences (e.g., powershell.exe spawning rundll32.exe to make a network call).
  - ✓**Static File Analysis:** Uses ML models to analyze file attributes and code to predict if a file is malicious, even without a signature.
- ✓**Business Value:** Stops ransomware, fileless malware, and advanced persistent threats (APTs) that bypass traditional AV.

TAPSI

انجمن رمزایران
Iranian Society of Cryptology

AI in Malware Detection

Malware Classification Models

- ✓**Core Technique:** Supervised Machine Learning (Classification) and Static Analysis.
- ✓**Feature Engineering (The "What to Look For"):**
  - ✓**PE Header Analysis:** (For Windows files) Checks sections, imports, timestamps.
  - ✓**N-gram Analysis:** Looks for sequences of bytes or assembly instructions common in malware.
  - ✓**Entropy Analysis:** Measures file randomness (high entropy can indicate encryption/packing).
- ✓**The ML Process:**
  - ✓**Training Dataset:** Millions of samples of known "goodware" and "malware."
  - ✓**Model Training:** Algorithms like Gradient Boosted Trees (XGBoost) or Neural Networks learn the patterns that distinguish malicious files.
  - ✓**Inference:** A new, unseen file is analyzed, its features are extracted, and the model gives a verdict: Benign: 10% | Malicious: 90%.

TAPSI

انجمن رمزایران
Iranian Society of Cryptology

AI in SOAR

Security Automation (SOAR)

- ✓**The Problem:** SOC analysts are overwhelmed with alerts, leading to slow response times and burnout.
- ✓**How AI is Used:**
  - ✓**Alert Triage & Prioritization:** NLP models read and understand alert text, correlating related alerts and assigning a risk-based priority.
  - ✓**Playbook Automation:** AI recommends or triggers pre-defined response playbooks (e.g., "This alert matches the 'Credential Dumping' playbook. Run containment steps Y and Z.").
- ✓**Business Value:** Drastically reduces Mean Time to Respond (MTTR) and frees analysts for complex investigation tasks.

TAPSI

انجمن رمزایران
iranian Society of Cryptology

Prompt Engineering for Cyber Security

✓Effective prompts are **specific**, **contextual**, and **technically accurate**.

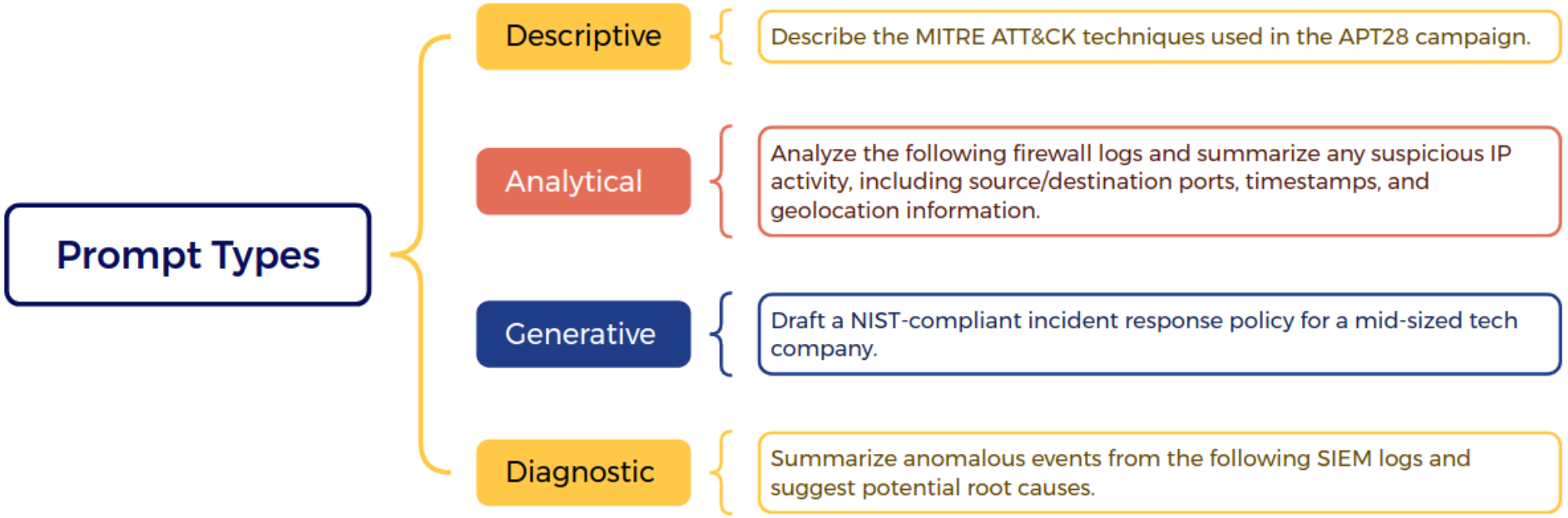
✓Few-Shot Prompting: Giving the AI a few examples of what you want before asking your actual question.

✓Prompt-Chaining: Breaking a complex task into smaller steps, where the output of the first prompt becomes the input for the next.

TAPSI

انجمن رمزایران
iranian Society of Cryptology

Prompt Engineering for Cyber Security



TAPSI

انجمن رمزایران
iranian Society of Cryptology

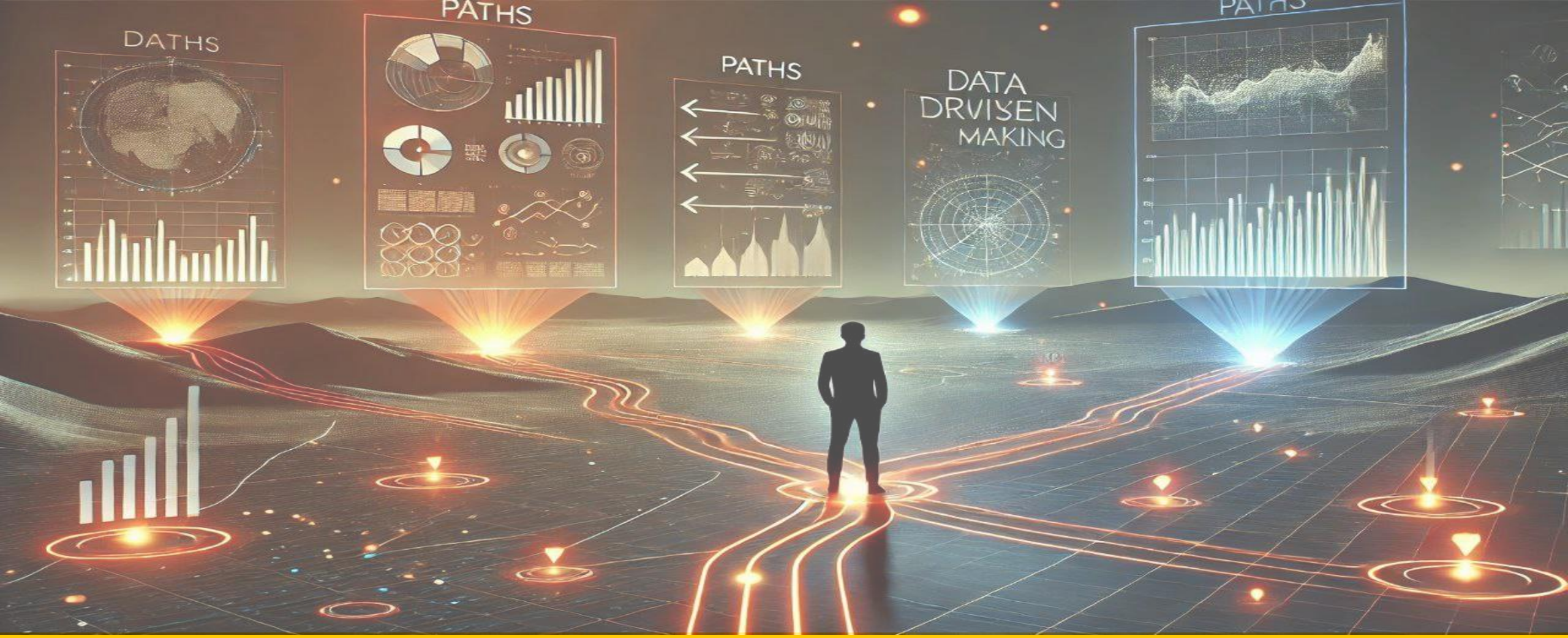
Supercharge Your AI Workflow

Get 1000+ Ready-to-Use Prompts.

