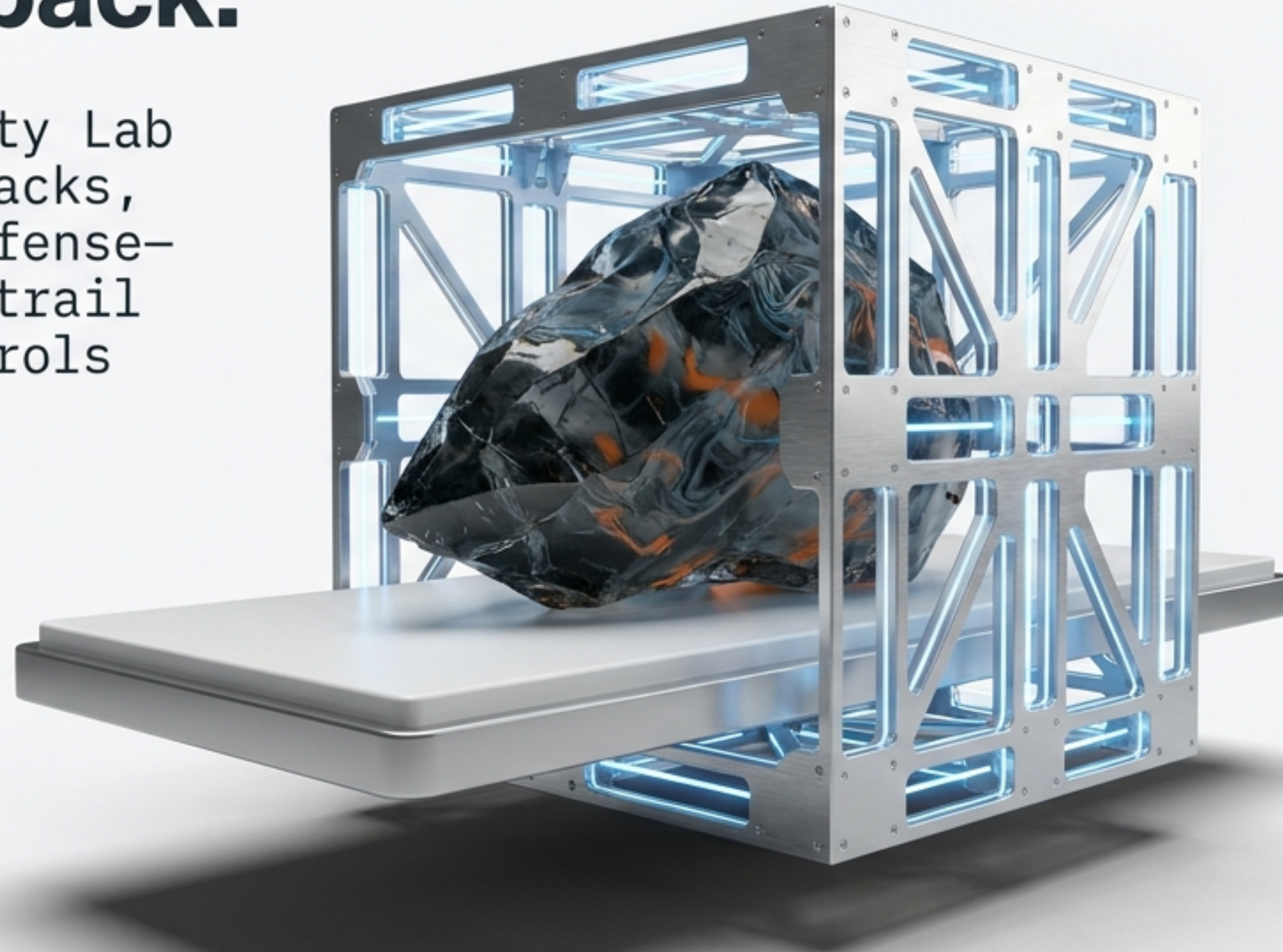
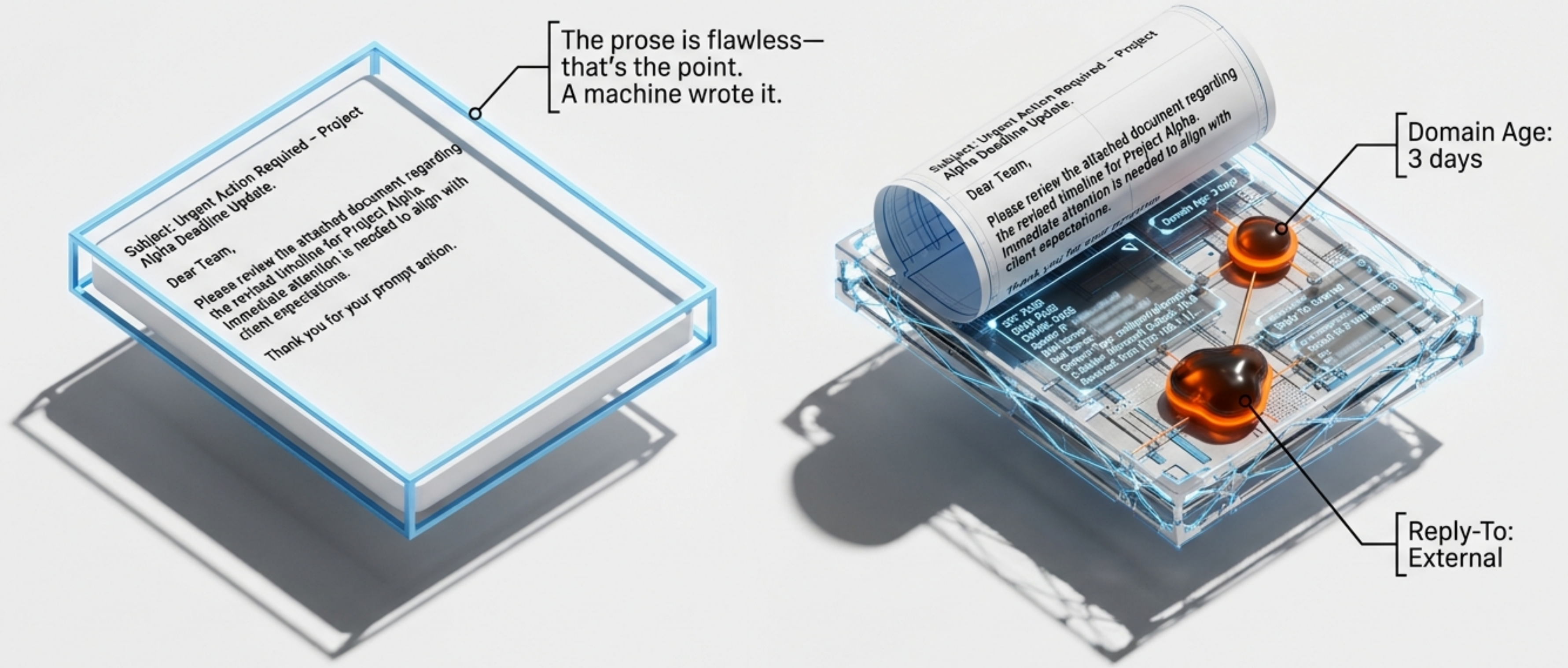


