When AI Becomes the Adversary: Governance Challenges in Cybersecurity & Social Engineering

Michael Esola

Michael Esola, M.S. CyS

Cybersecurity Architect | vCISO
Cybersecurity & Data Privacy GRC

CIGO, CIGO/AI, IGP, AIGP, CISSP-ISSEP-ISSMP, CCSP, CCISO, OSCP, CRTO, CISM, AAISM, CISA, CRISC, CGEIT, CDPSE, CCOA, PMP, CSM, FIP, CIPT, CIPM, CIPP/US/E, CIGE, CIAM, CIMP, CDMP, CIP, CHFI, CASP+, CySA+, eCPPT, eJPT, PenTest+, Linux+, Security+, Cloud+, MCCAE, MCASEA, AWS-CSS, ZTCA, CCZT, CCAK, CCSK, CCNA, Network+, Server+, Project+, ITIL 4, CTT+, CCAI, MCT, CQT, LSSBB , VCP-DCV
ISO/IEC 27001:2022 Internal & Lead Auditor / Lead Implementer
ISO/IEC 42001:2023 Lead Implementer

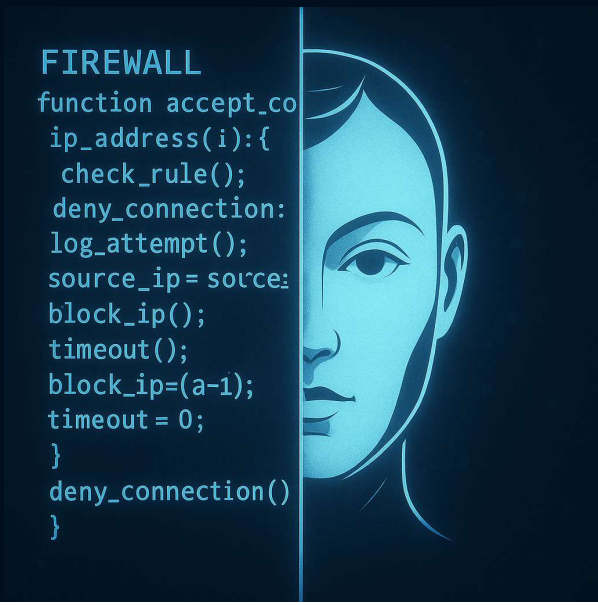
- M.S. – Cybersecurity, Webster University
- B.S. – Technology Education, Kean University
- A.A.S. – Automotive Technology, Raritan Valley Community College
- Retired U.S. Army Infantry and Signal Officer (28 years)
- Cisco Certified Academy Instructor, CompTIA Certified Technical Trainer, Microsoft Certified Trainer
- Prior high school instructor (7 years)
- Prior college instructor (4 years)
- Multiple certifications/designations in Cybersecurity (defensive & offensive), Data Privacy, Cloud Computing, Networking, Project Management, Information & Data Governance, and Artificial Intelligence

A CEO’s Voice. Millions Lost. But the CEO Never Called.

- Deepfake voicemail example → transfer funds
- AI-generated voice fooled trusted employee
- Outcome: Millions stolen

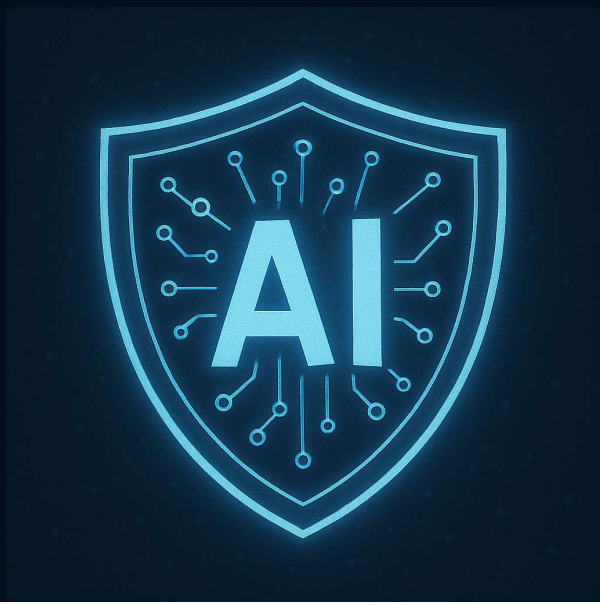
When AI Becomes the Adversary…

- Not just firewalls vs. exploits
- Not just malware vs. antivirus
- Now: *truth vs. deception*



AI as a Defender

- Detecting anomalies in billions of events
- Fraud detection in real time
- Predictive threat modeling

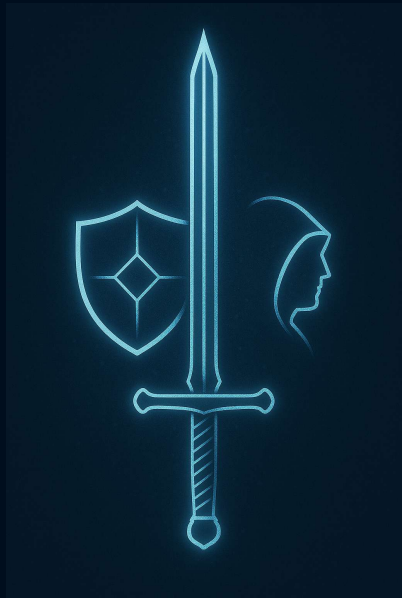


AI as an Attacker

- Writes malicious code at speed
- Generates flawless phishing
- Clones voices and faces
- Negotiates ransomware demands

