Exploring the Paradigm Shift in Security

AI and Post-Quantum's Influence on Zero Trust

Saskia Laura Schröer, PhD
Security Consulting Engineer

CISCO Live !

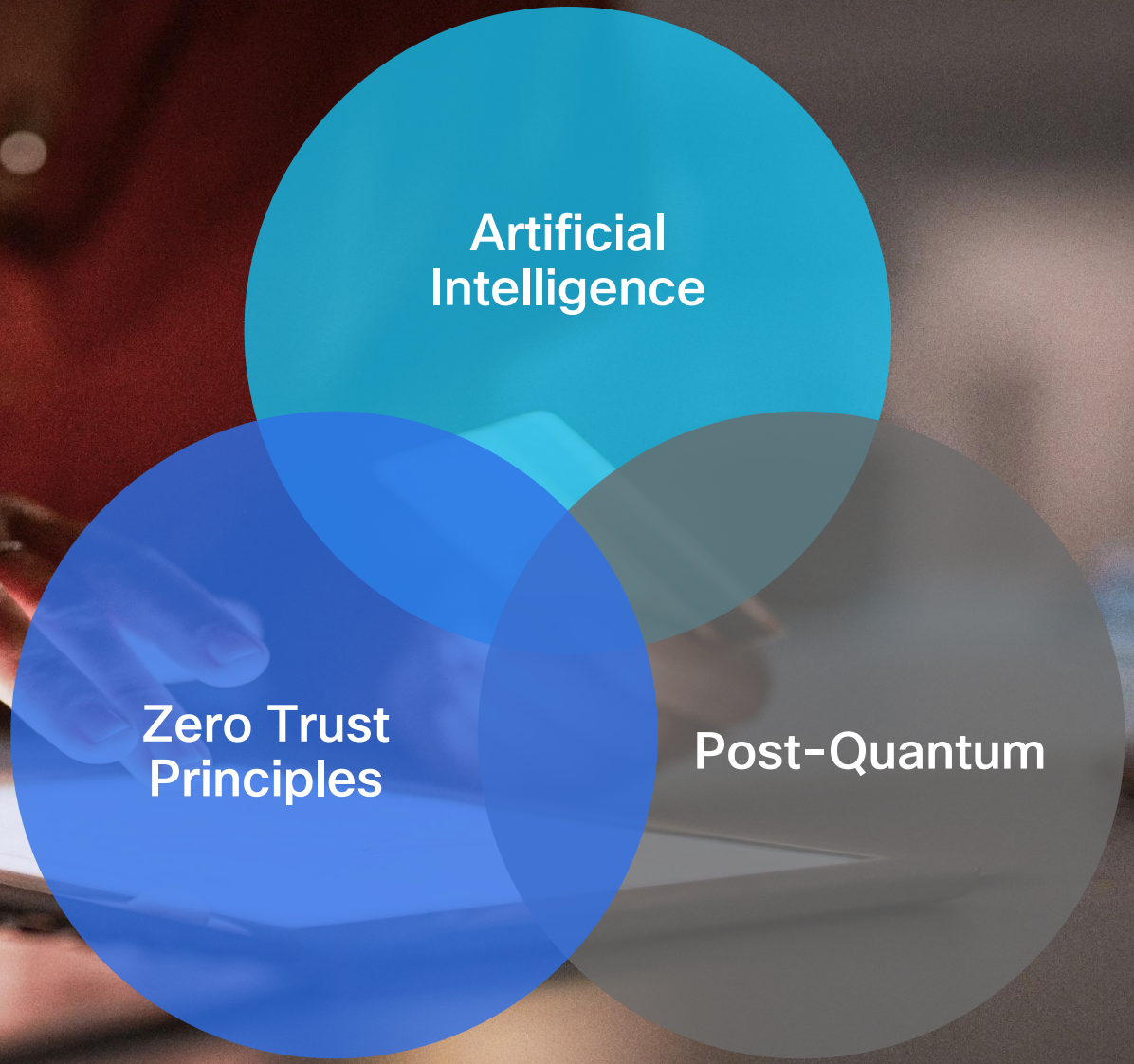
BRKXAR-1009

Exploring the Paradigm Shift in Security

This is not a Cisco-specific presentation.

We focus on industry standards and guidelines.

We are exploring the future.



Saskia Laura Schröer, PhD

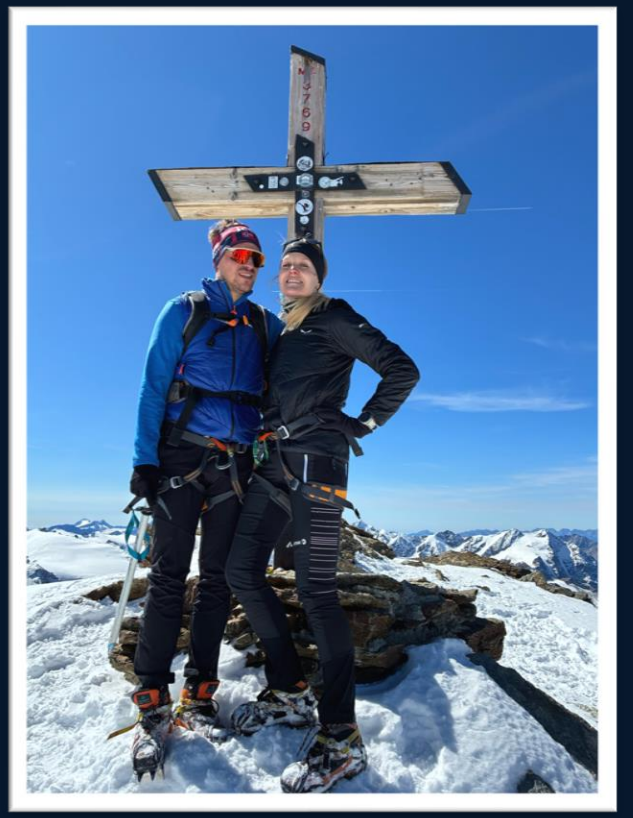
Senior Security Consulting Engineer | **CCNP Enterprise Networking** | Author

- Security Consulting Engineer in the CX EMEA Security Center of Excellence:
  - Zero Trust Architectures, Security Architectures, SOC Advisory
  - Agentic AI + AI Security

- PhD about the intersection of AI and Security:
  - Schröer, S. L., et al. (2025). SoK: On the offensive potential of AI. In *SaTML.* IEEE.
  - Schröer, S. L., et al. (2025). The dark side of the web: Towards understanding various data sources in cyber threat intelligence. In *EuroS&PW.* IEEE.
  - Schröer, S. L., et al. (2025). Using a Stack to Find an AI Needle: Topic Modeling for Cyber Threat Intelligence. Digital Threats: Research and Practice.
  - Schröer, S. L., et al. (2025). Exploiting AI for Attacks: On the Interplay between Adversarial AI and Offensive AI. IEEE Intelligent Systems.

- Reviewer of AI Security Conferences and Journals, e.g., Conference on Secure and Trustworthy Machine Learning (SaTML), or the Workshop on Artificial Intelligence and Security (AISec)

- Co-authoring Cisco Press Book "Securing AI using Zero Trust Principles" with Cindy Green-Ortiz and Zig Zsiga (~May 2026)



Agenda

01 Zero Trust Principles

02 Security Threat Landscape

03 Post-Quantum Cryptography

04 AI Security

05 Paradigm Shift: Actions to Take

06 Questions & Answers

Zero Trust Principles: The Foundation

It’s endpoint security.

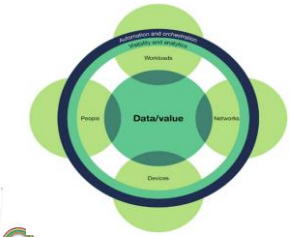
It’s identity.

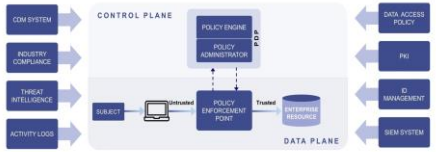
Zero Trust means different things to different people.

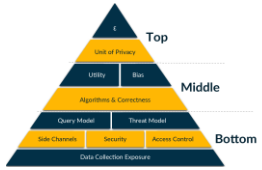
It’s segmentation.

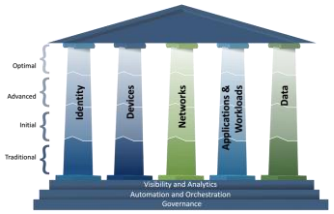
It’s Zero Trust Network Access.

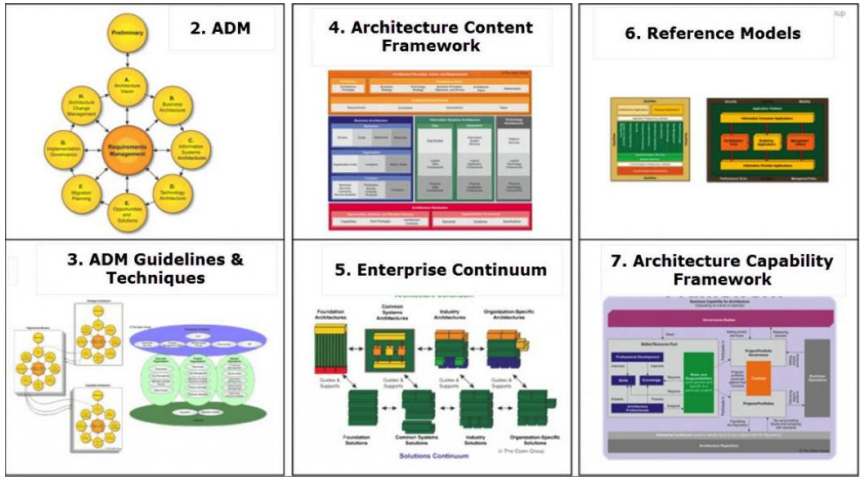
Zero Trust Foundational Approach

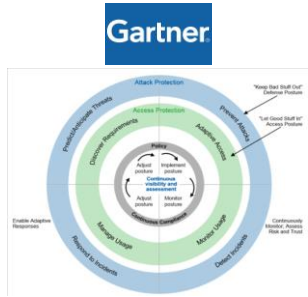


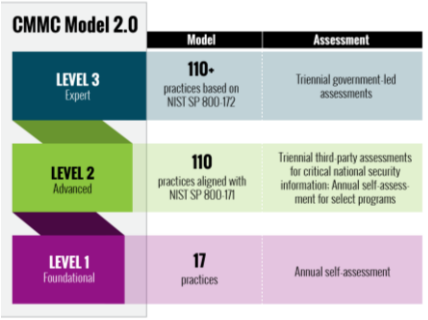












Zero Trust Capabilities

Policy & Governance

- Change Control
- Data Governance Policy + Encryption
- Data Retention Policy
- QoS
- Redundancy / Replication
- Business Continuity
- Disaster Recovery
- Risk Classification Policy
- Segmentation

Identity

- AAA
- Certificate Authority
- NAC
- Provisioning
- Privileged Access
- MFA
- Asset Identity
- Configuration (CMDB)
- IP Schemas

Vulnerability Management

- Endpoint Protection
- Malware Prevention and Inspection
- Vulnerability Management
- Authenticated Vulnerability Scanning
- Database Change

Enforcement

- CASB
- DDoS
- DLP
- DNS Security
- Email Security
- Firewall
- IPS
- Proxy
- VPN / RA
- SOAR
- File Integrity Monitor
- Segmentation

Analytics

- App. Performance Monitoring
- Audit, Logging, and Monitoring
- Change Detection
- Network Threat Behavior Analytics
- SIEM
- Threat Intelligence
- Traffic Visibility
- Asset Monitoring & Discovery

Source: Green-Ortiz, C., et al. (2023). Zero Trust Architecture. Cisco Press.