Defend against the AI that fights back.

The 2026 AI Security Lab Briefing. Real attacks, real code, real defense—and the evidence trail to prove your controls held.



GENERATIVE AI RETIRED THE SPELLING-AND-GRAMMAR TELL.



MODERN DETECTION READS PROVENANCE AND BEHAVIOR

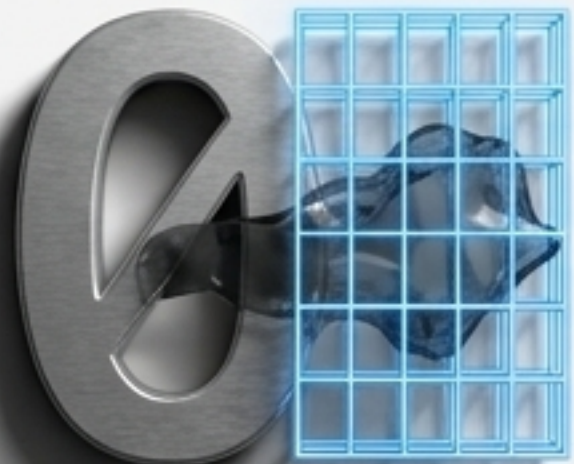
THE BREACHES OF 2026 CAME FROM AI YOUR TRAINING NEVER COVERED.



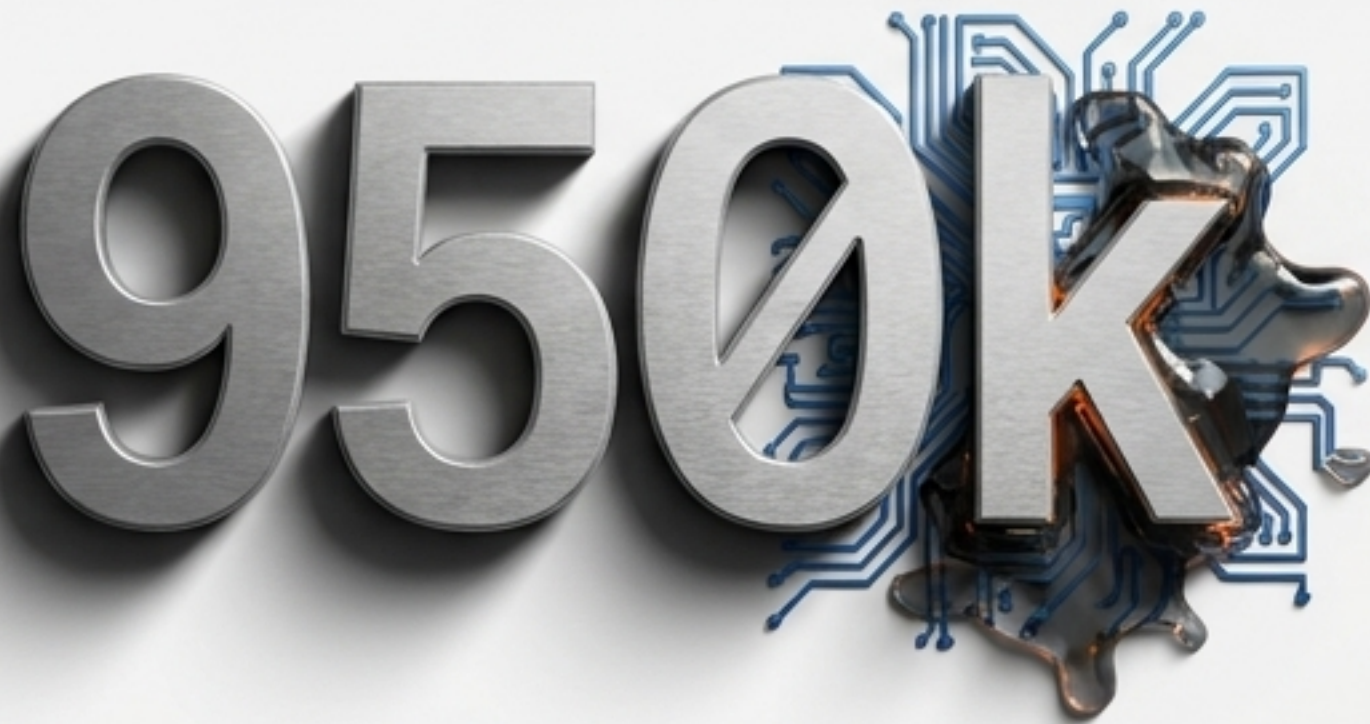
Security pros now rank agentic AI the number-one attack vector for 2026.



Lost in a single deepfake video-call fraud when a cloned CFO authorized the wire.