Example:
"Simulate a Kubernetes incident where a misconfigured pod allows privilege escalation and hostPath access to the underlying node."

TAPSI

انجمن رمزایران
Iranian Society of Cryptology



Global Adoption Trends

Global Leaders Using AI in Cybersecurity

Palo Alto Cortex XSIAM and Its Generative AI Integrations:

- **Transforming Autonomous Cybersecurity Operations**

Cortex XSIAM combines the power of:

- **Security data lake** (aggregating logs, alerts, telemetry)
- **Behavioral analytics**
- **Machine learning**
- **SOAR (Security Orchestration, Automation, and Response)**
- **EDR/XDR capabilities (via Cortex XDR)**

…into one unified platform. The addition of **generative AI** brings **language-level intelligence and automation** into the workflow.

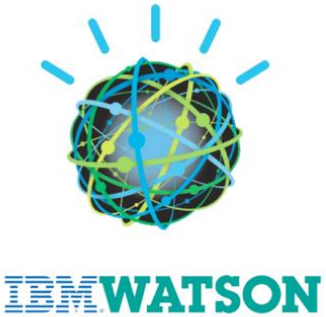
TAPSI

انجمن رمزایران
iranian Society of Cryptology

Global Leaders Using AI in Cybersecurity

IBM Security (QRadar & Watson)

- ✓AI Usage: Security Information and Event Management (SIEM) + AI reasoning.
- ✓Adaptation: Watson AI correlates logs, predicts threats, and recommends actions.
- ✓Impact: Speeds up incident response and reduces alert fatigue for analysts.



TAPSI

انجمن رمزایران
iranian Society of Cryptology

Global Leaders Using AI in Cybersecurity

Darktrace

- ✓AI Usage: Enterprise immune system using unsupervised ML & anomaly detection.
- ✓Adaptation: Monitors network behavior to detect insider threats, ransomware, and cloud risks.
- ✓Impact: Self-learning AI that adapts as the organization’s environment changes.

DARKTRACE

TAPSI

انجمن رمزایران
iranian Society of Cryptology

Global Leaders Using AI in Cybersecurity

CrowdStrike

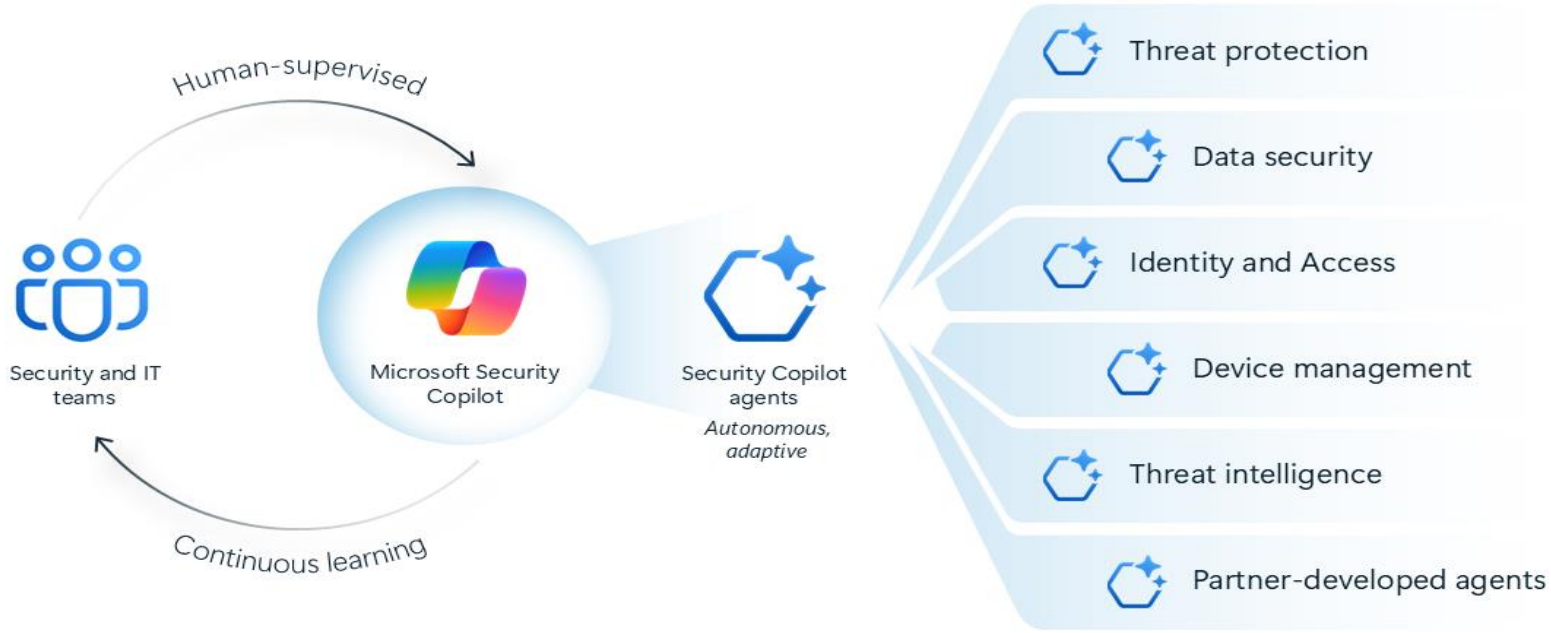
- ✓AI Usage: Endpoint detection & response (EDR) powered by Falcon AI.
- ✓Adaptation: Behavioral modeling, anomaly detection, and attack graph analysis.
- ✓Impact: Enables near real-time autonomous threat hunting at scale.

TAPSI

انجمن رمزایران
iranian Society of Cryptology

Global Leaders Using AI in Cybersecurity

Microsoft Security Copilot



TAPSI

انجمن رمزایران
Iranian Society of Cryptology

Global Trends in Academia

CONNECTED PAPERS

TAPSI

انجمن رمزایران
Iranian Society of Cryptology

Risk and Challenges of Using Agentic AI

Challenges of Using AI Agents

Transparency

- ✓AI-driven decisions can be **opaque** — hard to see why an action was taken.
- ✓Lack of visibility → reduced trust by analysts, regulators, and executives.
- ✓Risk of hidden biases influencing detection or response.

TAPSI

انجمن رمزایران
iranian Society of Cryptology

Challenges of Using AI Agents

Explainability

- ✓Security teams need to understand **why** an alert was triggered or a response was automated.
- ✓Complex models (deep learning, reinforcement learning) often lack **human-readable reasoning**.
- ✓Without explainability, it’s hard to **validate, audit, or contest AI decisions.**

TAPSI

انجمن رمزایران
iranian Society of Cryptology

Challenges of Using AI Agents

Governance

- ✓Who is **accountable** when an AI agent makes the wrong decision?
- ✓Compliance challenges: GDPR, NIS2, U.S. AI Bill of Rights, upcoming EU AI Act.
- ✓Need for **ethical frameworks and governance policies**:
  - ✓Human-in-the-loop oversight.
  - ✓Continuous monitoring of agentic AI behavior.
  - ✓Clear accountability and audit trails.