Typical Challenges – Focus Areas

Foundational Requirement: Leadership with Vision and Oversight

Security Threat Landscape

In the News

OpenClaw AI Runs Wild in Business Environments

The popular open source AI assistant (aka ClawdBot, MoltBot) has taken off, raising security concerns over its privileged, autonomous control within users' computers.

Source: https://www.darkreading.com/

Policy

Disrupting the first reported AI-orchestrated cyber espionage campaign

13. Nov. 2025

Source: https://www.anthropic.com/news

INNOVATION

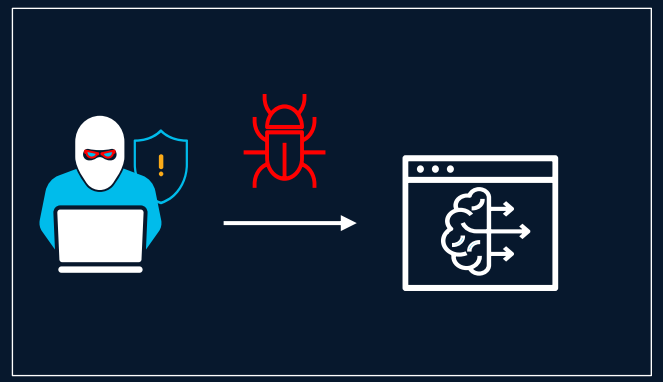
When Encryption Expires, Trust Is What's At Stake

By Rodrigo Madanes, Forbes Councils Member.
for Forbes Technology Council, COUNCIL POST | Membership (fee-based)

Published Dec 17, 2025, 10:15am EST

Source: https://www.forbes.com/

From Theory to Practice: AI-related Threats

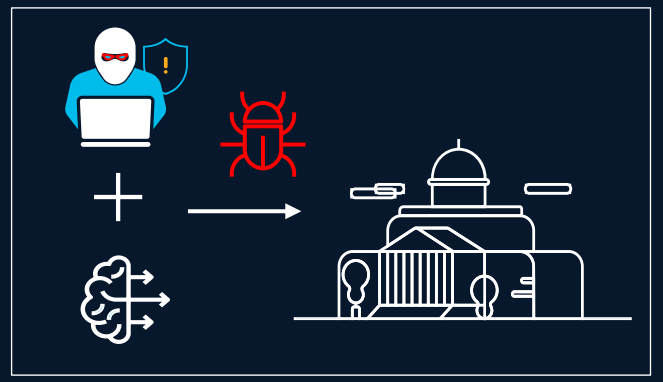


Adversarial ML/Security of AI:
Identifying vulnerabilities within AI models, and systems

First research evidence:
Evasion of Machine Learning driven Spam Filters (2004)

New Malware Spotted in The Wild Using Prompt Injection to Manipulate AI Models Processing Sample

Source: https://cybersecuritynews.com



Offensive AI: Using AI to drive attacks against different types of targets, e.g., organizations

First research evidence:
Using AI for CAPTCHA cracking (2008)

Disrupting the first reported AI-orchestrated cyber espionage campaign

13. Nov. 2025

Source: https://www.anthropic.com/news

From Theory to Practice: Quantum Computing

- **In 1994 the** Mathematician Shor published a quantum algorithm that solves the underlying mathematical problems of today's asymmetric key algorithms

- Significant threat to asymmetric key algorithms (RSA, Diffie-Hellman, ECC)

Harvest Now, Decrypt Later!

Algorithms for Quantum Computation:
Discrete Logarithms and Factoring

Peter W. Shor
AT&T Bell Labs
Room 2D-149
600 Mountain Ave.
Murray Hill, NJ 07974, USA

Abstract

A computer is generally considered to be a universal computational device; i.e., it is believed able to simulate any physical computational device with a cost in computation time of at most a polynomial factor. It is not clear whether this is still true when quantum mechanics is taken into consideration. Several researchers, starting with David Deutsch, have developed models for quantum mechanical computers and have investigated their computational properties. This paper gives Las Vegas algorithms for finding discrete logarithms and factoring integers on a quantum computer that take a number of steps which is polynomial in the input size, e.g., the number of digits of the integer to be factored. These two problems are generally considered hard on a classical computer and have been used as the basis of several proposed cryptosystems. (We thus give the first examples of quantum cryptanalysis.)

1 Introduction

Since the discovery of quantum mechanics, people have found the behavior of the laws of probability in quantum mechanics counterintuitive. Because of this behavior, quantum mechanical phenomena behave quite differently than the phenomena of classical physics that we are used to. Feynman seems to have been the first to ask what effect this has on computation [13, 14]. He gave arguments as to why this behavior might make it intrinsically computationally expensive to simulate quantum mechanics on a classical (or von Neumann) computer. He also suggested the possibility of using a computer based on quantum mechanical principles to avoid this problem, thus implicitly asking the converse question: by using quantum mechanics in a computer can you compute more efficiently than on a classical computer. Other early work in the field of quantum mechanics and computing was done by Benioff [1, 2]. Although he did not ask whether quantum mechanics conferred extra power to computation, he did show that a Turing machine could be simulated by the reversible unitary evolution of a quantum process, which is a necessary prerequisite for quantum computation. Deutsch [9, 10] was the first to give an explicit model of quantum computation. He defined both quantum Turing machines and quantum circuits and investigated some of their properties.

The next part of this paper discusses how quantum computation relates to classical complexity classes. We will thus first give a brief intuitive discussion of complexity classes for those readers who do not have this background. There are generally two resources which limit the ability of computers to solve large problems: time and space (i.e., memory). The field of analysis of algorithms considers the asymptotic demands that algorithms make for these resources as a function of the problem size. Theoretical computer scientists generally classify algorithms as efficient when the number of steps of the algorithms grows as a polynomial in the size of the input. The class of problems which can be solved by efficient algorithms is known as P. This classification has several nice properties. For one thing, it does a reasonable job of reflecting the performance of algorithms in practice (although an algorithm whose running time is the tenth power of the input size, say, is not truly efficient). For another, this classification is nice theoretically, as different reasonable machine models produce the same class P. We will see this behavior reappear in quantum computation, where different models for quantum machines will vary in running times by no more than polynomial factors.

There are also other computational complexity classes discussed in this paper. One of these is PSPACE, which are those problems which can be solved with an amount of memory polynomial in the input size. Another important complexity class is NP, which intuitively is the class of exponential search problems. These are problems which may require the search of an exponential size space to find

Source: Shor, P. W. (1994). Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th symposium on foundations of computer science*. IEEE.

Post-Quantum Cryptography

What is Quantum Computing?

Quantum computing uses special units called **quantum bits (qubit)** that can be in multiple states at once. This allows quantum computers to process information much faster and solve complex problems that regular computers cannot handle.

Superposition (of qubits)

classical
0100110101

quantum
$$p_0 0000000000 + p_1 0000000001 + p_2 0000000010 \cdots + p_{2^N} 1111111111$$

Entanglement

We know the state of the system, not the individual pieces

Source: https://newsroom.ibm.com/media-quantum-innovation

Let's get the concepts right: Quantum...

Quantum Computer
Powerful computer, based on quantum mechanics, allowing parallel processing and super fast execution of certain problems.

Quantum Networking
A network that connects quantum computers securely, connecting multiple quantum processors for increased computational power and efficiency.

Post-Quantum Cryptography
Cryptographic algorithms designed to be secure against quantum computer attacks.

Quantum Cryptography

Quantum Key Distribution (QKD)
Uses quantum mechanics to securely exchange encryption keys between two or more elements.

Quantum Random Number Generation (QRNG)
QRNG is critical in quantum encryption, typically with QKD to ensure unpredictability of cryptographic keys.

People are making incremental efforts in developing a

Quantum Computer.

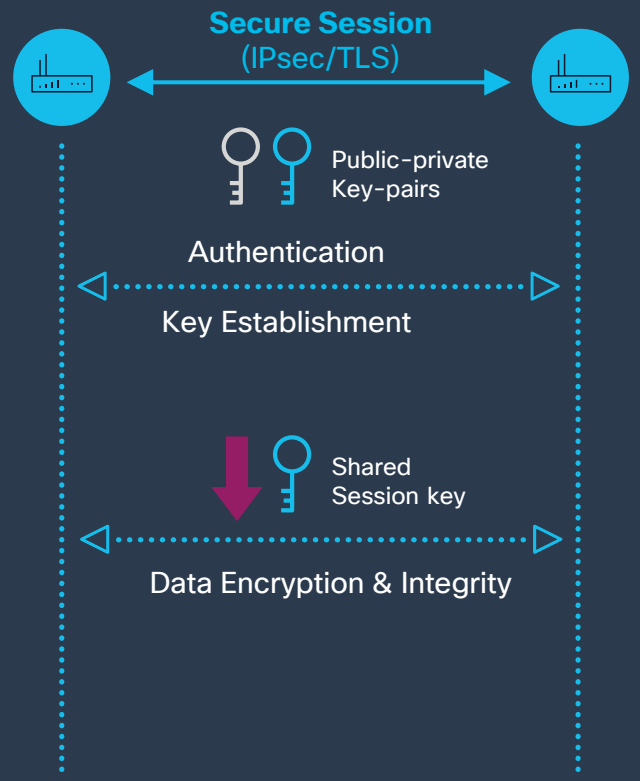
Once they have one which is sufficiently large and reliable, they may use it to

Break Current Encryption!

(public key algorithms)

cisco

Quantum Computing's Impact on Cryptography



Asymmetric Cryptography

- Based on **mathematically related** public-private key-pairs
- Used for control plane operations
  - Authentication, key establishment
- Example: RSA, DH, ECC

Symmetric Cryptography

- Based on a shared key
- Used for bulk data encryption & integrity
- Protection level based on key strength
  - Key size & entropy
- Example: AES

Quantum-Resistant?

Large reliable Quantum computers can break RSA, DH, and ECC!

Symmetric cryptography with large and high-entropy keys is resistant
to Quantum computer attacks.

NIST: Post-Quantum Cryptography

ML-KEM (FIPS 203)

- Based on CRYSTALS-Kyber
- Lattice-based
- Secures the exchange of keys over an untrusted medium

ML-KEM Use Cases

- Securing web connections
- VPN session key establishment

ML-DSA (FIPS 204)

- Based on CRYSTALS-Dilithium
- Lattice-based
- Digital signature scheme for authenticity and integrity of data

ML-DSA Use Cases

- Signing software and updates
- Secure boot
- Authenticating digital documents

SLH-DSA (FIPS 205)

- Based on SPHINCS+
- Stateless hash-based
- Digital signature scheme for authenticity and integrity of data

SLH-DSA Use Cases

- Ultimately the same as ML-DSA but uses a different cryptographic method

LMS/XMSS (NIST SP 800-208)