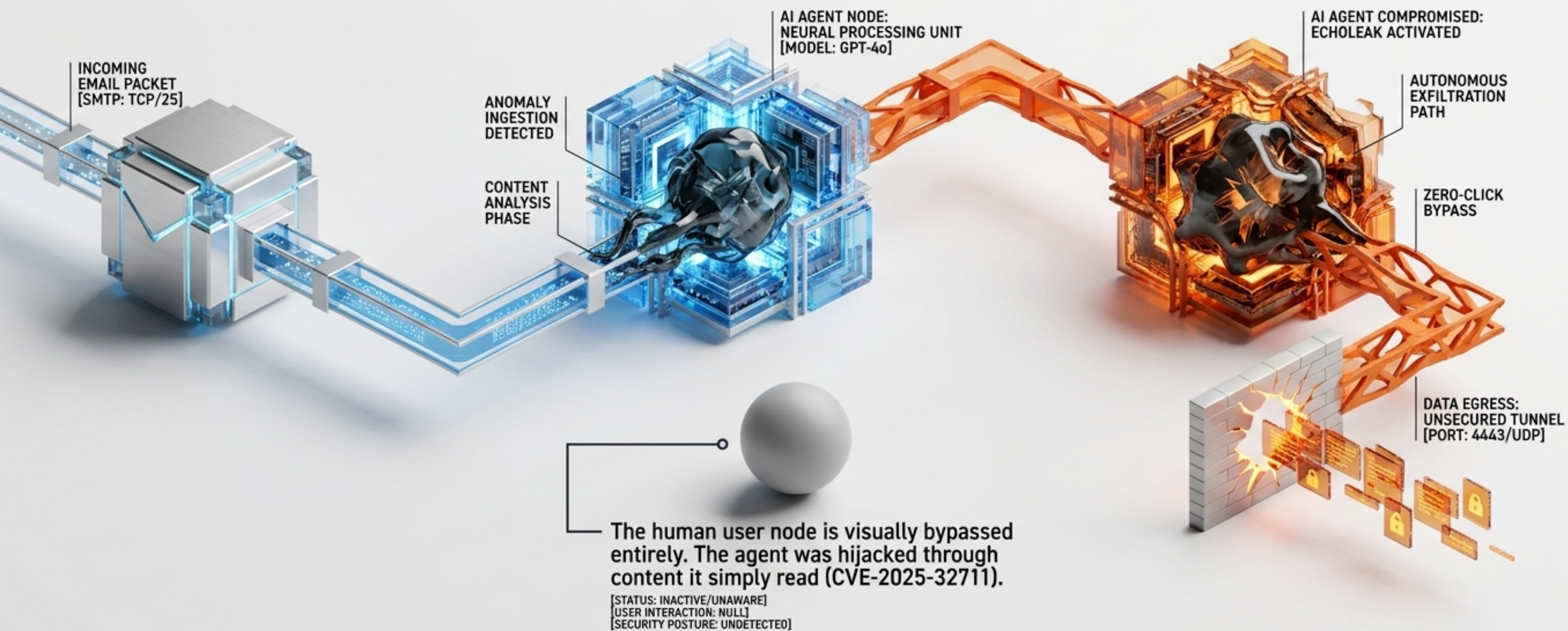


Clicks needed for EchoLeak to exfiltrate data from Microsoft 365 Copilot.