TAPSI

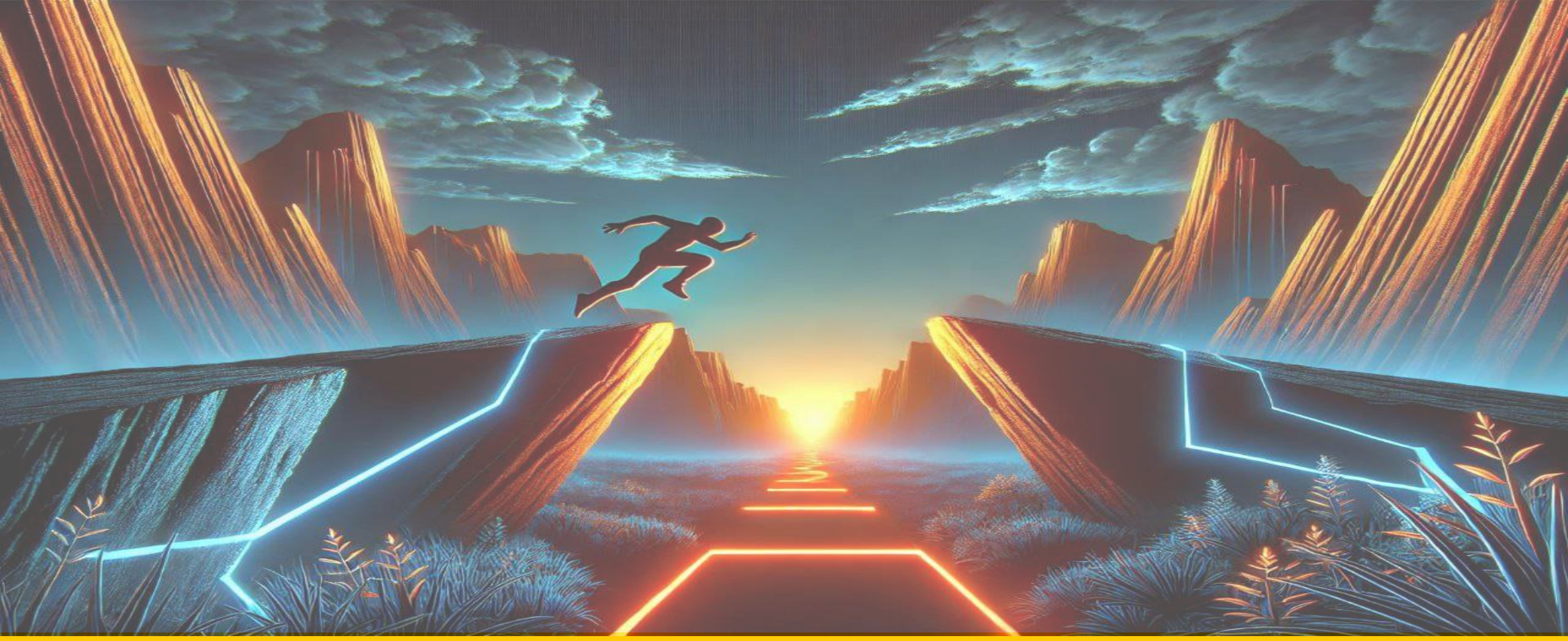
انجمن رمزایران
iranian Society of Cryptology

Key Threat Vectors and Attack Scenarios

- ✓**Weaponized Prompt Injection:** Manipulating a company's own customer service AI agent to reveal internal data or execute unauthorized actions .
- ✓**Overprivileged Agent Exploitation:** An attacker hijacks an IT automation agent with broad system permissions, causing massive disruption or data theft .
- ✓**Identity Spoofing & Fraud:** A malicious agent impersonates a legitimate procurement bot in a multi-agent workflow to authorize fraudulent payments .
- ✓**AI-Powered Reconnaissance:** Autonomously scanning and profiling an organization's digital attack surface to identify soft targets .

TAPSI

انجمن رمزایران
iranian Society of Cryptology

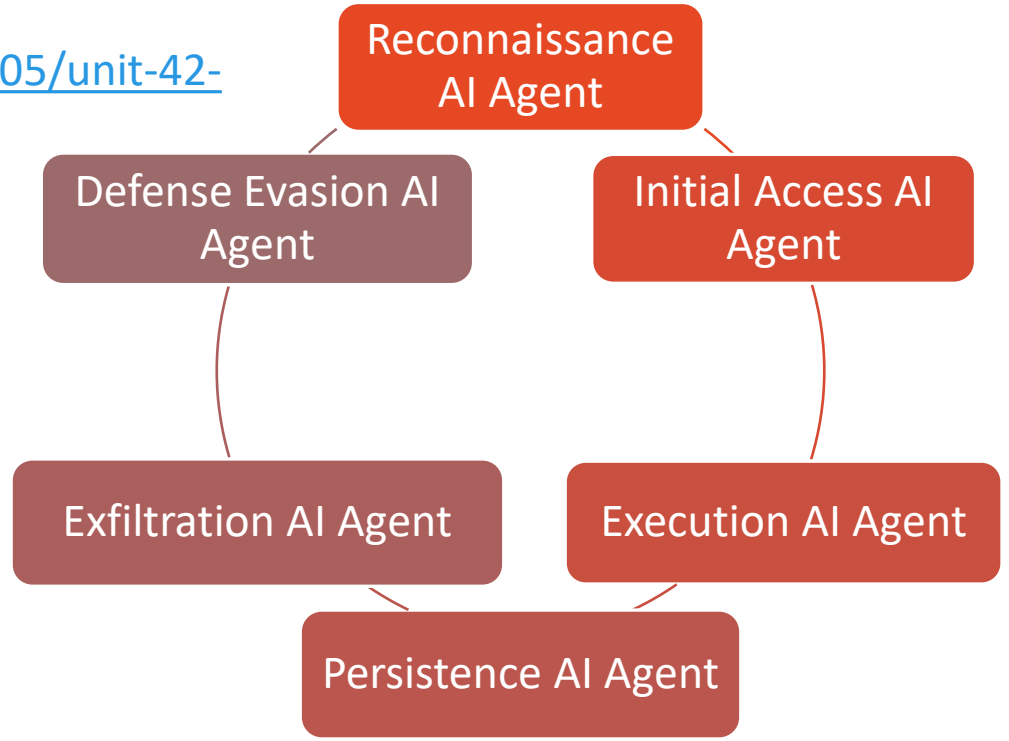


Risks of Agentic AI Adversarial Use

Adversarial Use of Agentic AI

The **Unit 42 Attack Framework** is a comprehensive model developed by Palo Alto Networks' Unit 42 threat intelligence team. It outlines how adversaries can leverage **Agentic AI**

https://www.paloaltonetworks.com/blog/2025/05/unit-42-develops-agentic-ai-attack-framework/



TAPSI

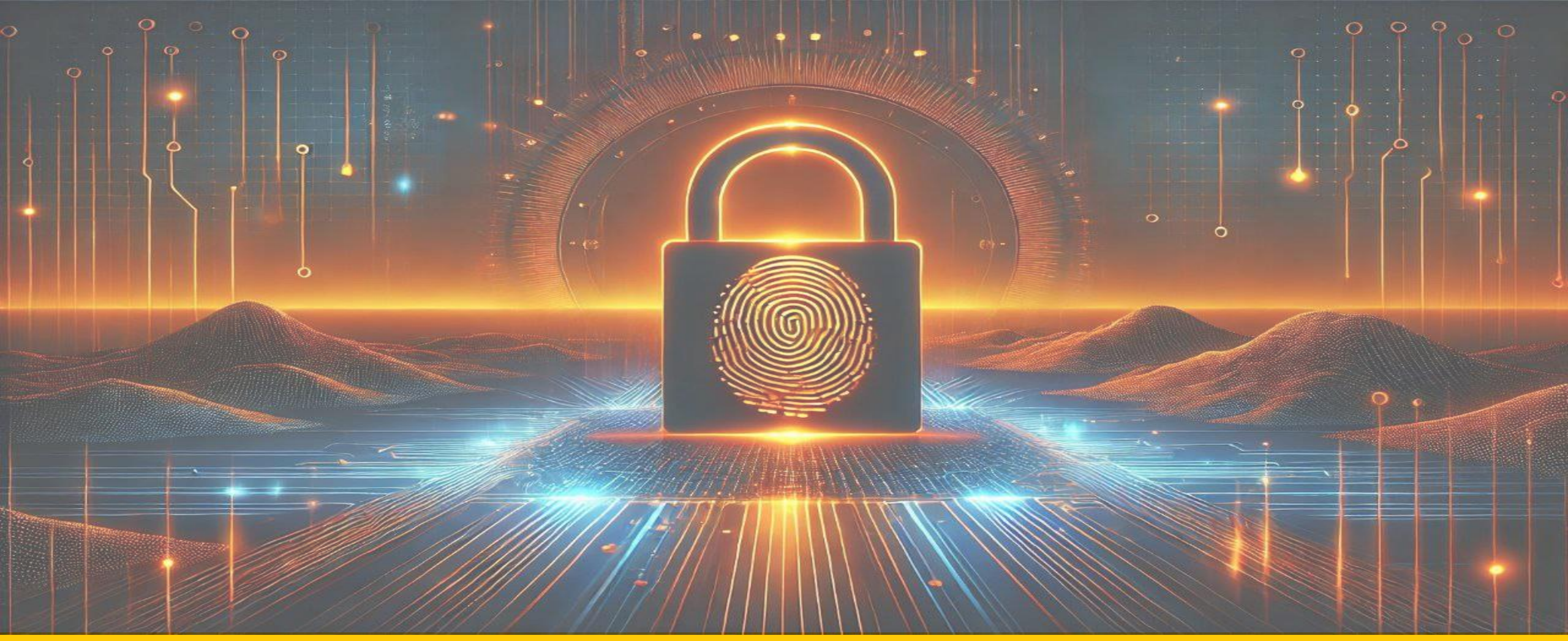
انجمن رمزایران
iranian Society of Cryptology