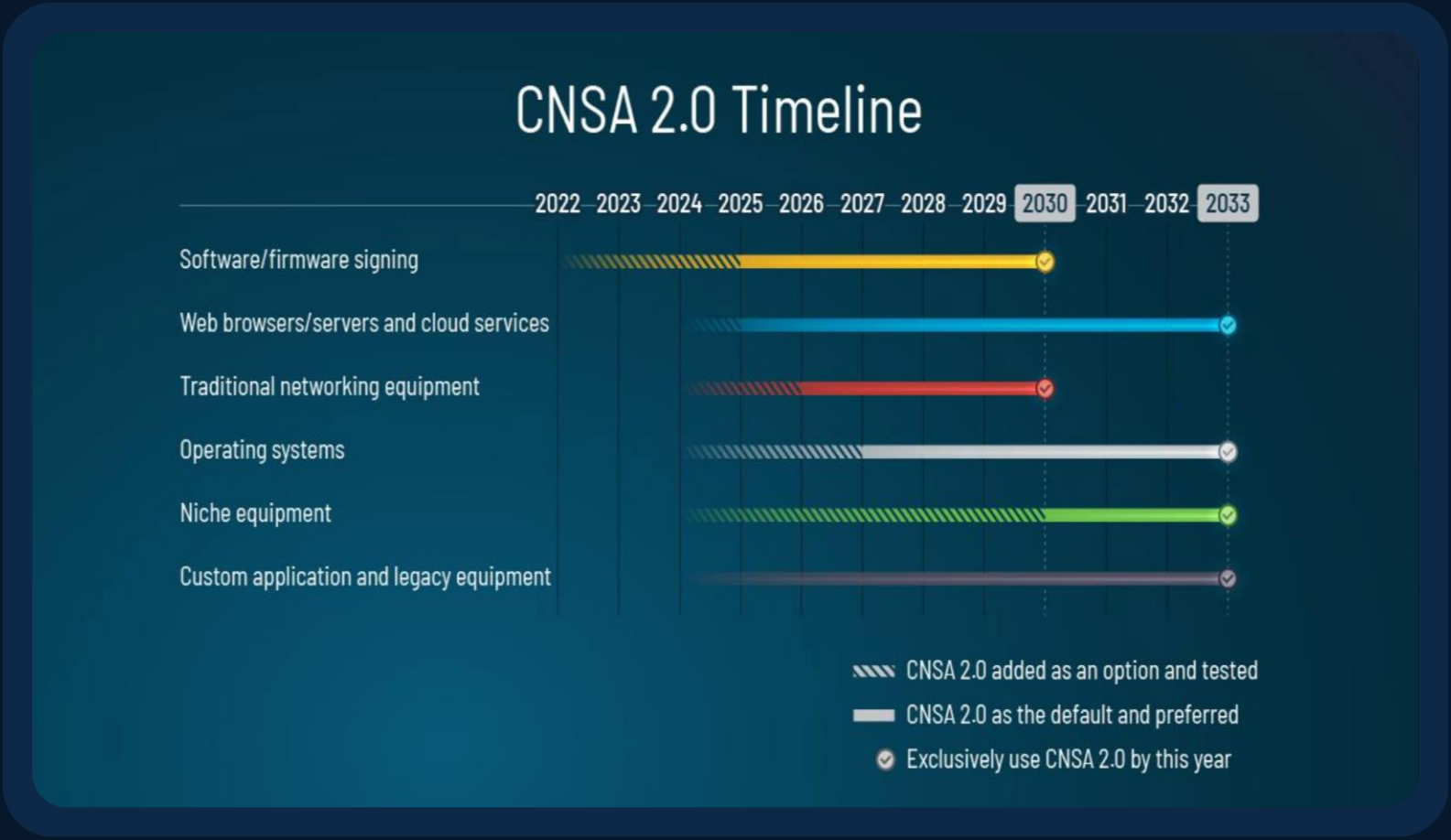
- Two digital signature schemes based on LMS (Leighton-Micali Signature) and XMSS (eXtended Merkle Signature Scheme)
- Stateful hash-based

LMS/XMSS (NIST SP 800-208)

- Asymmetric algorithm for digitally signing firmware and software

Source: https://csrc.nist.gov/Projects/post-quantum-cryptography/post-quantum-cryptography-standardization/selected-algorithms

Commercial National Security Algorithm Suite 2.0



CNSA FAQ update December 2024 version 2.1:

Required-by date accelerated to **January 2027.**

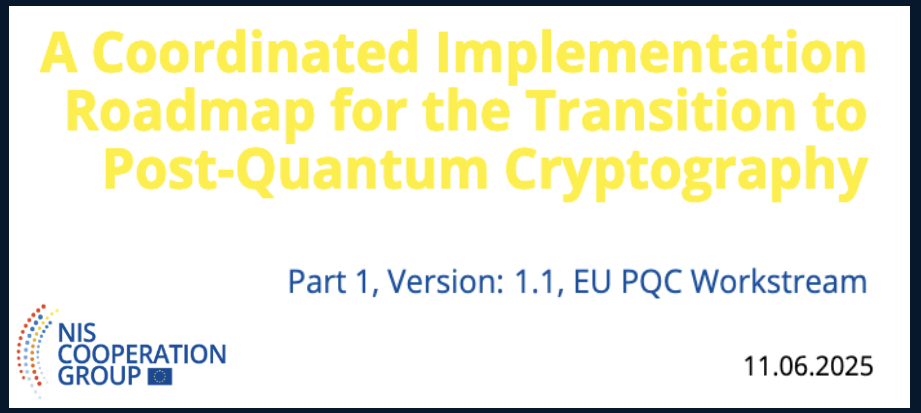
Only PQC allowed in NSS (National Security Systems) after **December 2031.**

Source: National Security Agency, Commercial National Security Algorithm Suite 2.0

European Perspective on Post-Quantum Cryptography

EU timeline:

- By 2026: Member states to outline national PQC transition roadmaps + start PQC transition planning, incl. pilots (e.g., cryptographic asset management, risk analysis)
- By 2030: PQC transition for high-risk use cases to be completed, support cryptographic agility
- By 2035: PQC transition for medium-risk use cases to be completed



NIS2 Directive

DORA

NIS2 demands the use of use “state-of-the-art” security and DORA to follow a flexible approach to address the dynamic landscape of cryptographic threats

More about the UK PQC Migration Timeline

Preparing for Quantum Computing Security Threats

Hybrid: Combination of traditional cryptographic and PQC algorithms.

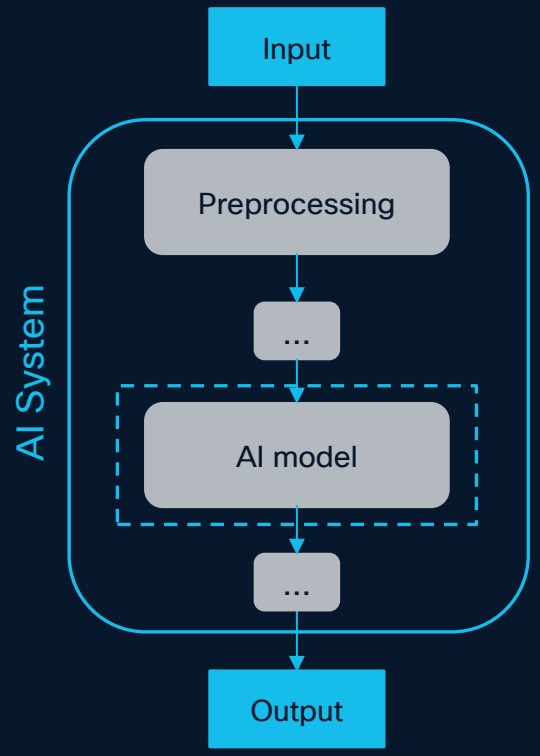
Prioritizing asset inventory (CBOM) and focusing on crypto agility aligns with Zero Trust to drive adaptability and growth!

Have you already started preparing for quantum-related threats?

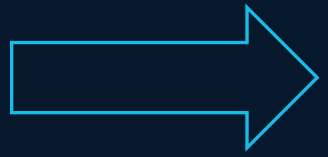
AI Security

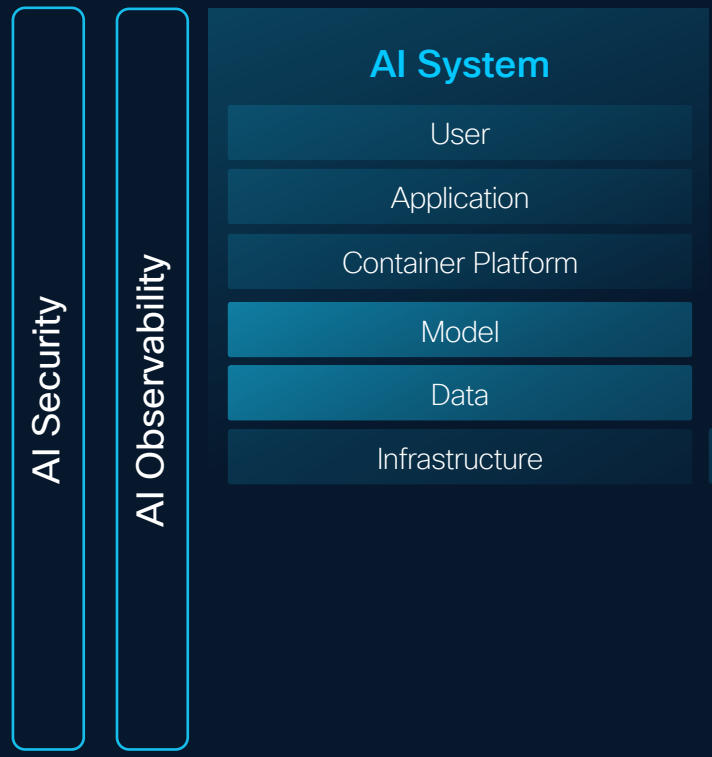
What is “new” about AI Security?
Let’s have a look at what AI is first...

From AI Models to AI Systems



AI security encompasses the entire AI technology stack.

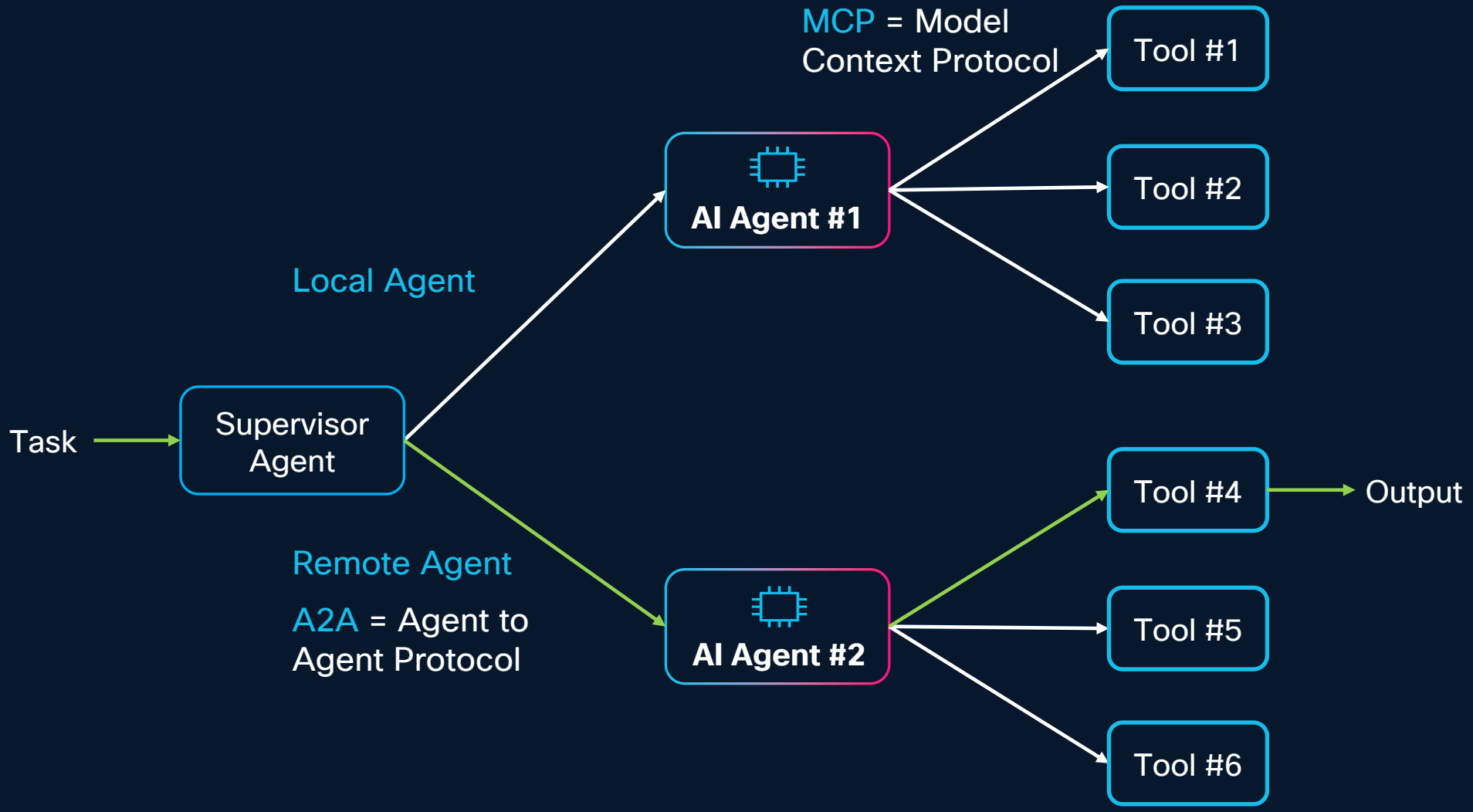




Predictive AI vs. Generative AI

Source: Apruzzese, G., et al. (2023). “Real attackers don't compute gradients”: bridging the gap between adversarial ml research and practice. In SaTML. IEEE.