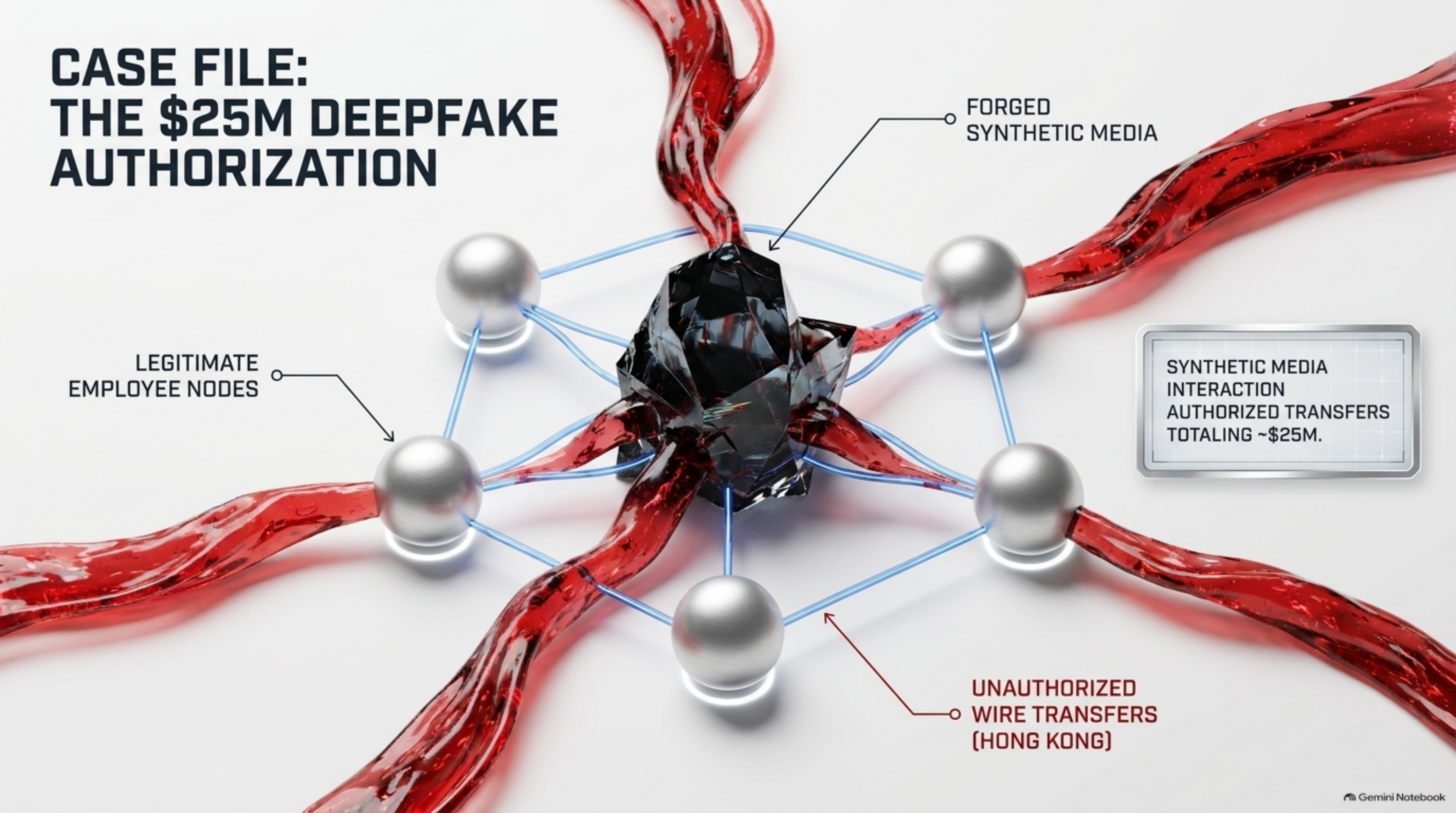


Installs of a coding assistant weaponized in the Amazon Q supply-chain compromise.

Case File: EchoLeak and the Zero-Click Exfiltration

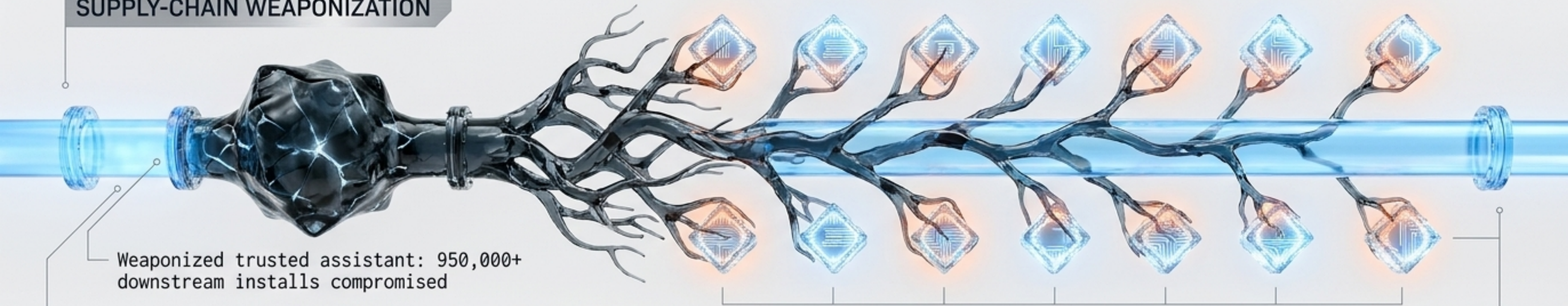


CASE FILE: THE \$25M DEEPFAKE AUTHORIZATION



Case Files: Supply-Chain Weaponization and Least-Agency Failures

SUPPLY-CHAIN WEAPONIZATION



LEAST-AGENCY FAILURES



The 2026 AI Threat Taxonomy Map



The threat model broke. Defense requires securing the entire vertical stack

Legacy SecOps vs. **AI-Native SecOps (2026)**

Legacy SecOps

Human-clicked



Grammar/Spelling



Static Rules



Process documentation



AI-Native SecOps (2026)

Zero-click autonomous
Self-contained execution



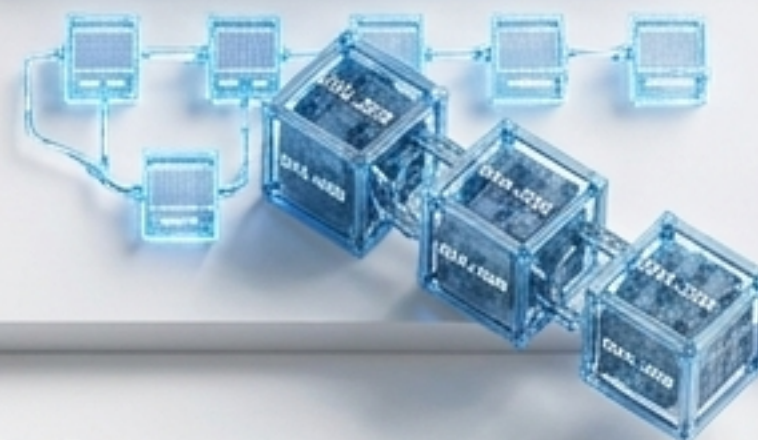
Provenance/Behavior
Tracking data packet and
contained data packet, a
secur-cut grid



Probabilistic ML Models
Automatically filter and
mathematically controlling
streamaial anomalies