What Makes Agentic AI a Potent Adversarial Tool

- ✓**Autonomy:** Operates independently to achieve complex objectives without constant human oversight .
- ✓**Reasoning and Planning:** Uses LLMs to break down high-level goals (e.g., "steal financial data") into a sequence of actionable steps .
- ✓**Tool Use:** Can interact with and exploit IT environments through APIs, system calls, and other tools, just as a benign agent would .
- ✓**Adaptation:** Learns from its environment and the success/failure of its actions to improve its attack methods over time .

TAPSI

انجمن رمزایران
iranian Society of Cryptology



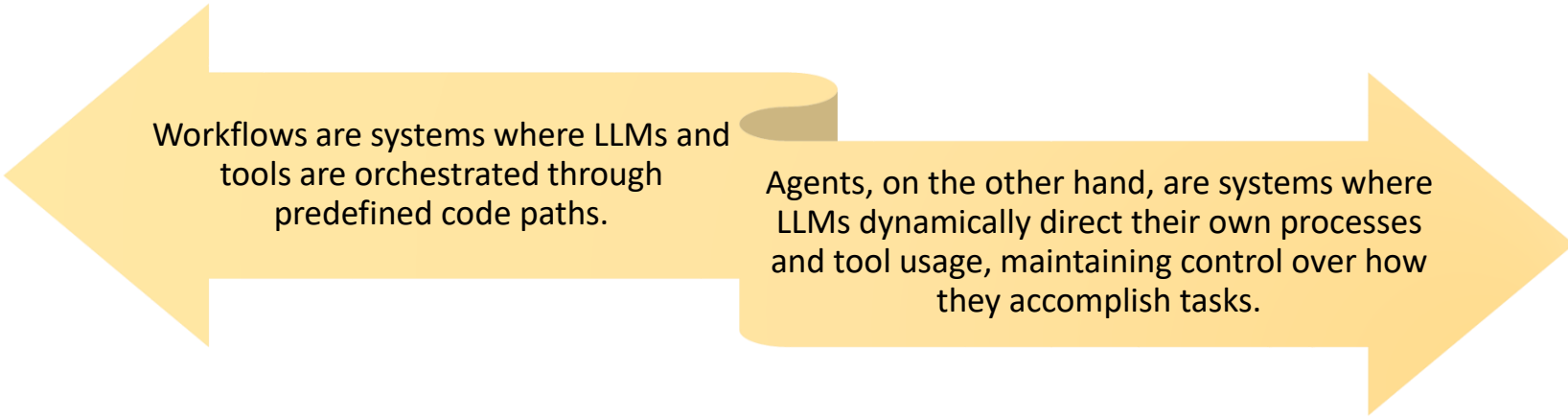
Designing AI Agent for Cyber Security

Designing AI Agent

What are agents?

"Agent" can be defined in several ways.

- ✓Some customers define agents as fully autonomous systems that operate independently over extended periods, using various tools to accomplish complex tasks.
- ✓Others use the term to describe more prescriptive implementations that follow predefined workflows.

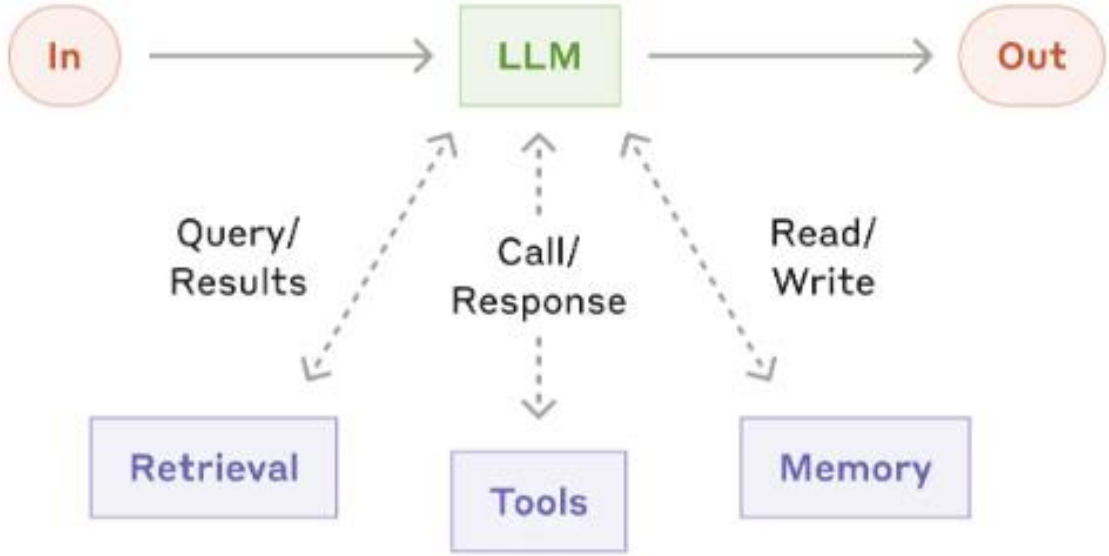


TAPSI

انجمن رمزایران
iranian Society of Cryptology

Designing AI Agent

Building block: The augmented LLM



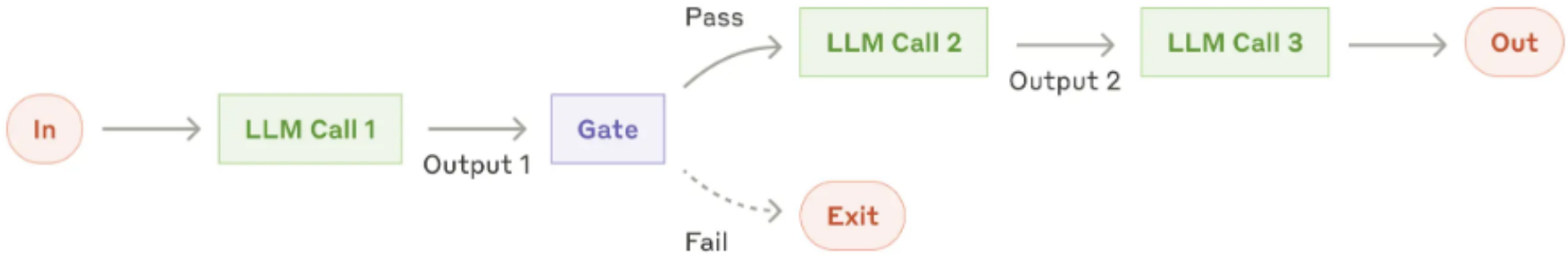
TAPSI

انجمن رمزایران
iranian Society of Cryptology

Designing AI Agent

Workflow: Prompt chaining

Prompt chaining decomposes a task into a sequence of steps, where each LLM call processes the output of the previous one.