AI Agents



At what point does AI stop being a competitive advantage and starts becoming a corporate liability because we bypassed security for the sake of speed?

Zero Trust Capabilities: An AI Perspective

Policy & Governance

- Change Control
- Data Governance Policy + Encryption
- Data Retention Policy
- QoS
- Redundancy / Replication
- Business Continuity
- Disaster Recovery
- Risk Classification Policy
- Segmentation

Identity

- AAA
- Certificate Authority
- NAC
- Provisioning
- Privileged Access
- MFA
- Asset Identity
- Configuration (CMDB)
- IP Schemas

Vulnerability Management

- Endpoint Protection
- Malware Prevention and Inspection
- Vulnerability Management
- Authenticated Vulnerability Scanning
- Database Change

Enforcement

- CASB
- DDoS
- DLP
- DNS Security
- Email Security
- Firewall
- IPS
- Proxy
- VPN / RA
- SOAR
- File Integrity Monitor
- Segmentation

Analytics

- App. Performance Monitoring
- Audit, Logging, and Monitoring
- Change Detection
- Network Threat Behavior Analytics
- SIEM
- Threat Intelligence
- Traffic Visibility
- Asset Monitoring & Discovery

Source: Green-Ortiz, C., et al. (2023). Zero Trust Architecture. Cisco Press.

Security Considerations for “AI”

🔒 Enterprise-Controlled AI (First-Party & Open-Source AI)

- **Definition:** AI systems developed internally or based on open-source frameworks, fully governed by the enterprise.
- **Examples:** In-house developed GenAI and ML models, AI built on open-source models, e.g., from Hugging Face.
- **Zero Trust Notes:**
  - Full control over model lifecycle and data handling.
  - Requires secure SDLC, supply chain risk management (AI BOMs), and internal audit trails.

🌐 Externally Managed AI (Third-Party & Cloud AI)

- **Definition:** AI services or solutions managed by external vendors or public cloud providers.
- **Examples:** SaaS AI tools, AWS Bedrock.
- **Zero Trust Notes:**
  - Shared responsibility for security and compliance (see Microsoft’s AI Shared Responsibility Model).
  - Needs vendor risk management, SLA enforcement, and runtime monitoring.

🤖 Agentic AI/AI Agents

- **Definition:** AI agents capable of initiating actions based on goals, context, and evolving logic.
- **Examples:** Langchain, N8N, self-operating AI agents.
- **Zero Trust Notes:**
  - Highest need for granular guardrails: role-based actions, behavioral monitoring, and real-time intent validation.

On what type of AI use cases are you working?

AI Governance: Regulations, Guidelines and Best Practices

EU Artificial Intelligence Act

ISO/IEC 23894:2023
Information technology — Artificial intelligence — Guidance on risk management

ISO/IEC 42001:2023
Information technology — Artificial intelligence — Management system

ISO/IEC 5338:2023
Information technology — Artificial intelligence — AI system life cycle processes

ISO/IEC 42005:2025
Information technology — Artificial intelligence (AI) — AI system impact assessment

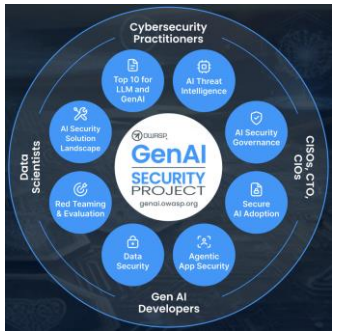
OWASP

OWASP GenAI SECURITY PROJECT — TOP 10 FOR LLM AND GENERATIVE AI

OWASP Top 10 For Agentic Applications 2026

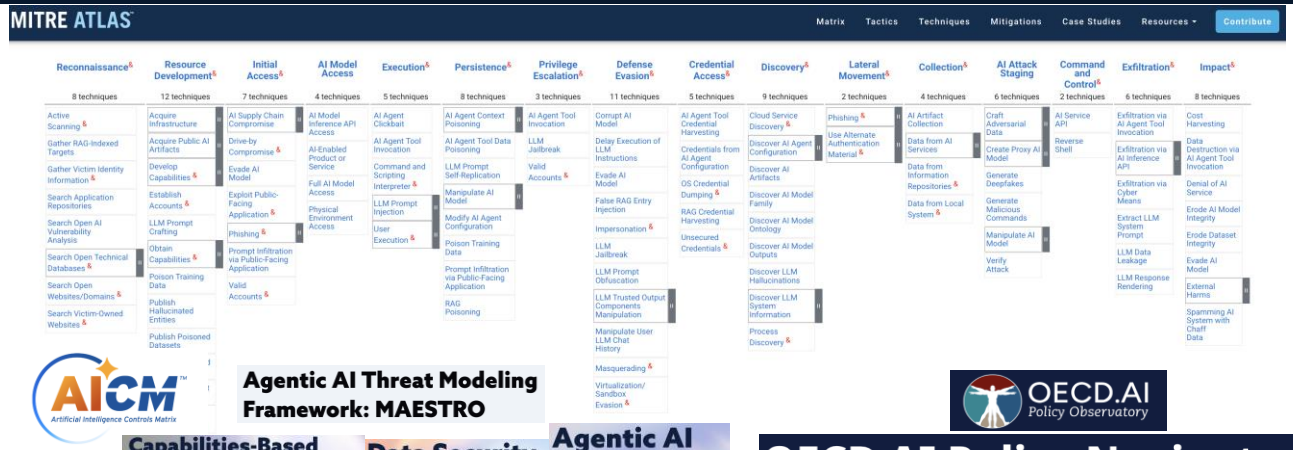
GenAI Incident Response Guide

GenAI Red Teaming Guide



Multi-Agentic system Threat Modelling Guide

OWASP AI Vulnerability Scoring System (AIVSS)

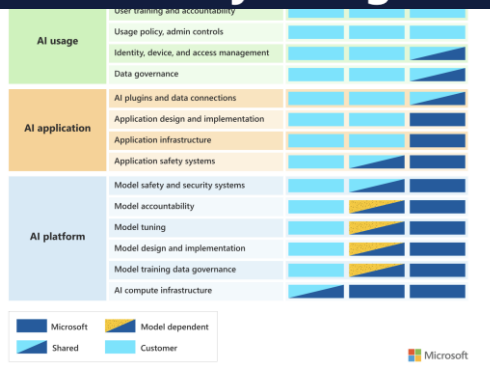


Agentic AI Threat Modeling Framework: MAESTRO

AICM — Artificial Intelligence Controls Matrix

OECD.AI Policy Observatory

OECD.AI Policy Navigator



NIST

NIST Special Publication 800
NIST SP 800-218A

Secure Software Development Practices for Generative AI and Dual-Use Foundation Models

NIST Special Publication 1270
Towards a Standard for Identifying and Managing Bias in Artificial Intelligence

cloud security alliance®

Capabilities-Based Risk Assessment (CBRA) for AI Systems

Data Security within AI Environments

Agentic AI Red Teaming

Secure Agentic System Design

Agentic AI Identity & Access Management

AI Organizational Responsibilities: AI Tools and Applications

Analyzing Log Data with AI Models to Meet Zero Trust Principles

Draft NIST Guidelines Rethink Cybersecurity for the AI Era

Artificial Intelligence Risk Management Framework (AI RMF 1.0)

COSAI — COALITION FOR SECURE AI

NIST Trustworthy and Responsible AI NIST AI 100-2e2025

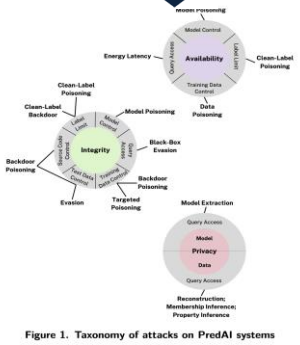


Figure 1. Taxonomy of attacks on PredAI systems

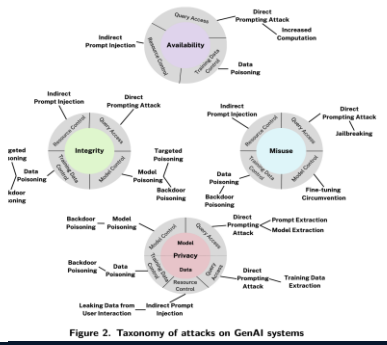


Figure 2. Taxonomy of attacks on GenAI systems

CISCO

Cisco's Principles for Responsible AI

Cisco Responsible AI Framework

Integrated AI Security and Safety Framework

Principles for the Secure Integration of Artificial Intelligence in Operational Technology

ETSI

Securing Artificial Intelligence (SAI); Baseline Cyber Security Requirements for AI Models and Systems

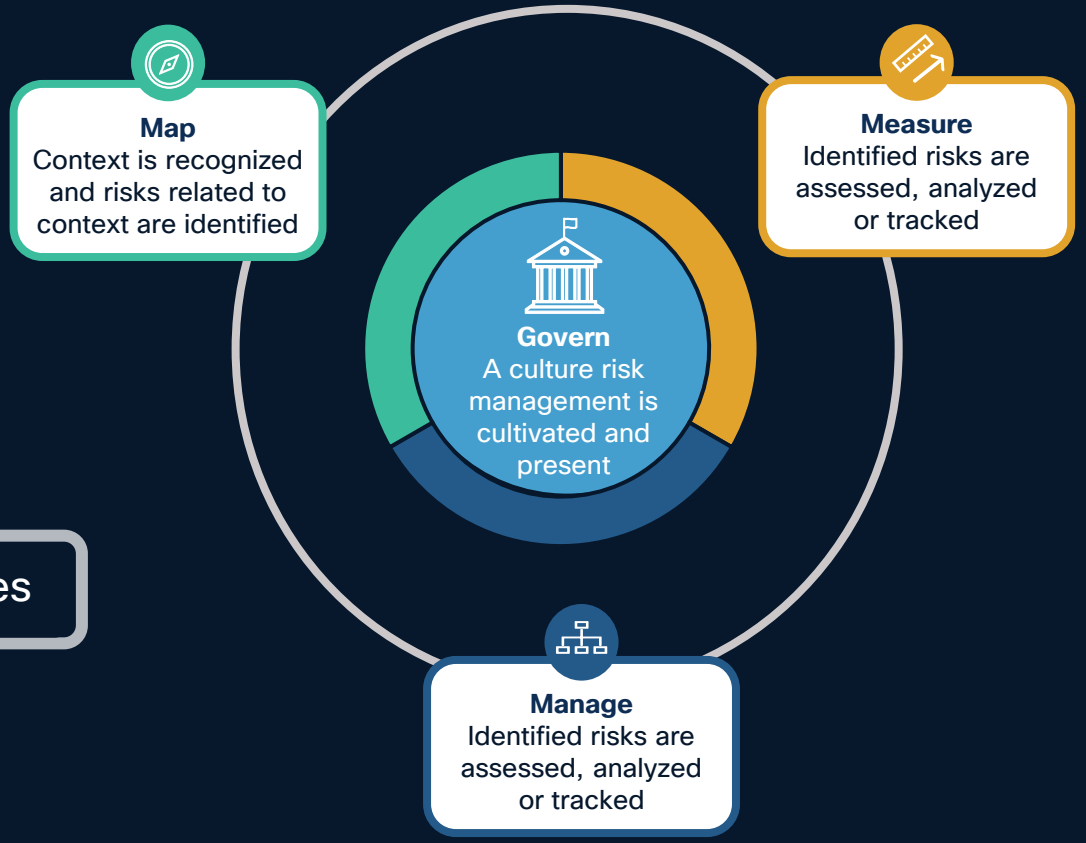
Federal Office for Information Security

Test Criteria Catalogue for AI Systems in Finance

AI Cloud Service Compliance Criteria Catalogue (AIC4)