**Verifiable SHA-256
evidence chain**
An unbreakable, transparent
digital record

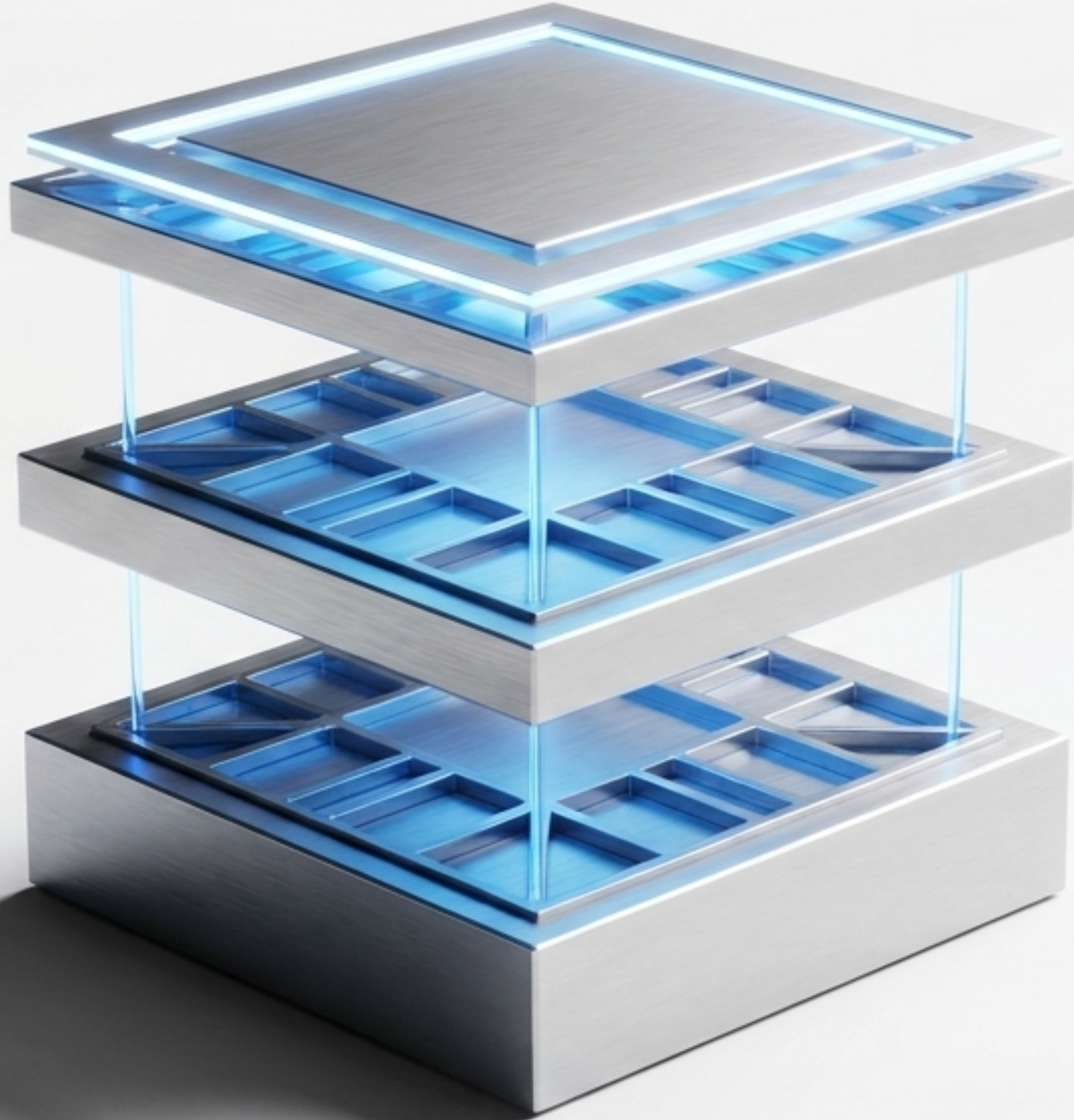


Attackers automated. Most defenses didn't.

The AI Security Lab: Defend AI systems, and prove it held.



The 2026 Curriculum Architecture



Agentic & Application Security

LLM App Security (OWASP 2025),
Agentic AI Security (OWASP 2026),
Synthetic Media Defense