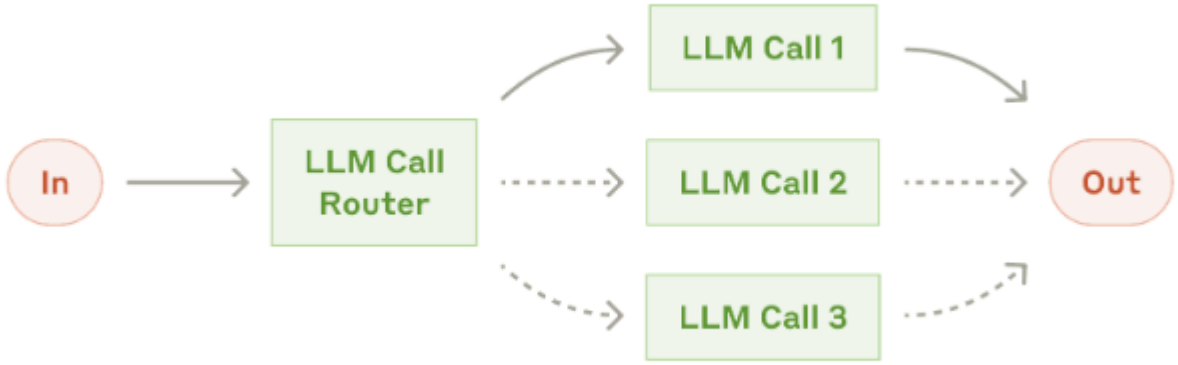
TAPSI

انجمن رمزایران
iranian Society of Cryptology

Designing AI Agent

Workflow: Routing

Routing classifies an input and directs it to a specialized followup task. This workflow allows for separation of concerns, and building more specialized prompts.



TAPSI

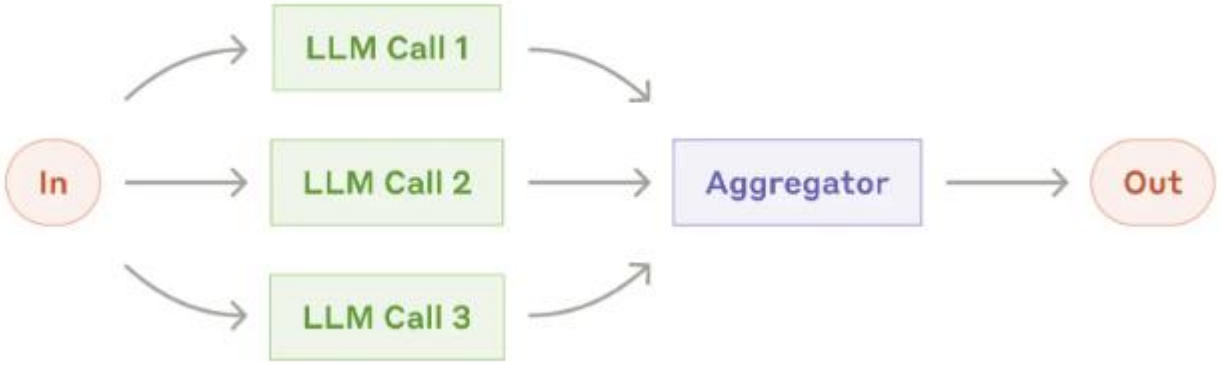
انجمن رمزایران
iranian Society of Cryptology

Designing AI Agent

Workflow: Parallelization

LLMs can sometimes work simultaneously on a task and have their outputs aggregated programmatically. It has two key variations:

- ✓Sectioning: Breaking a task into independent subtasks run in parallel.
- ✓Voting: Running the same task multiple times to get diverse outputs.



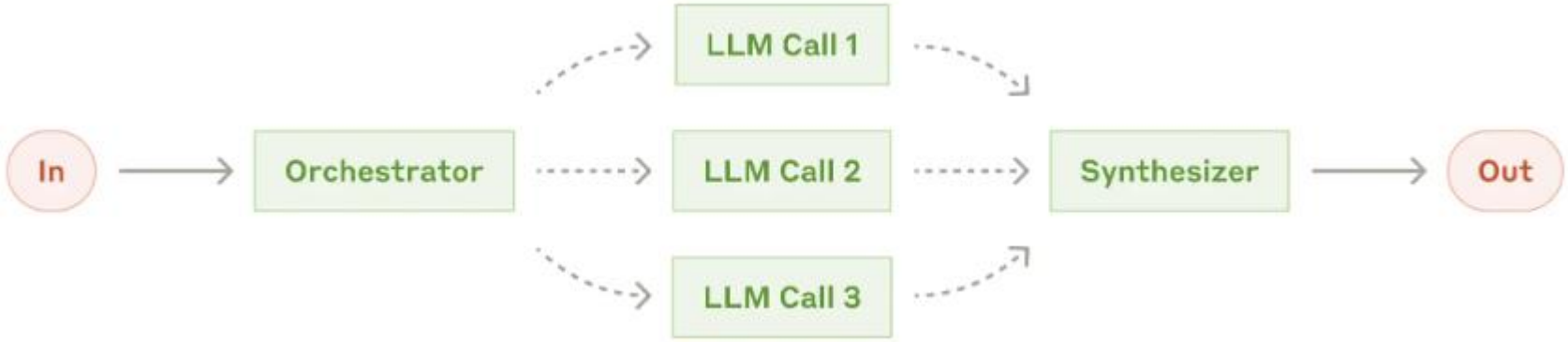
TAPSI

انجمن رمزایران
iranian Society of Cryptology

Designing AI Agent

Workflow: Orchestrator-workers

In the orchestrator-workers workflow, a central LLM dynamically breaks down tasks, delegates them to worker LLMs, and synthesizes their results.



TAPSI

انجمن رمزایران
iranian Society of Cryptology

Designing AI Agent

Workflow: Evaluator-optimizer

In the evaluator-optimizer workflow, one LLM call generates a response while another provides evaluation and feedback in a loop.

TAPSI

انجمن رمزایران
iranian Society of Cryptology

Key Terminology: The Language of Agents

Tools: Functions or external resources (APIs, databases) that an agent uses to interact with the world .

Memory: The ability to store and recall past interactions. This can be short-term (for the current task) or long-term (across sessions) .

Reasoning & Planning: The process of breaking down a high-level goal into a sequence of actionable steps .

Autonomy: Ability to operate without constant human control.

Environment: The system or data sources the agent interacts with.

TAPSI

انجمن رمزایران
iranian Society of Cryptology

The Framework Landscape: Building Agentic Systems

Framework	Primary Strength	Key Feature
LangGraph	Complex, stateful workflows	Graph-based architecture for fine-grained control
CrewAI	Production-grade multi-agent teams	High-level, role-based abstraction
AutoGen	Research & flexible prototyping	Free-form message passing between agents
OpenAI Agent SDK	Lightweight, single-agent tasks	Routine-based execution using docstrings

TAPSI

انجمن رمزایران
iranian Society of Cryptology

The Framework Landscape: Building Agentic Systems

- ✓n8n is an open-source workflow automation platform.
- ✓Enables low-code/no-code integration of APIs, databases, AI models, and services.
- ✓Think of it as an automation backbone for connecting tools and orchestrating tasks.

https://n8n.io/workflows/categories/secops/

TAPSI

انجمن رمزایران
iranian Society of Cryptology

Introduction to LangChain

- What it is:

✓A framework to build AI applications by chaining together reasoning, tools, and memory.

- Why it matters:

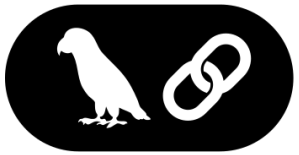
Allows cybersecurity agents to:

✓Query threat intel databases.

✓Automate incident response workflows.

✓Generate reports and recommendations.

- Features: Tool integration, memory, reasoning chains.



LangChain

TAPSI

انجمن رمزایران
Iranian Society of Cryptology

Introduction to LangGraph

- What it is:
  - ✓A framework for designing stateful, graph-based AI workflows.
- Why it matters:
  - ✓Supports multi-agent collaboration (SOC-like teams of AI agents).
- Applications:
  - ✓Threat hunting with distributed agents.
  - ✓Attack simulation with adversary vs. defender agents.
- Strength: Clear control over agent states and handoffs.