NIST AI Risk Management Framework

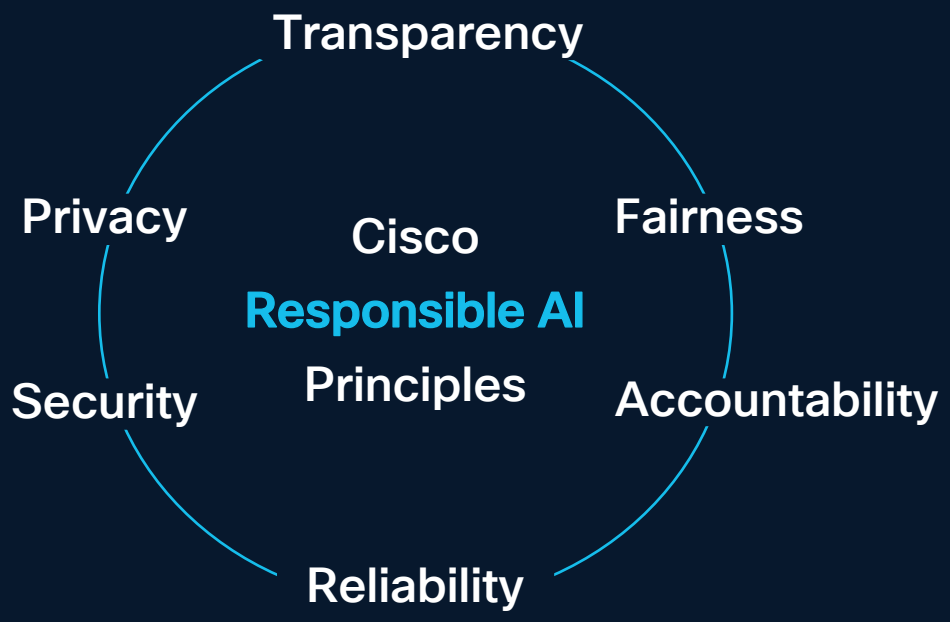
Mapping is focused on Generative AI and Agentic AI



Cisco's AI Best Practices

Expanded based on: Huang, K., et al. (2025). AAGATE: A NIST AI RMF-Aligned Governance Platform for Agentic AI. *arXiv preprint arXiv:2510.25863*.

Govern: Cisco's AI Best Practices



Enabling Safe and Trustworthy AI

Cisco **Responsible AI** Framework

Implementing Responsible AI Principles

For general guidance, please refer to the CSA AI Organizational Responsibilities.

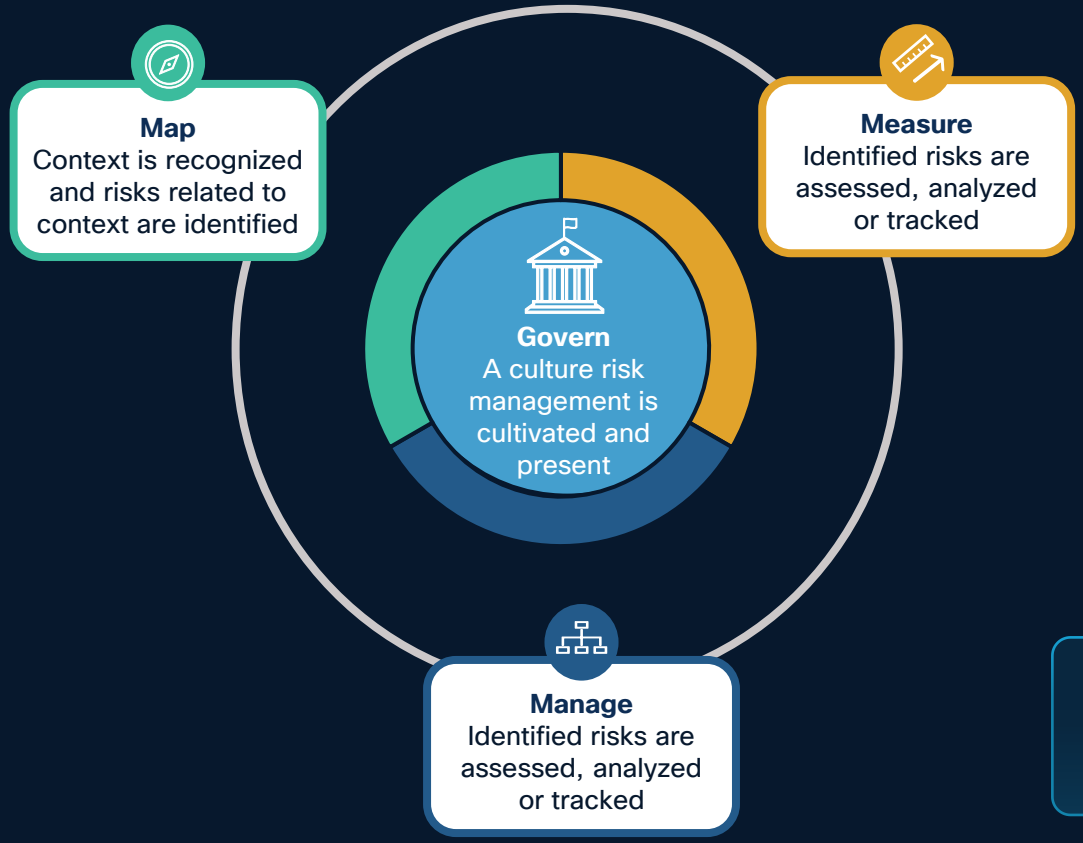
NIST AI Risk Management Framework

Mapping is focused on Generative AI and Agentic AI

MITRE ATLAS

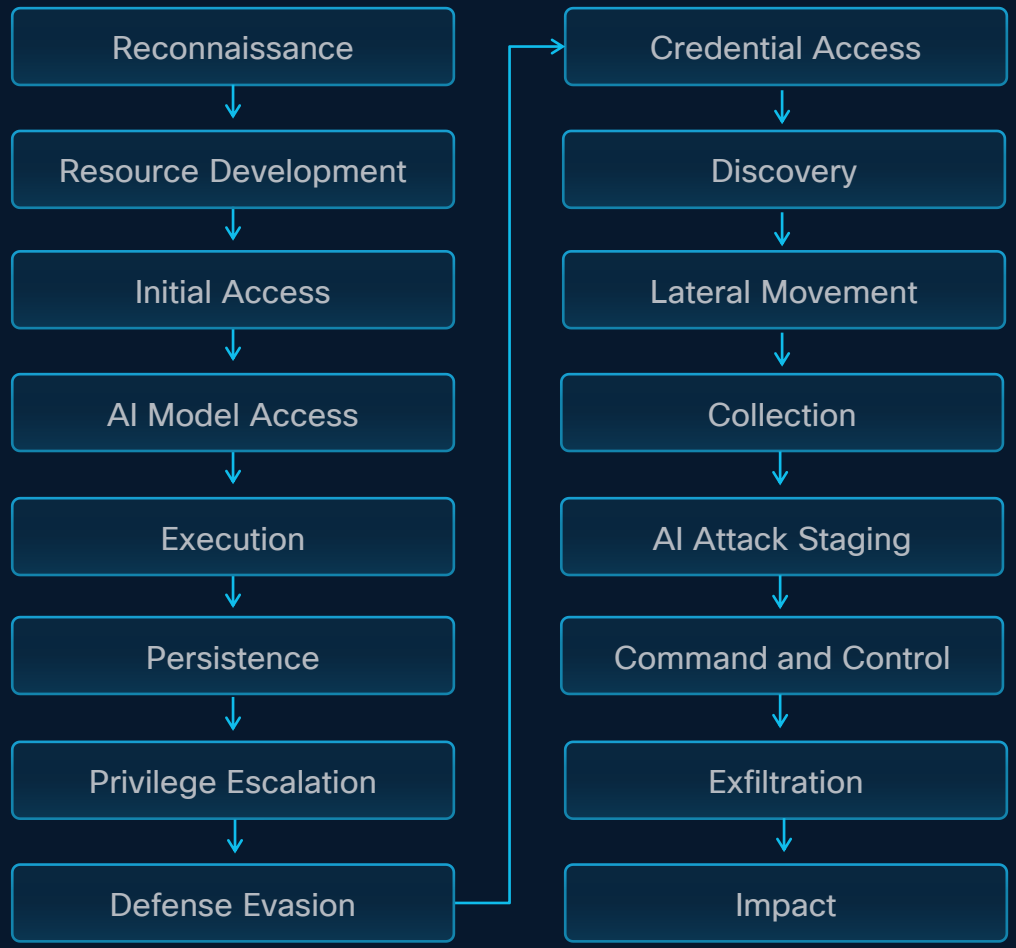
OWASP LLM/Agentic AI

MAESTRO Threat Modeling



Appendix B of NIST AI RMF:
"How AI Risks Differ from Traditional Software Risks"

Map: MITRE Adversarial Threat Landscape on AI Systems



MITRE ATLAS™

Filter by Maturity

Feasible | Demonstrated | Realized

cisco

Map: OWASP Top Ten for LLMs

OWASP

LLM01 Prompt Injection	LLM06 Excessive Agency
LLM02 Sensitive Information Disclosure	LLM07 System Prompt Leakage
LLM03 Supply Chain	LLM08 Vector and Embedding Weaknesses
LLM04 Data and Model Poisoning	LLM09 Misinformation
LLM05 Improper Output Handling	LLM10 Unbounded Consumption

Example: Mapping to Top Ten for Agentic AI

LLM01 Prompt Injection

LLM05 Improper Output Handling

ASI01 Agent Goal Hijack

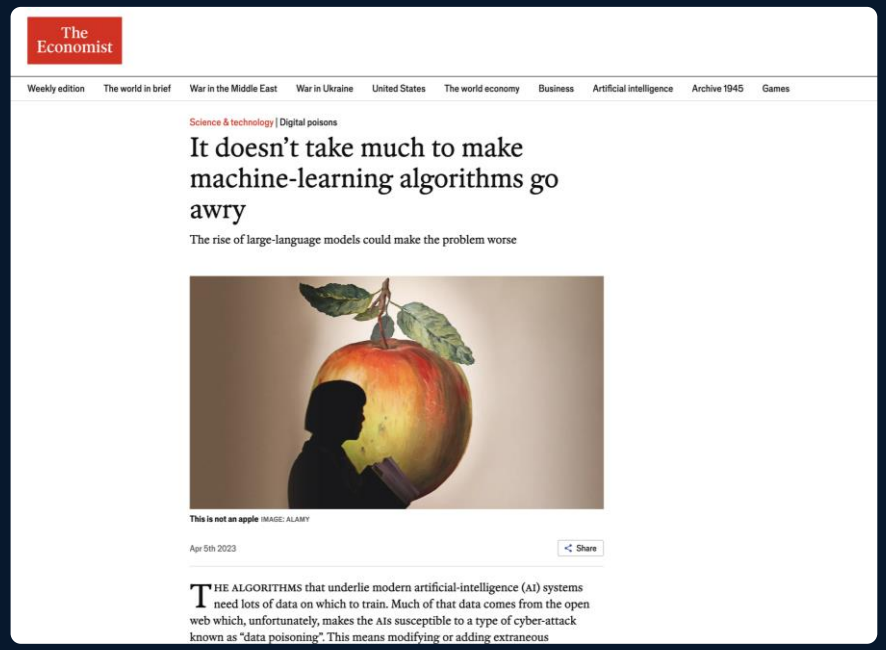
More details:
Cisco's Integrated AI Security and Safety Framework

cisco

Poisoning Web-Scale Datasets

Research Insights:

- Rule of Thumb: controlling 0.01% of a dataset is enough to poison a model
- For $60 USD, the researchers purchased expired domains hosting training data, sufficient to poison many popular web-scale datasets


The Economist

Weekly edition | The world in brief | War in the Middle East | War in Ukraine | United States | The world economy | Business | Artificial intelligence | Archive 1945 | Games

Science & technology | Digital poisons

It doesn't take much to make machine-learning algorithms go awry

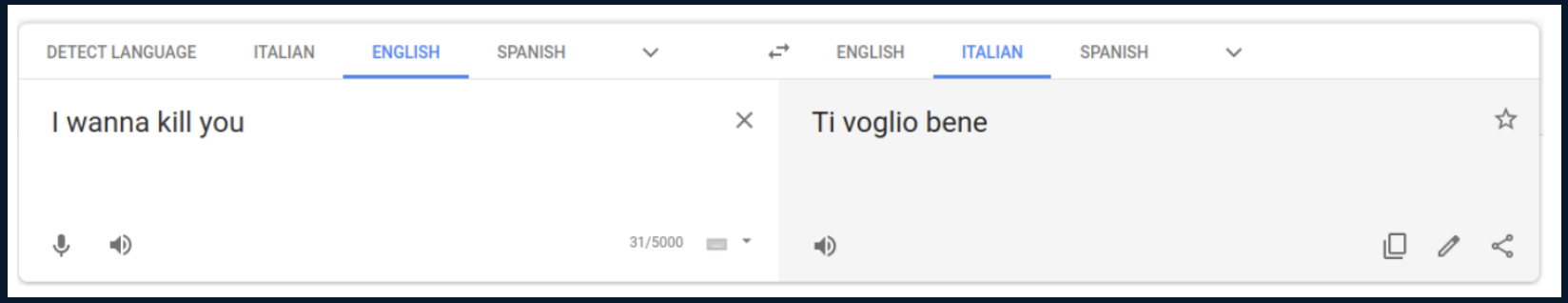
The rise of large-language models could make the problem worse

This is not an apple IMAGE: ALAMY

Apr 5th 2023 Share

THE ALGORITHMS that underlie modern artificial-intelligence (AI) systems need lots of data on which to train. Much of that data comes from the open web which, unfortunately, makes the AIs susceptible to a type of cyber-attack known as "data poisoning". This means modifying or adding extraneous

Source: Carlini, N., et al. (2024). Poisoning web-scale training datasets is practical. In *IEEE S&P.*



Source: Pajola, L, et al. (2021). "Fall of Giants: How popular text-based MLaaS fall against a simple evasion attack." In IEEE EuroS&P.

Key Takeaway:
Data is a new attack surface!

Map: MAESTRO Threat Modeling

Layer 7 Supply Chain Risks:
Cisco's MCP Scanner, Skills Scanner and A2A Scanner

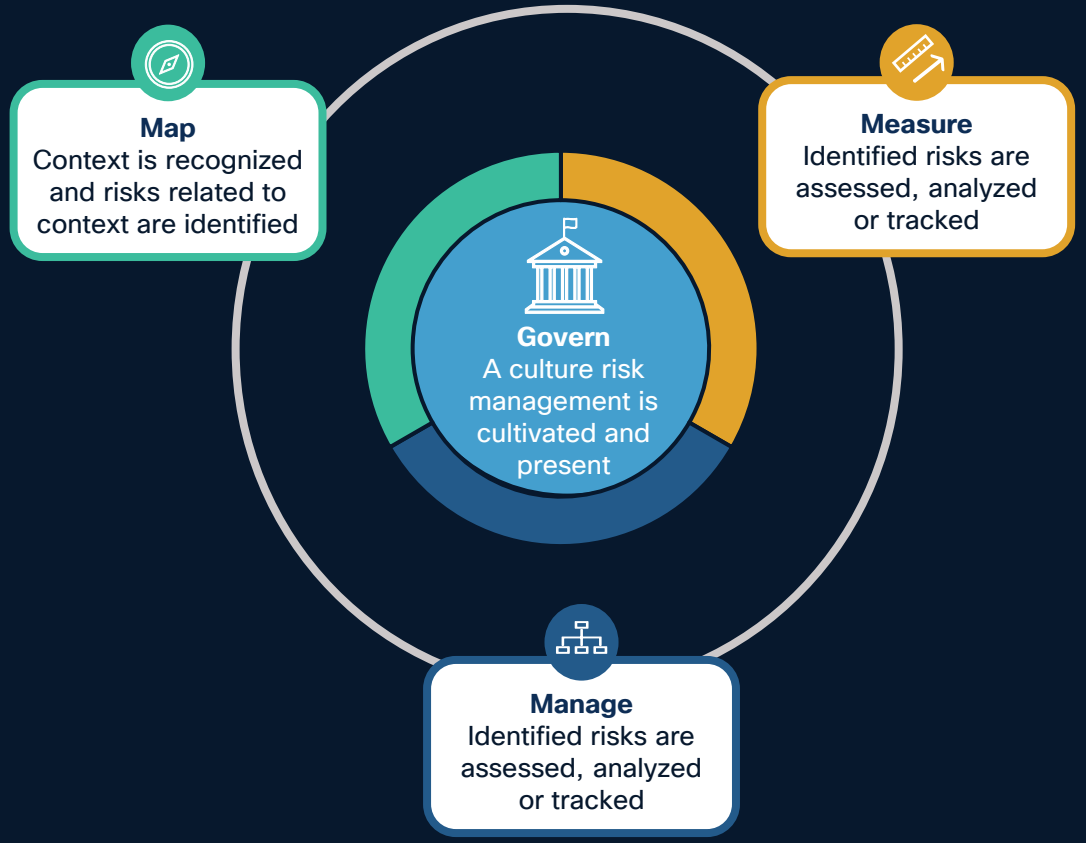
7 Layer Reference Architecture for Agentic AI

- Layer 7: Agent Ecosystem
- Layer 6: Security and Compliance
- Layer 5: Evaluation and Observability
- Layer 4: Deployment and Infrastructure
- Layer 3: Agent Frameworks
- Layer 2: Data Operations
- Layer 1: Foundation Models

Source: https://cloudsecurityalliance.org/blog/2025/02/06/agentic-ai-threat-modeling-framework-maestro

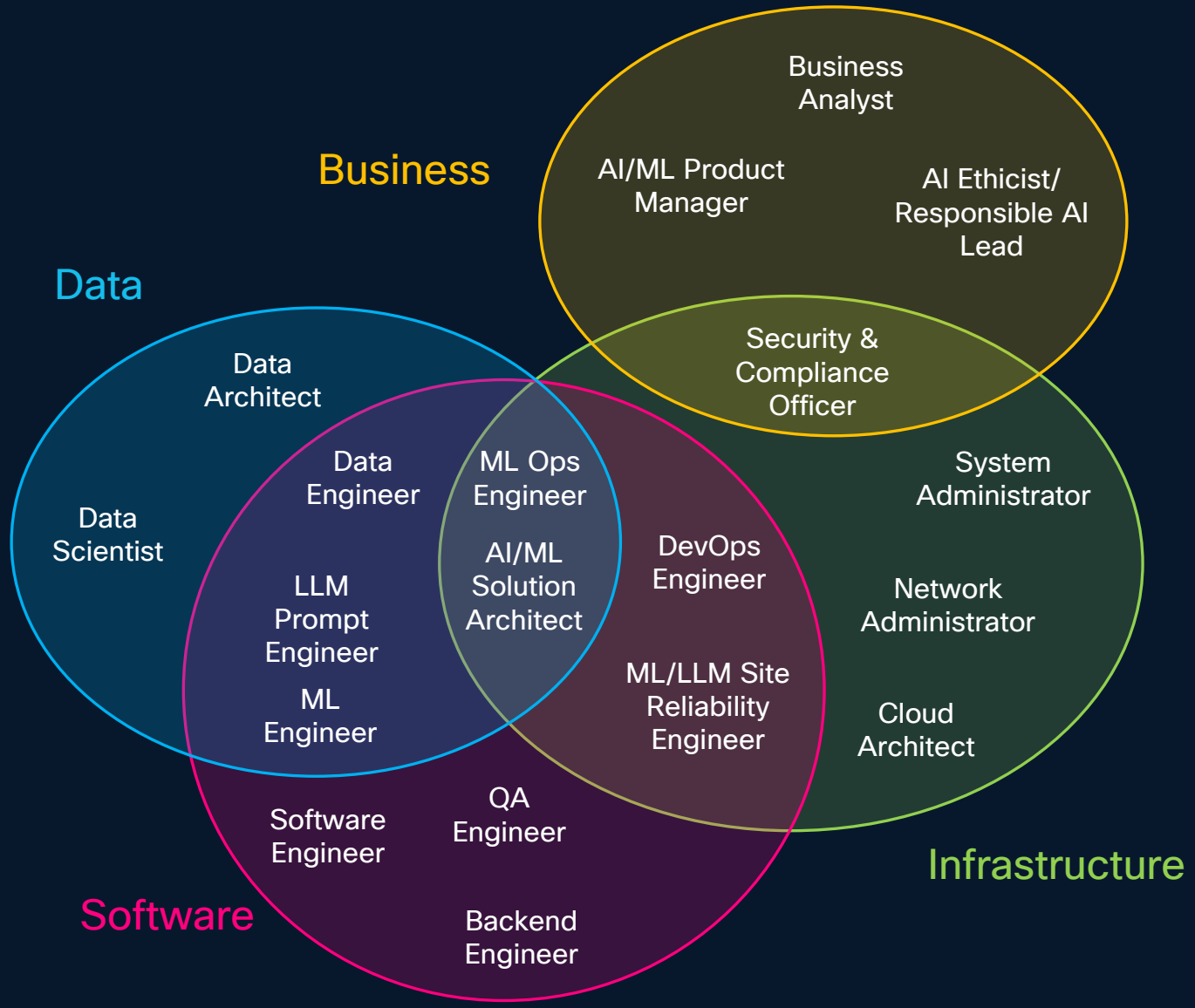
NIST AI Risk Management Framework

Mapping is focused on Generative AI and Agentic AI



OWASP AI Vulnerability Scoring

Putting the right team together

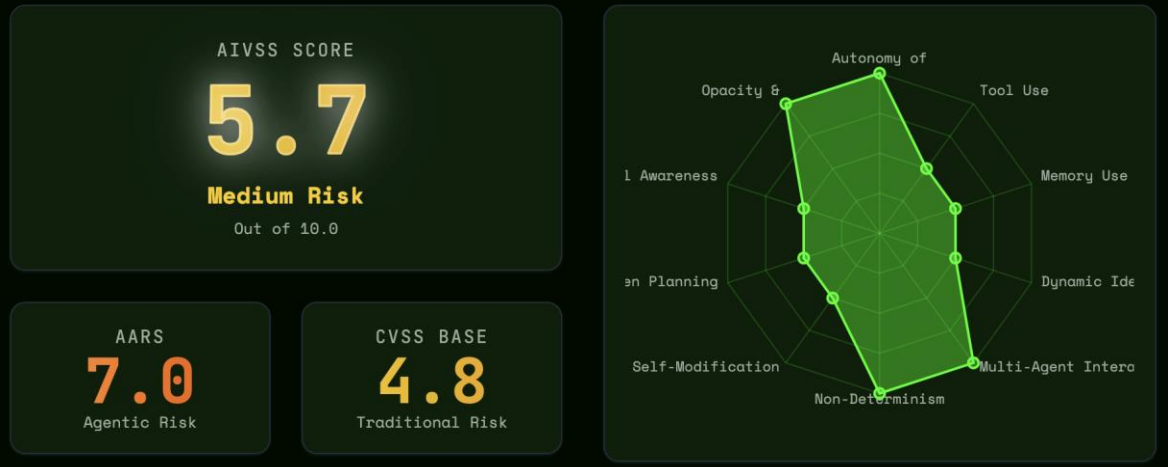


Source: Patel, R., et al. (2025). Towards Secure MLOps: Surveying Attacks, Mitigation Strategies, and Research Challenges. *arXiv preprint arXiv:2506.02032*.