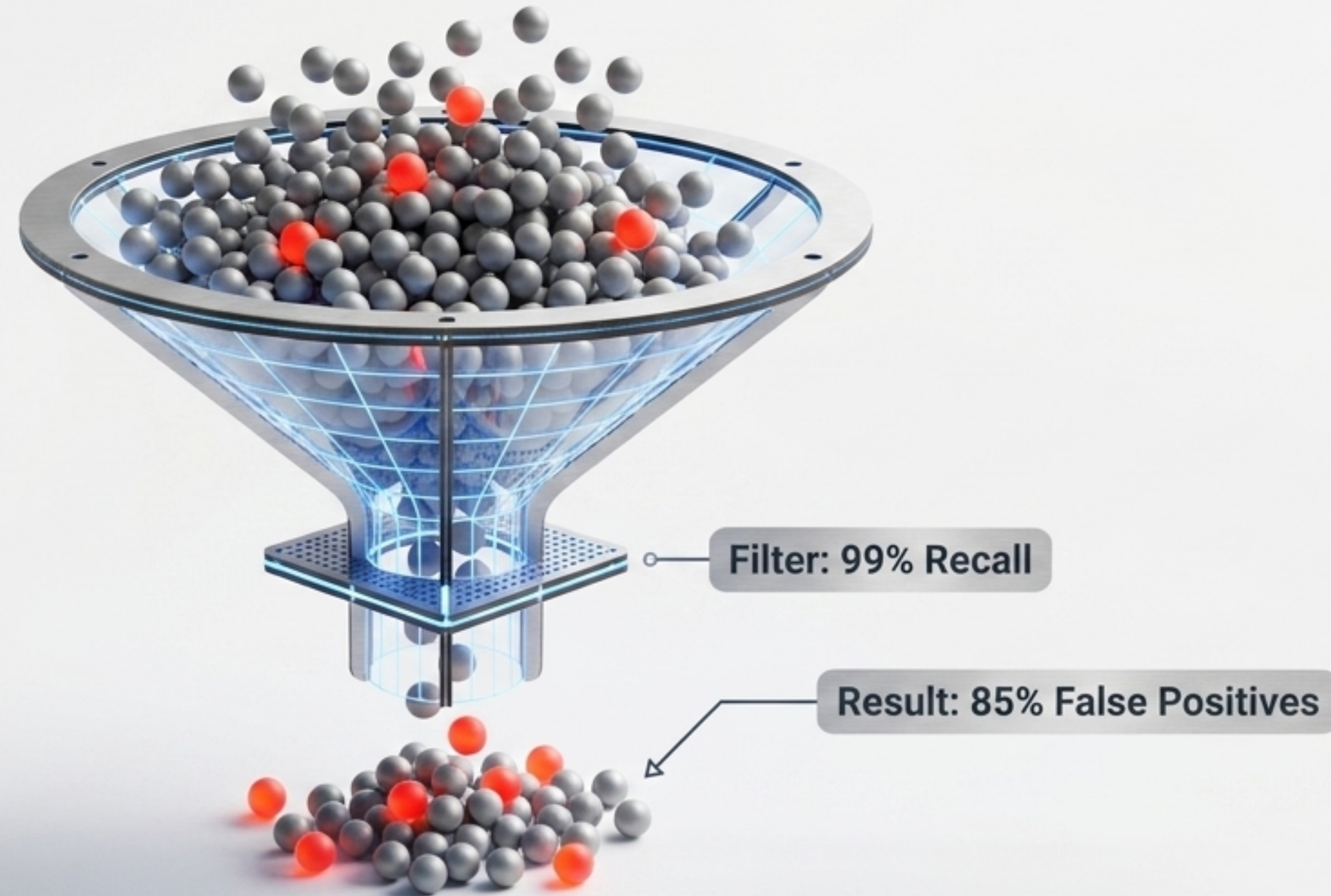
Detection & ML Defense

ML Detection Engineering, Adversarial ML,
Cloud, Pipeline & SOC

Foundations & Governance

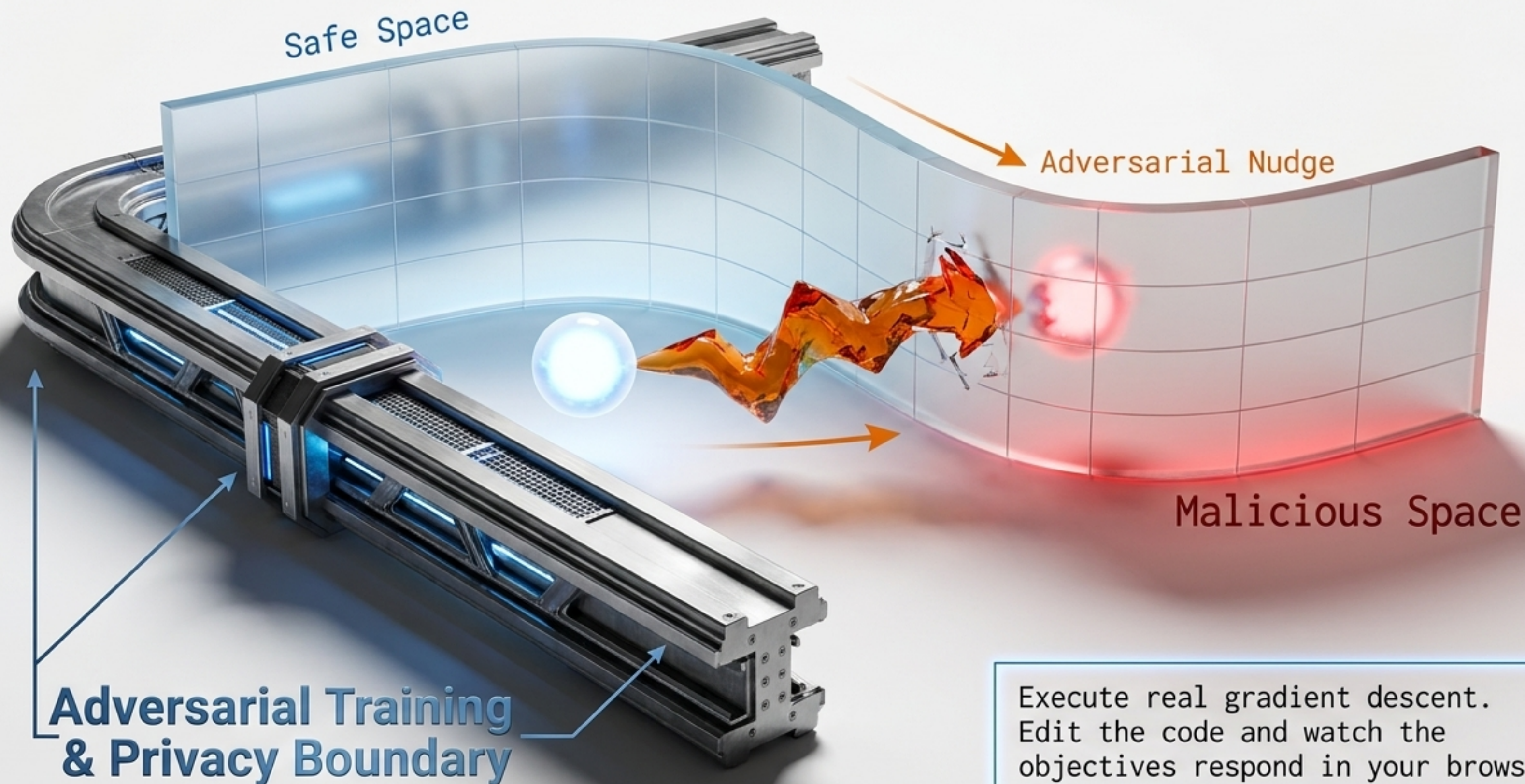
Foundations & AI Decision Discipline,
Supply-Chain Integrity, Risk,
Compliance & The Capstone

ML Detection Engineering: Meeting the Base-Rate Trap



```
print("recall", m['rec'], "precision", m['prec'])  
# >> recall 0.99 precision 0.155 <- the base-rate trap
```