TAPSI

انجمن رمزایران
iranian Society of Cryptology

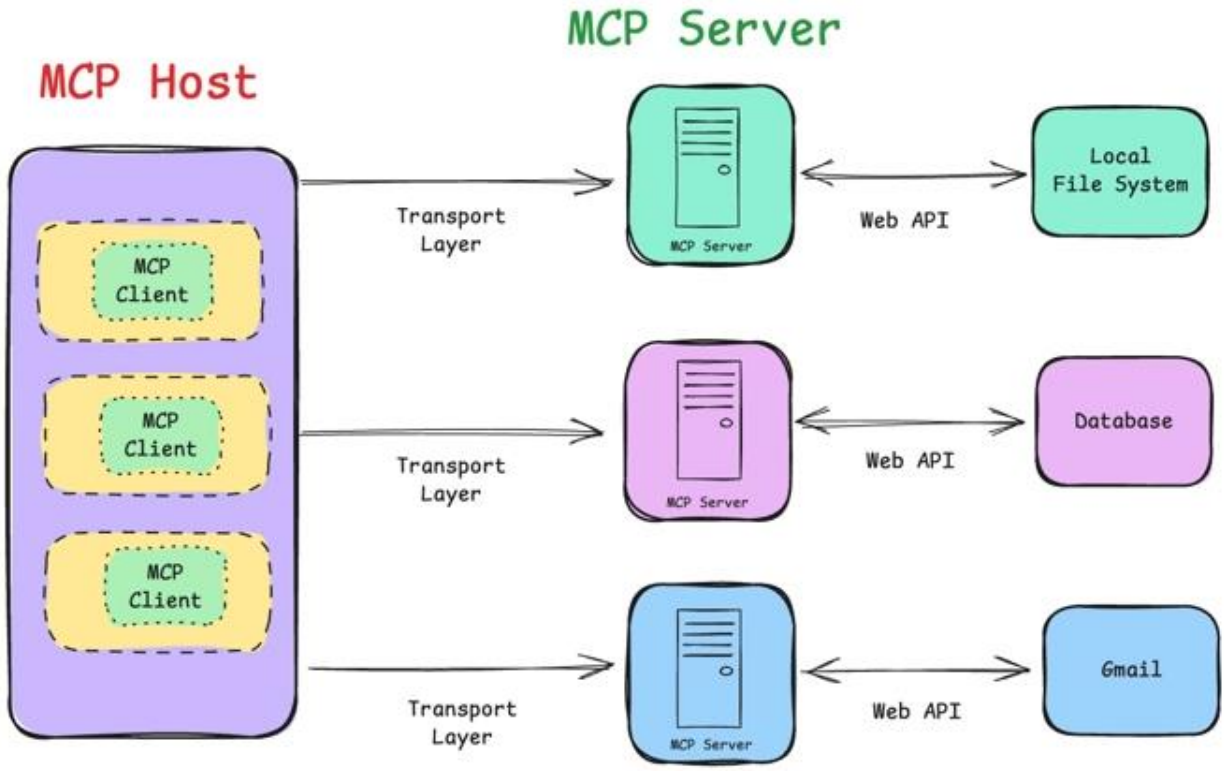
Model Context Protocol (MCP)

- What it is:
  - ✓ An open standard for connecting AI agents to external tools, APIs, and data.
- Why it matters:
  - ✓ Ensures interoperability between security tools and AI agents.
  - ✓ Reduces vendor lock-in and improves modularity.
- Cybersecurity Use Case:
  - ✓ An agent can use MCP to query a SIEM, fetch vulnerability data, and trigger a response action.

TAPSI

انجمن رمزایران
iranian Society of Cryptology

Model Context Protocol (MCP)



TAPSI

انجمن رمزایران
Iranian Society of Cryptology

Building effective agents

Agentic AI frameworks make it easy to get started by simplifying standard low-level tasks

- ✓ Calling LLMs
- ✓ Defining and parsing tools
- ✓ Chaining calls together

ANTHROP\C

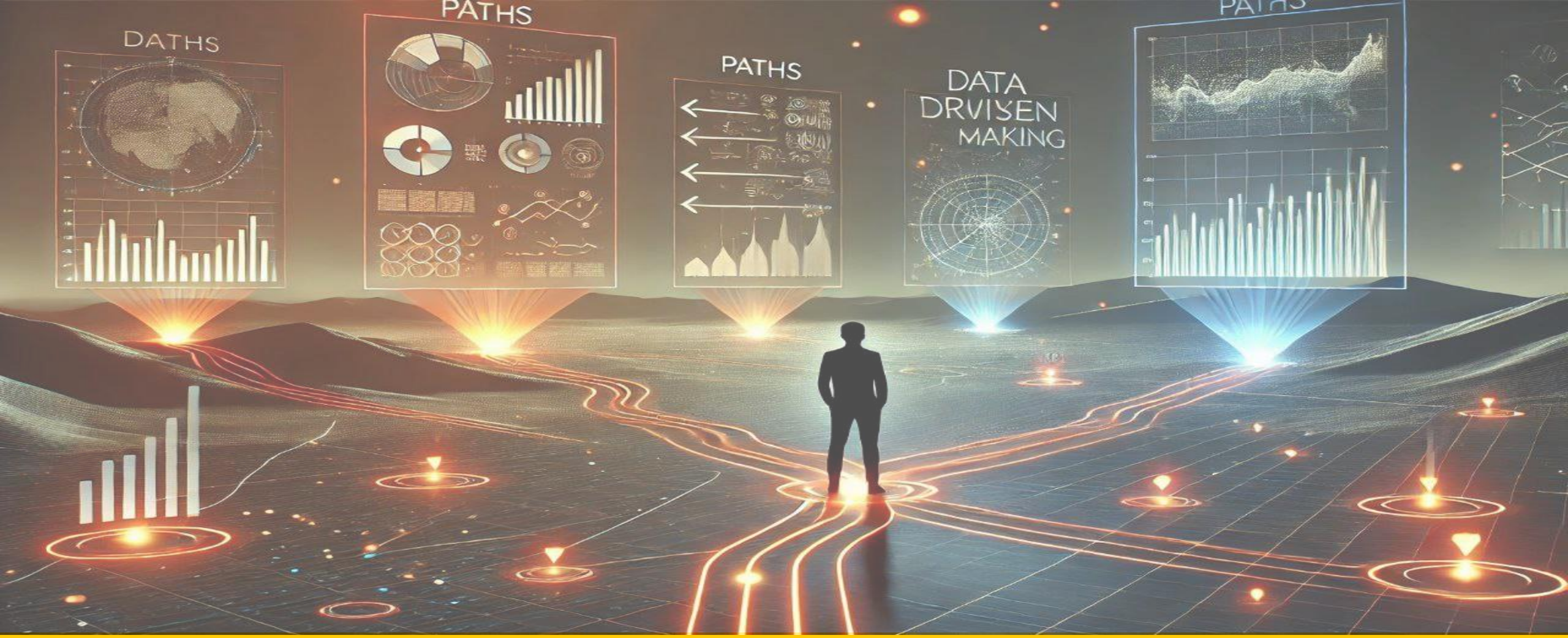
They often create extra layers of abstraction that can:

- ✓ Obscure the underlying prompts and responses
- ✓ Making them harder to debug

https://www.anthropic.com/engineering/building-effective-agents

TAPSI

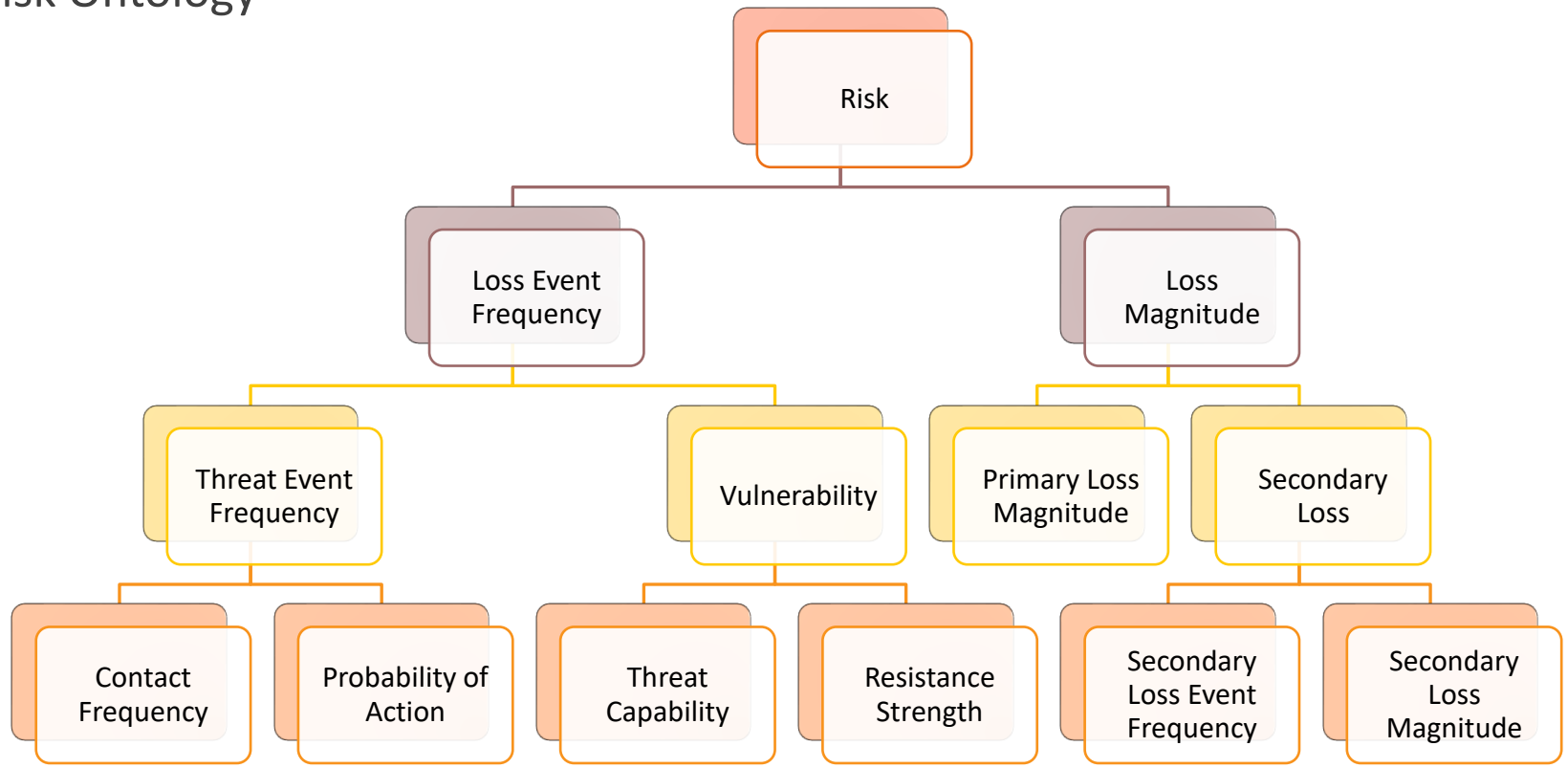
انجمن رمزایران
Iranian Society of Cryptology



Step by Step Implementation

Recap. FAIR Framework

The FAIR Risk Ontology



TAPSI

انجمن رمزایران
Iranian Society of Cryptology

Step by Step Implementation

Top Apps

Today

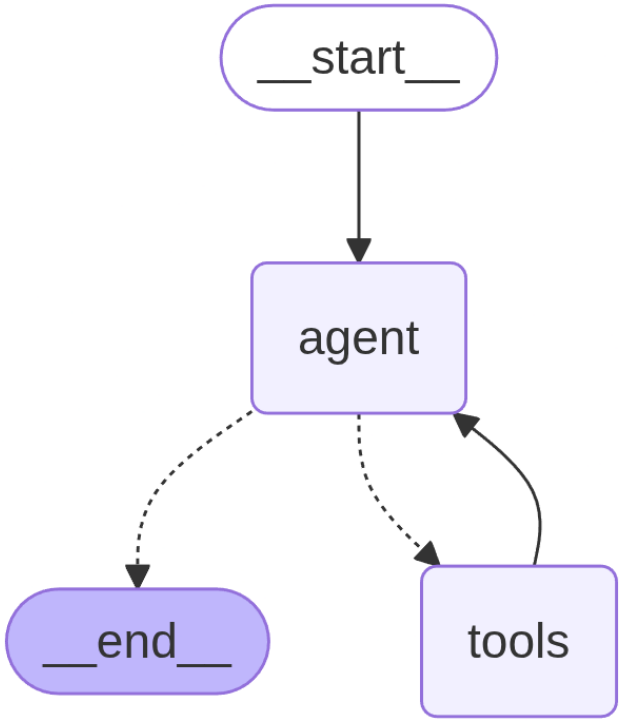
Largest public apps opting into usage tracking on OpenRouter

1.	Kilo Code AI coding agent for VS Code	99.2B tokens	11.	HammerAI Chat with AI characters for free	3.54B tokens
2.	Cline Autonomous coding agent right in...	64.6B tokens	12.	Sophias Lorebary Unofficial JanitorAI extension with...	3.15B tokens
3.	BLACKBOXAI AI agent for builders	52.7B tokens	13.	GR Geo Ranking System ... new	3.04B tokens
4.	janitorai.com new	44.2B tokens	14.	OpenRouter: Chatroom Chat with multiple LLMs at once	2.81B tokens
5.	liteLLM Open-source library to simplify L...	36.8B tokens	15.	Relace Evaluation Frame... new	2.59B tokens
6.	Roo Code A whole dev team of AI agents in ...	29.7B tokens	16.	OpenCode AI coding agent built for the termi...	2.54B tokens
7.	SillyTavern LLM frontend for power users	8.47B tokens	17.	New API LLM gateway, fork of One API	2.13B tokens
8.	Pax Historia An alternate history sandbox game	6.94B tokens	18.	New API new	2.04B tokens
9.	Chub AI GenAI for everyone	4.43B tokens	19.	GDevelop AI-powered game engine	1.63B tokens
10.	Zread.AI AI code wiki with multilingual guid...	4.02B tokens	20.	Portkey AI Control panel for AI apps	1.53B tokens

TAPSI

انجمن رمزایران
Iranian Society of Cryptology

Step by Step Implementation



TAPSI

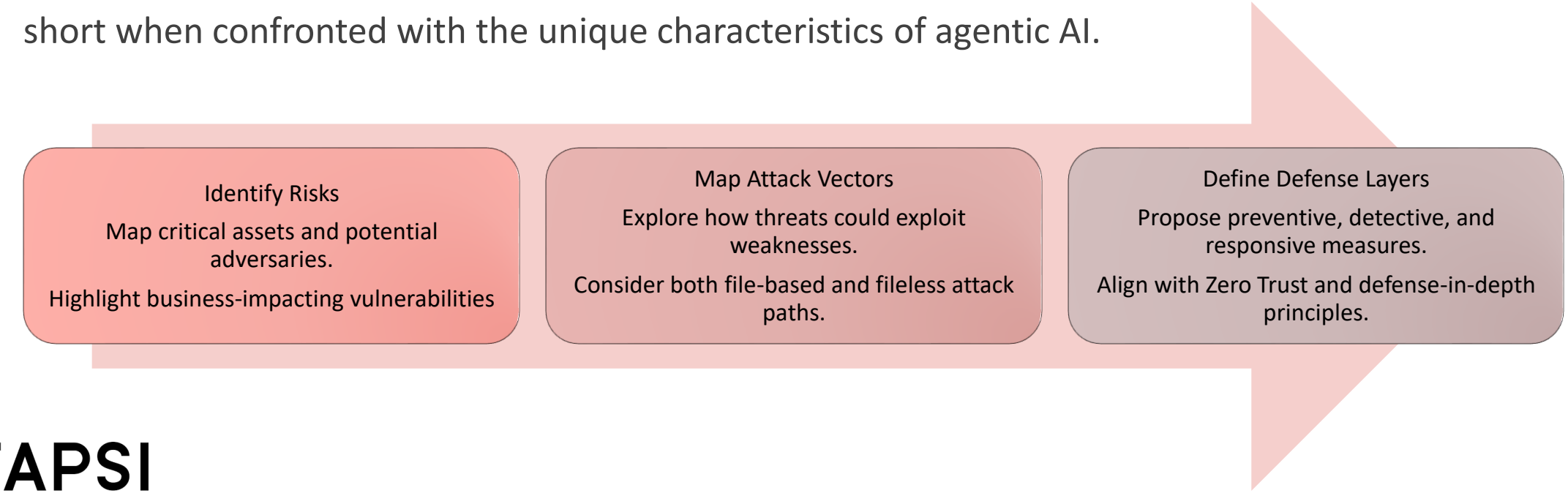
انجمن رمزایران
Iranian Society of Cryptology