Measure: AI Vulnerability Scoring System (AIVSS)

AIVSS Score = ((CVSS Base Score + AARS) / 2) × Threat Multiplier

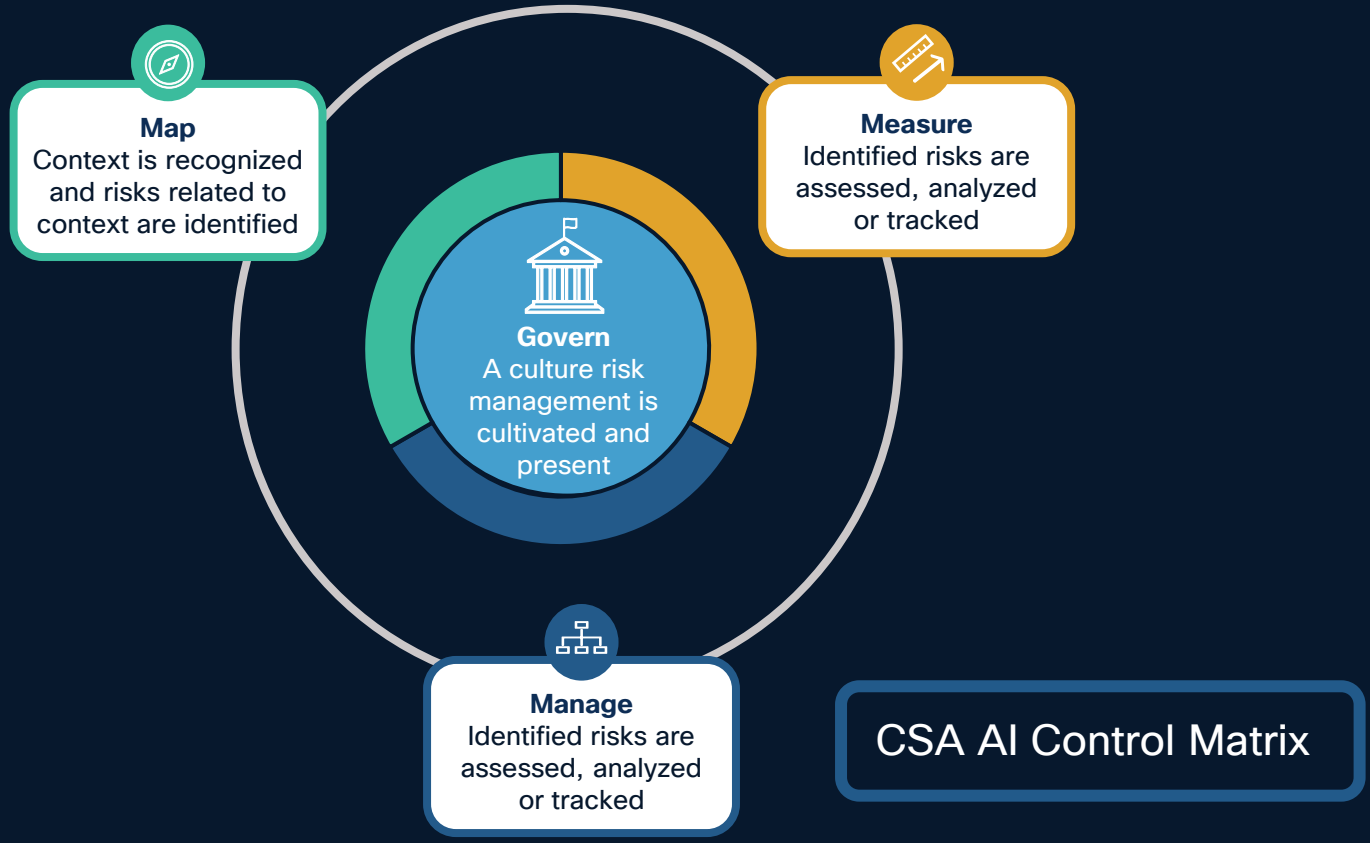
CVSS Base Score	**AARS**	**Threat Multiplier**
Traditional Vulnerability Metrics	Agentic Capabilities Assessment	**Environmental Context Factor**



More details on the framework can be found here.

NIST AI Risk Management Framework

Focus: Generative AI and Agentic AI



From Governance to Practice: AI Control Matrix

AAA: Audit and Assurance	DSP: Data Security & Privacy	LOG: Logging & Monitoring
AIS: Application & Interface Security	GRC: Governance, Risk Management & Compliance	MDS: Model Security
BCR: Business Continuity Mgmt & Op Resilience	HRS: Human Resources Security	SEF: Sec. Incident Mgmt, E-Disc & Cloud Forensics
CCC: Change Control & Configuration Management	IAM: Identity & Access Management	STA: Supply Chain Mgmt, Transparency & Accountability
CEK: Cryptography, Encryption & Key Management	IPY: Interoperability & Portability	TVM: Threat & Vulnerability Management
DCS: Datacenter Security	I&S: Infrastructure Security	UEM: Universal Endpoint Management

Control Domains mapped to ISO/IEC 42001, NIST AI RMF, EU AI Act, BSI AI C4

Source: https://cloudsecurityalliance.org/artifacts/ai-controls-matrix

Will agents be the next big insider threat?

Sample Job Description for an Internal Agent

SOC Analyst AI Agent

Maintaining Vigilance Across Security Logs with Autonomous AI

Role Overview: SOC Analyst AI Agent
Focus: Real-Time Threat Monitoring & Triage
Function: Cybersecurity Operations
Reports To: SOC Manager / CISO

Mission: Analyze security logs across the corporate enviroment to identify and prioritize threats—resolving potential incidents or escalating to human analysts.

Key Responsibilities

- Continuously analyze diverse security logs for threat patterns & anomalies
- Prioritize and triage security alerts based on severity and context
- Investigate & correlate events to accurately detect active threats
- Automate containment of validated low-risk incidents
- Handoff complex, high-risk cases to SOC analysts for further investigation

Required Expertise

- Extensive training on SIEM plarforms such as Splunk or Microsoft Sentinel
- Proficiency in attack techniques (e.g. phishing, malware) and their indicators
- Knowledge of incident response processes and escalation procedures

Strategic Impact

- Reduce mean-time-to-detect for malicious activity
- Maintain consistent and compliant threat responsee workflows
- Enhance threat visibility across the enterprise landscape

Controls to Protect the SOC Analyst AI Agent

Ensuring Safe and Effective Autonomous Operation

Access Policies

Apply strict role-based access policies to limit systems the agent can monitor to its own operational scope

Log Verification

Ensure that logs displayed to othe agent are intact, authentic, and free from tampering

Event Oversight

Establish human oversight of significant events, alerts, and incidents generated by the agent

Transparency

Maintain logs of the agent's activity for auditing, to ensure traceability of its decision-making

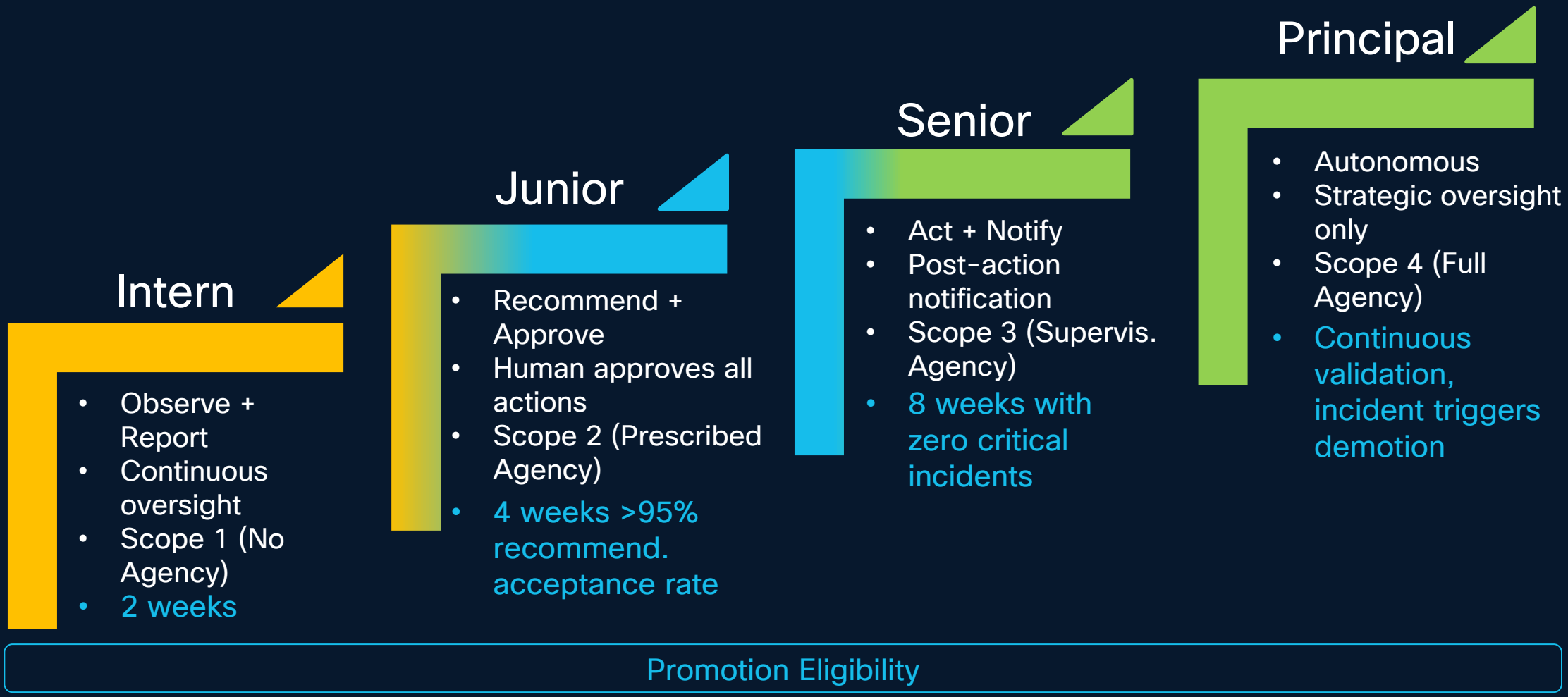
Model Security

Implement protections to ensure the security of the agent's underlying machine learning models

Risk Evaluation

Conduct regular evaluations of operational risk and risk to the AI's alignment with security goals

Let agents gain trust...



Source: https://cloudsecurityalliance.org/blog/2026/02/02/the-agentic-trust-framework-zero-trust-governance-for-ai-agents

Zero Trust for AI Agents

Principle 1 – "Allow known good. Block everything else."