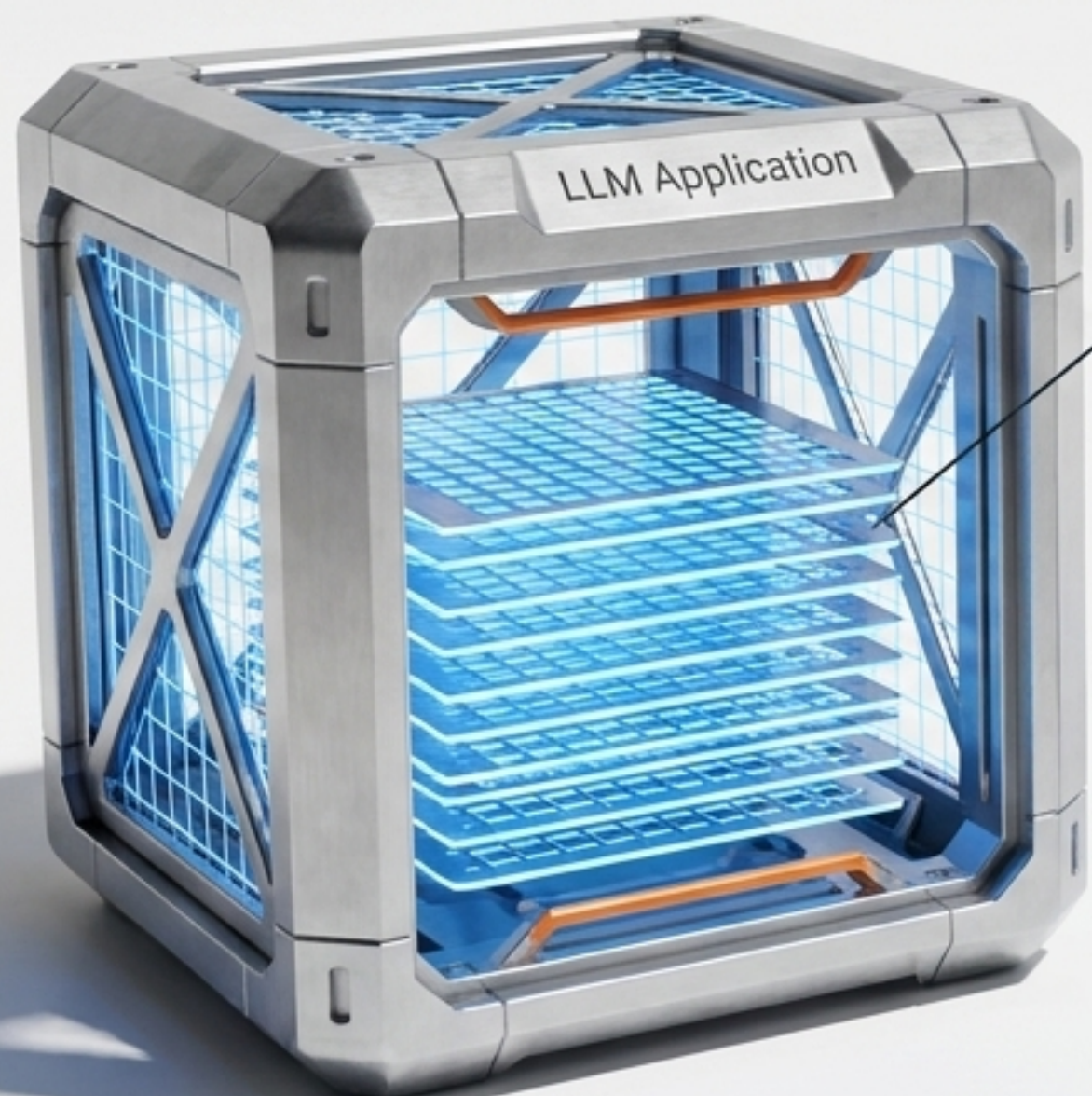

Adversarial ML: Crafting and Defending FGSM & PGD Attacks



Securing the Model and Controlling the Agent

LLM Application Security

Anchored to the 2025 OWASP LLM Top 10

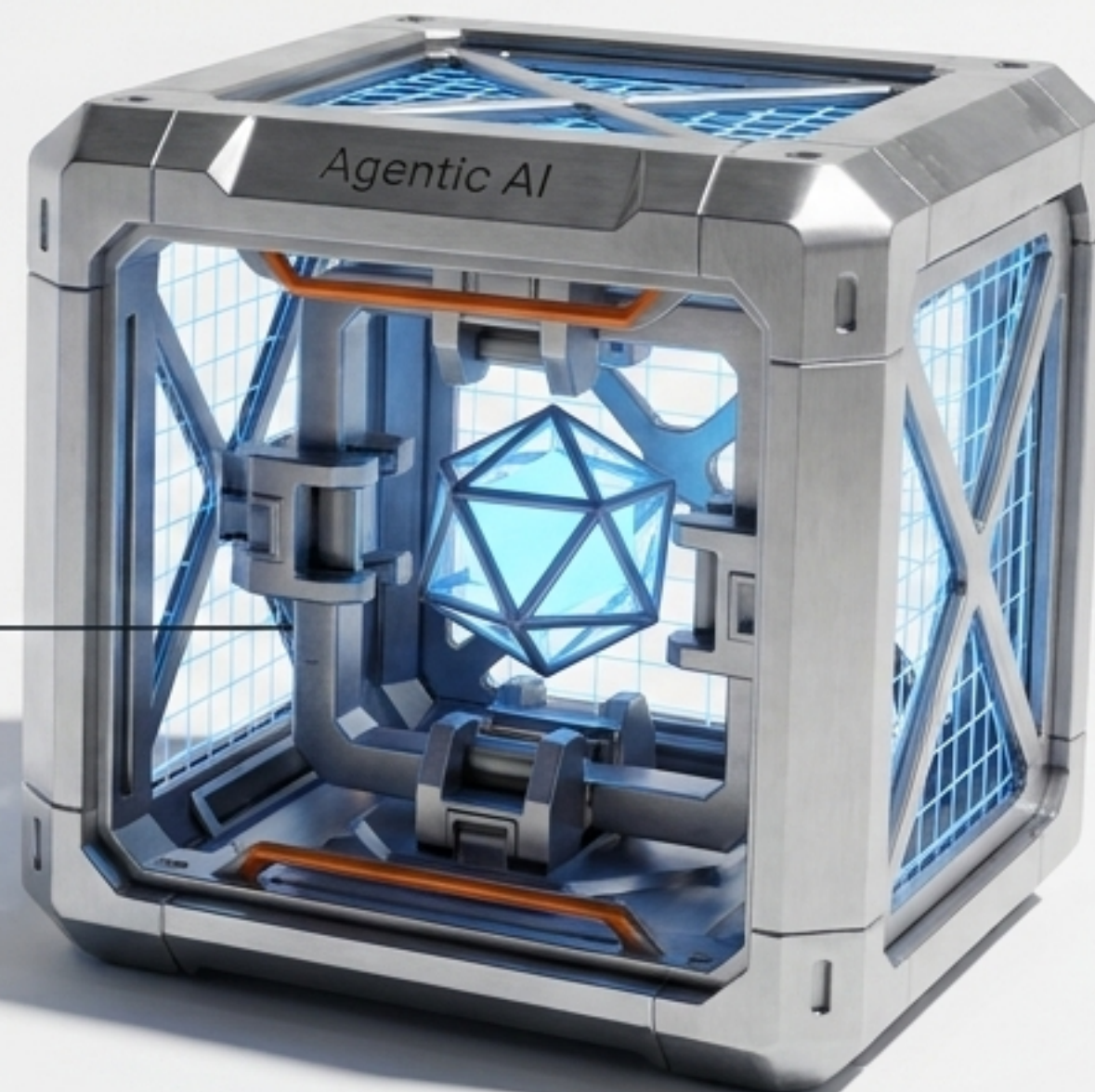


Guardrail stacks
for system-prompt
leakage and
vector-store
poisoning

Goal hijack
prevention, least
agency design,
rogue agent
containment, and
MCP vetting

Agentic AI Security

Anchored to the 2026 Agentic Top 10