Building Trustworthy Agents

Agentic AI Threat Modeling

MAESTRO (Multi-Agent Environment, Security, Threat, Risk, & Outcome)

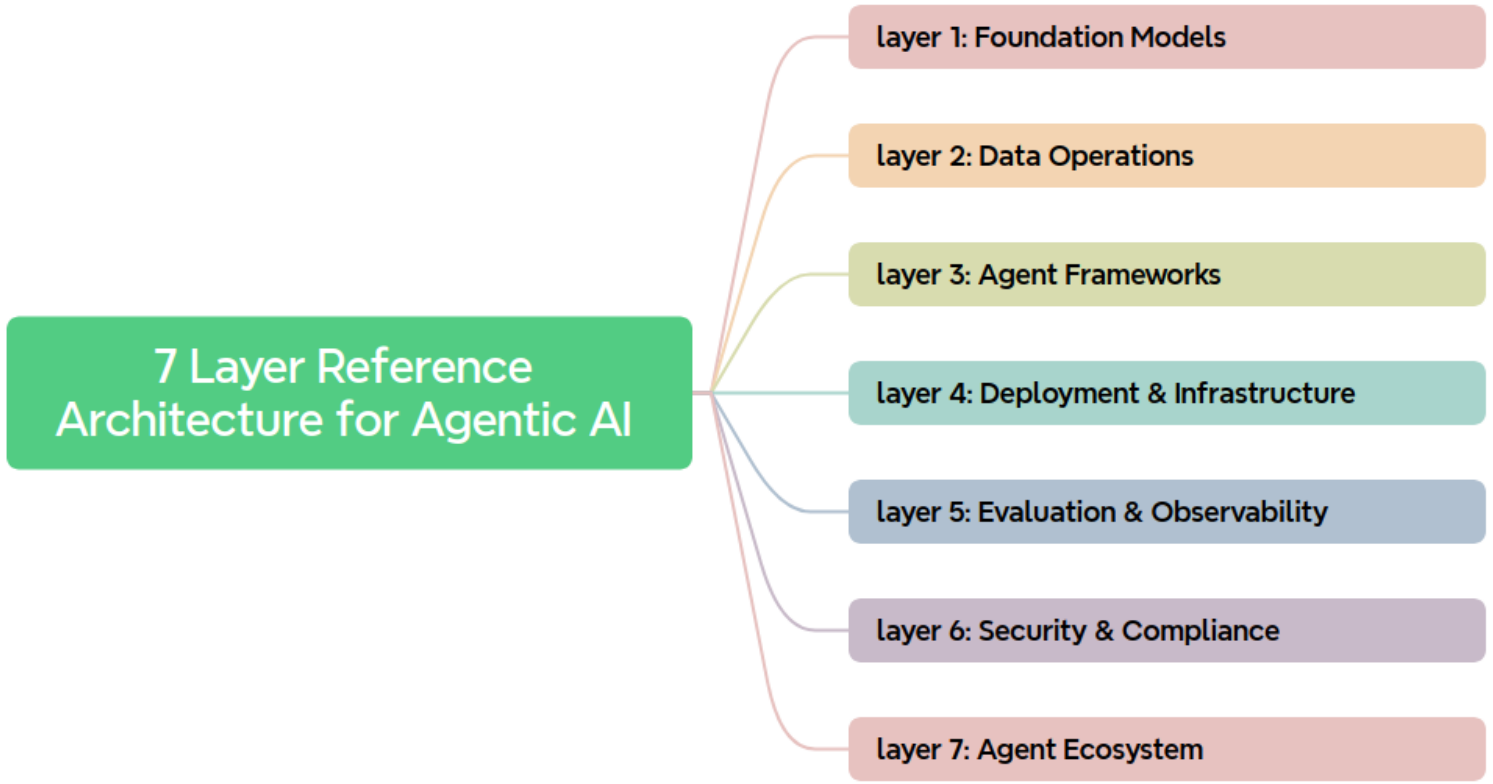
✓STRIDE, PASTA, LINDDUN, and OCTAVE — were built for predictable software logic. They fall short when confronted with the unique characteristics of agentic AI.



TAPSI

انجمن رمزایران
Iranian Society of Cryptology

Agentic AI Threat Modeling



https://medium.com/@oracle_43885/maestro-orchestrating-next-generation-security-for-the-agentic-ai-revolution-852a760606a5

Building Trustworthy AI Agents

Guardrails for Autonomy

✓**Action Restrictions:** Define boundaries on what agents can and cannot do.

✓**Human-in-the-Loop:** Require analyst approval for critical actions.

✓**Fail-Safes:** Automatic rollback or shutdown if unsafe behavior is detected.

TAPSI

انجمن رمزایران
iranian Society of Cryptology

Building Trustworthy AI Agents

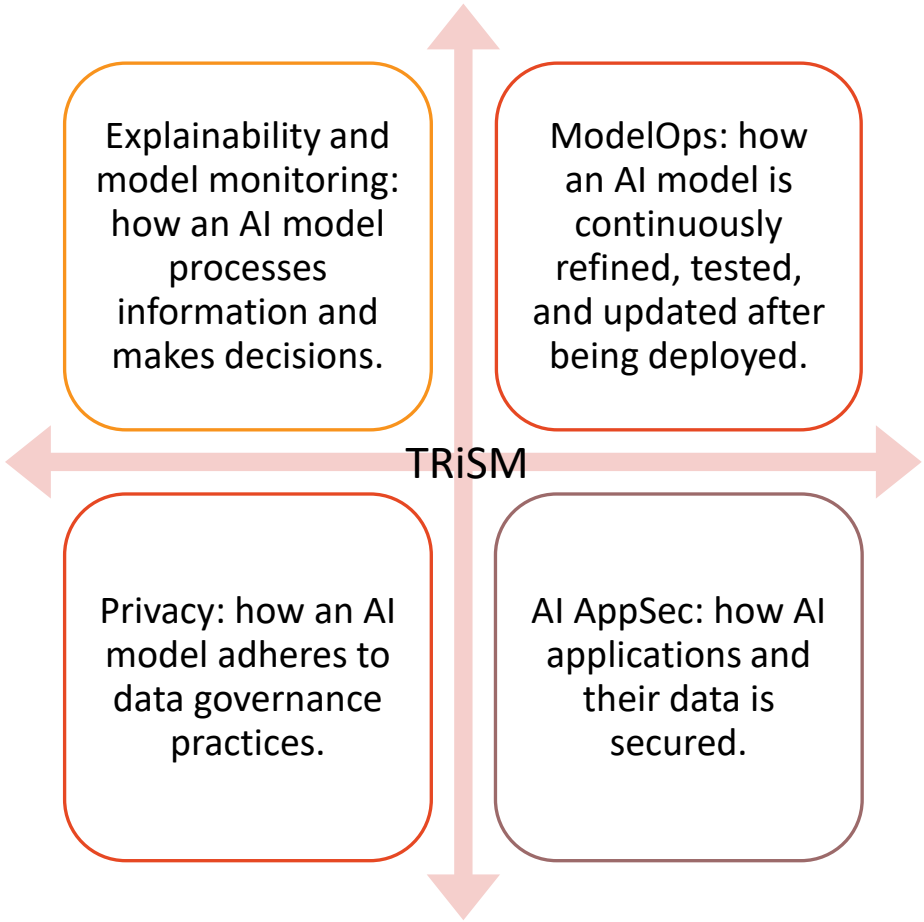
TRiSM Safeguards (Trust, Risk, and Security Management)

- ✓AI TRiSM is an acronym coined by Gartner
- ✓Refers to a framework for how organizations should identify and mitigate risks
- ✓Risks about: reliability, security, and trust within AI models and applications.

TAPSI

انجمن رمزایران
iranian Society of Cryptology

Building Trustworthy AI Agents



TAPSI

انجمن رمزایران
Iranian Society of Cryptology

“Now it’s your turn — what’s on your mind?”