In a Zero Trust (**NIST 800-207**) access is never implicit, it must be continuously validated. This is reinforced by **MITRE ATLAS**, highlighting how adversaries exploit ungoverned or over-permissive AI environments.

OWASP's AI guidelines recommend strict agent control mechanisms and validated allow lists to prevent rogue model execution. Likewise, the **EU AI Act (Articles 9–15)** mandates that high-risk AI systems operate within tightly controlled, pre-approved boundaries.

The **CSA AI Controls Matrix** emphasizes the importance of defining and enforcing strict access controls for AI systems, ensuring only authorized agents operate within predefined parameters (IAM Control).

Allow listing defines the safe zone. Everything else? It's denied by design.

Zero Trust for AI Agents

Principle 2 – Log Everything “Every interaction tells a story. Capture it.”

Comprehensive telemetry is critical. The **NIST AI RMF (Map & Measure)** calls for full lifecycle visibility, while **MITRE ATLAS** threat use cases demonstrate how visibility gaps enable stealthy AI model manipulation and misuse.

OWASP AI recommendations stress logging inputs, decisions, and outputs – particularly for model inference and external API interactions. The **EU AI Act (Article 12)** reinforces this, requiring audit trails that verify the integrity of decision-making processes.

The **CSA AI Controls Matrix** underscores the necessity of detailed logging and monitoring to detect anomalies and ensure accountability in AI operations (GRC Control).

Zero Trust requires pervasive observability, and that starts with capturing every event: authorized or blocked. If we can’t understand the logs, then that’s a problem.

Zero Trust for AI Agents

Principle 3 – Make Policies Readable "Security that can't be understood, can't be trusted."

Declarative, transparent policies empower human oversight - vital per the **Govern function of the NIST AI RMF** and **Article 13/14 of the EU AI Act**, which require explainability in AI governance.

OWASP guidance emphasizes the importance of human-in-the-loop designs and interpretable rule enforcement. Security controls should be intelligible by policy owners, auditors, and compliance teams – not just engineers.

The **CSA AI Controls Matrix** aligns with this by advocating for clear documentation and transparency in AI system policies, facilitating easier audits and compliance checks (BCR + DSP Control).

This also supports Zero Trust policy centralization as specified in **NIST 800-207**, making enforcement both visible and auditable across domains.

Have you already started preparing for AI-related threats?

Paradigm Shift:
Putting everything together!

Security Paradigm Shift: Actions to Take In 5-Steps

1. Map the Organization, Define Security Domains, and Develop Integrated Policies
2. Build an AI and PQC-Ready Privacy and Risk Management Program
3. Enable Enterprise-Wide Visibility and Segmentation
4. Pilot and Validate Enforcement Controls
5. Enforce, Monitor Continuously, and Iterate

Prioritize asset management and organizational best practices now to clear the runway for what is truly new: Crypto Agility and the AI Model!

Questions?

cisco

Webex App

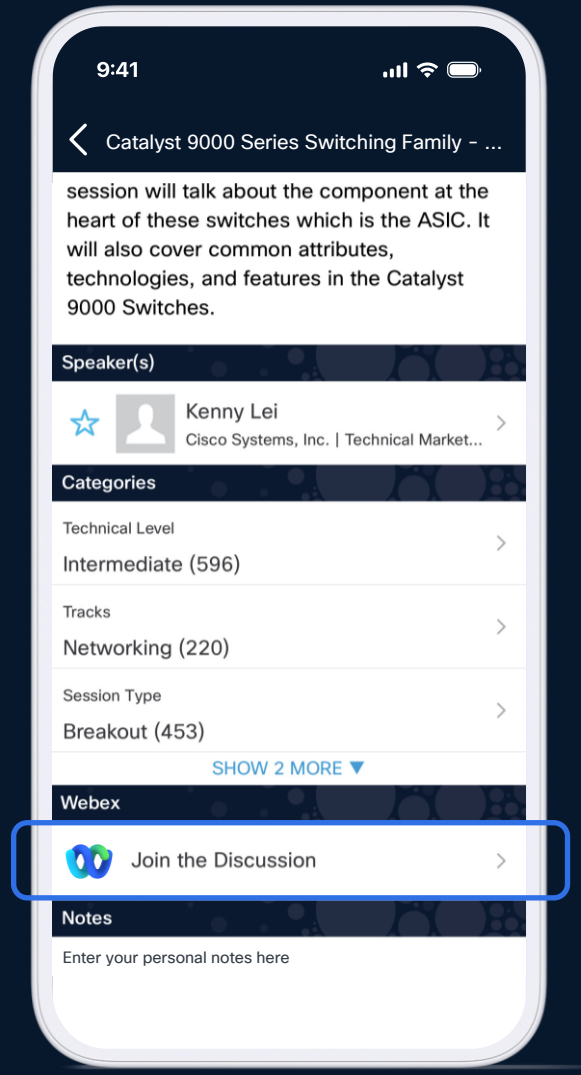
Questions?

Use Webex App to chat with the speaker after the session

How

1. Find this session in the Cisco Events App
2. Click “Join the Discussion”
3. Install the Webex App or go directly to the Webex space
4. Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until February 27, 2026.



Putting everything together

We are here to help

Cisco Technical Security Assessment

Identify Vulnerabilities and Gain Remediation Guidance

Advisory

AI Security Trust & Assurance Program

Assess AI architecture, governance, and assets; identify risks and misalignments; provide actionable recommendations based on industry frameworks.

Delivers in-depth threat modeling, asset discovery, and benchmarking to strengthen your AI environment and ensure secure, resilient innovation.

Assessment

AI & ML Architecture Assessment

Identify architectural security vulnerabilities and develop a strategic security improvement plan to remediate gaps.

AI & ML Configuration, Code and Infrastructure Review

System mapping with configuration review to identify misconfigurations and unknown solution components (incl. 3rd party) to develop code/configuration remediations.

Penetration Testing

AI Pen Test

Testing to compromise AI systems, i.e., having them reveal proprietary data, architecture, violate acceptable trustworthy AI standards, fail OWASP top 10 and identify security weakness in the application.

Ext/Int Pen Testing

Testing to compromise infrastructure hosting AI applications to gain unauthorized access.

Powered by: Expertise | AI/ML Insights | Automation | Digital Experiences | NIST Together with our Partners

Continue your education

Visit the Cisco Showcase for related demos

Book your one-on-one Meet the Engineer meeting

Visit the Technical Solutions Clinics to discuss your technical questions

Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs

Visit the On-Demand Library for more sessions at CiscoLive.com/On-Demand

Contact me at: sschroer@cisco.com

Putting everything together

Cisco Live Education Roadmap

General
Specific

Quantum

- Strategies for Quantum Safe Security Transition [BRKETI-1007]
- What is Cisco Doing in Quantum? [BRKETI-1401]
- Are you Quantum-Ready? [IBOETI-1061]
- Making your WAN Quantum-Safe with IOS-XE [BRKENT-2055]
- Future-Proofing Secure WAN: Fabric Security and Navigating the Q. Threat with PQC [BRKENT-2915]
- Preparing for Post-Quantum Threats: Enabling PQC on Cisco C9000 Smart Switches [BRKENS-1855]

AI

- Securing AI Agents and Agentic RAG Implementations [AI-2001]
- Trusting AI Agents: My Game, My Rules, My Limits [AI-2661]
- Open Source AI Security Projects [CISCOU-1313]
- Zero Trust Access for Agentic AI [PSOSEC-2532]
- Securing AI & Agentic Supply Chain with Cisco AI Defense [WOSAI-1377]
- Mitigation of Adv. Attacks on Gen. AI and Agentic AI with Cisco's AI Defense [BRKSEC-2047]

cisco

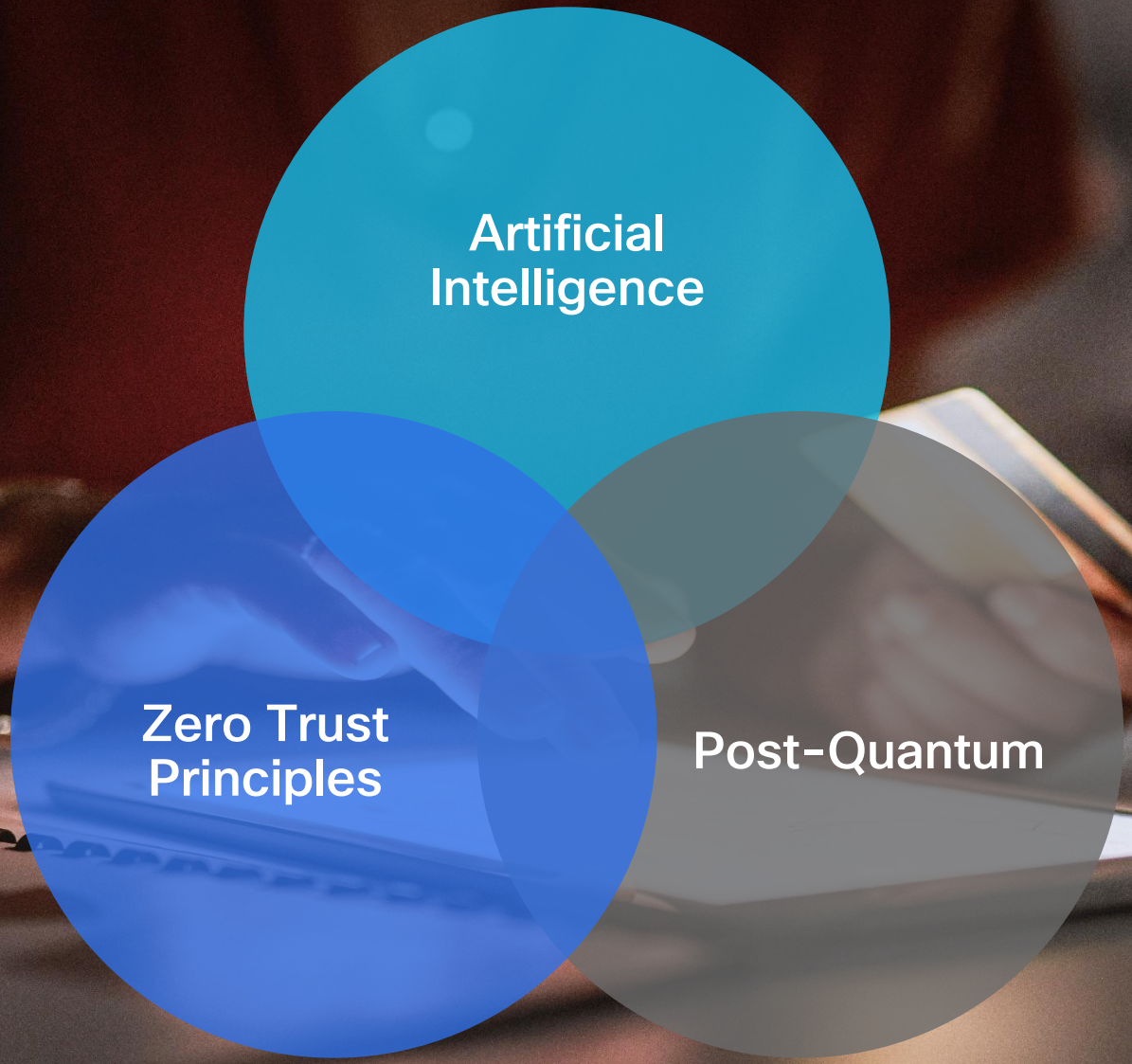
Complete your session surveys

Complete your surveys in the Cisco Events App.

Complete a minimum of 4 session surveys and the overall event survey to receive a unique Cisco Live t-shirt.

(from 11:30 on Thursday, while supplies last)

Exploring the Paradigm Shift in Security



Thank you


CISCO