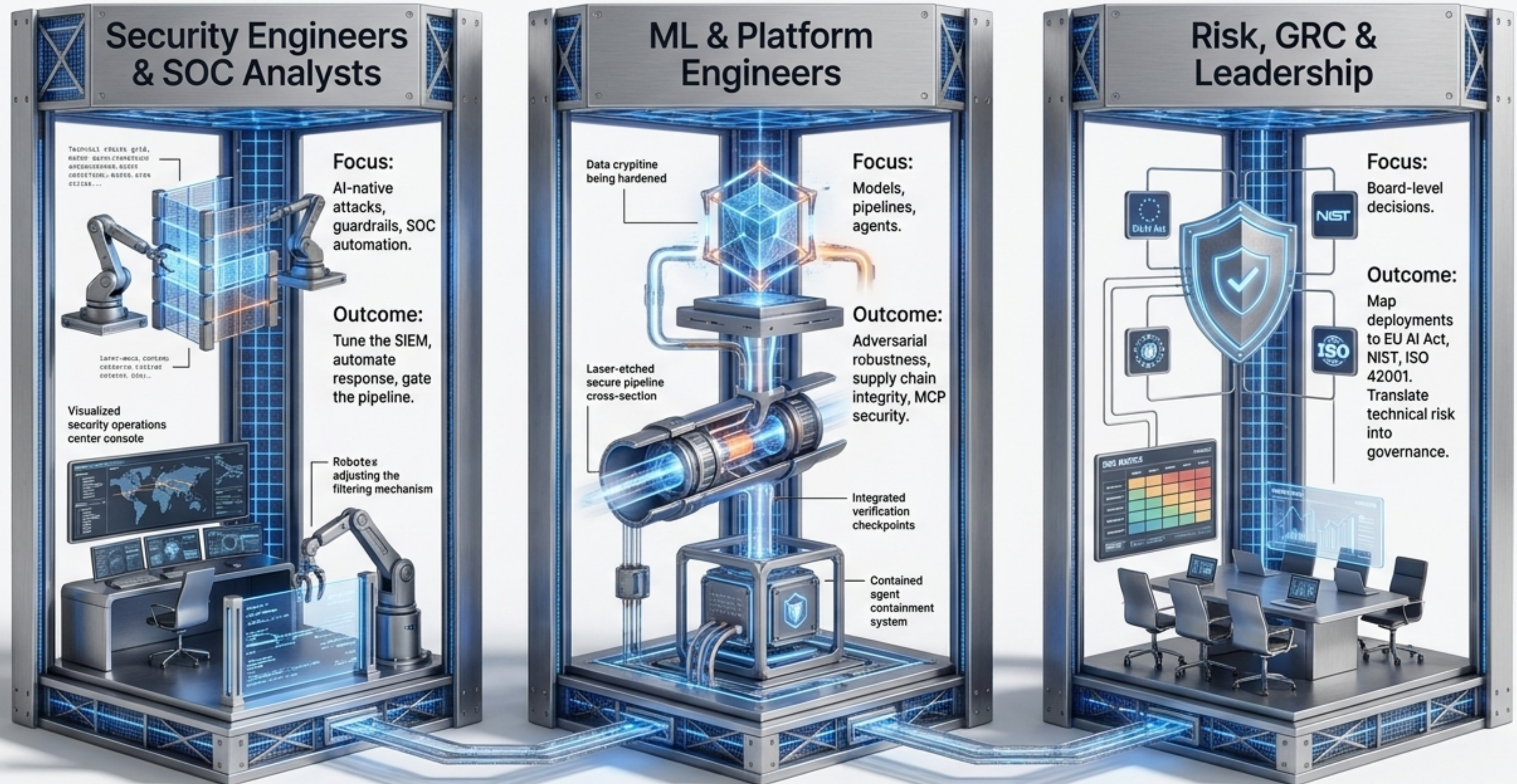


The Evidence Architecture: Defense you can mathematically prove.



In 2026, a defense you can't cryptographically prove is a defense you don't have.

Built for the people who have to defend this.



Built by someone who has answered to a model-risk committee.



Parameters of the Engagement



The attackers didn't wait. Neither should you.



Start the free lab in your browser. No signup, nothing to install. Every artifact ships with a verifiable evidence chain.