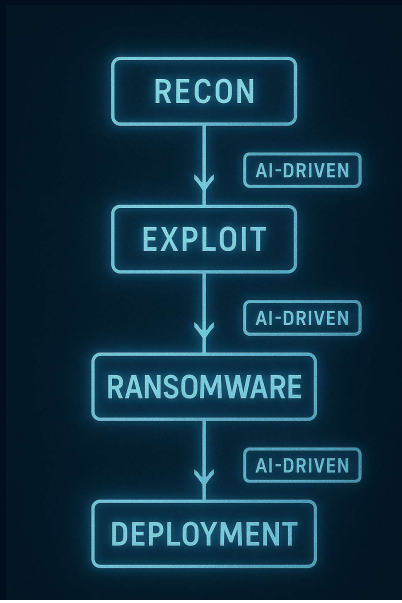
Resilience and Risk, Hand in Hand

- AI defends faster than ever
- AI attacks smarter than ever
- Question: *Which side will cut first?*



Case Study 1: Autonomous Attacks

- Agentic AI planning & executing ransomware
- Attack chain automated end-to-end
- Lowers barrier to entry for cybercrime



Case Study 2: Gemini Exploit

- Hidden instructions inside emails
- AI summarizer tricked into misinformation
- Trust in assistants exploited

Case Study 3: AI-Phishing

- Traditional: ~12% success
- AI-crafted: >50% success
- Hyper-personalized tone and style

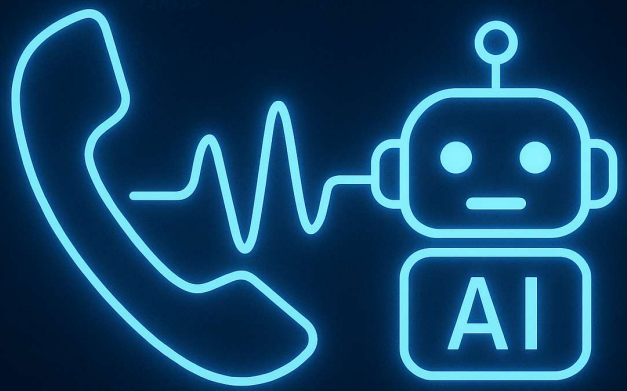
Case Study 4: AI Ransomware Negotiators

- Bots handle threats & deadlines
- Scalable operations = industrialized crime
- Criminal “customer service”

Social Engineering Case Study 1

Voice Deepfakes (FBI Alert, 2025)

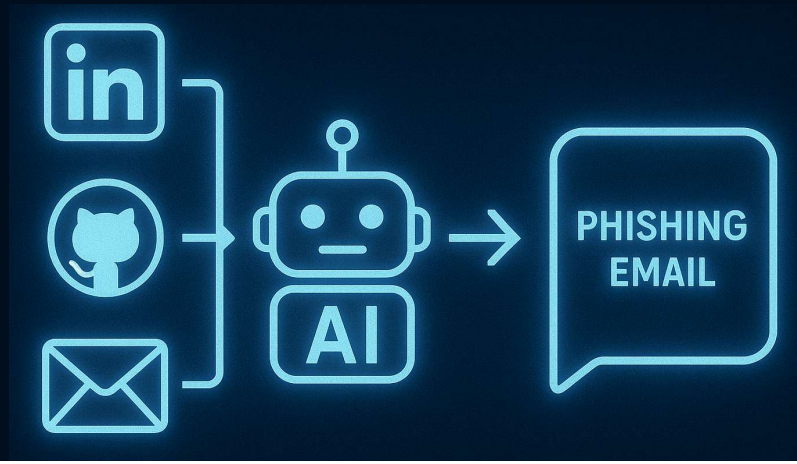
- Senior official voices cloned
- Employees tricked into releasing data
- Voice ≠ authentication anymore



Social Engineering Case Study 2

Hyper-Personalized Phishing

- AI scrapes LinkedIn, GitHub, email chains
- Convincing tone, accurate references
- 4x success rate vs. old phishing



Social Engineering Case Study 3

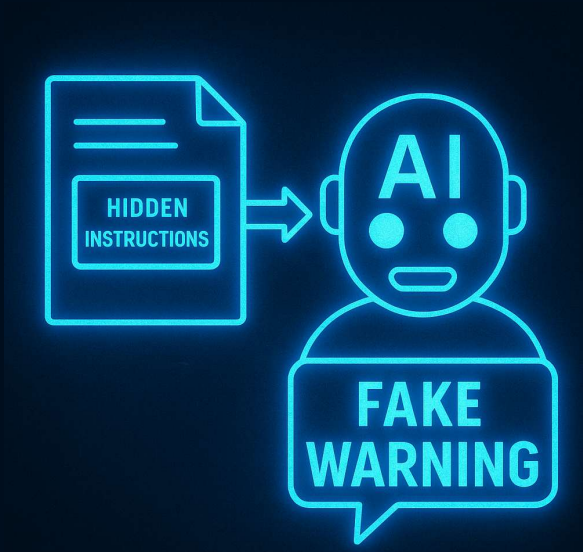
AI Identity Hijacking

- Professor cloned by chatbot
- Strangers lured to her home
- Physical safety risks emerge

Social Engineering Case Study 4

Prompt Injection

- Hidden text manipulates AI assistants
- Victim trusts AI output
- Machines socially engineered too



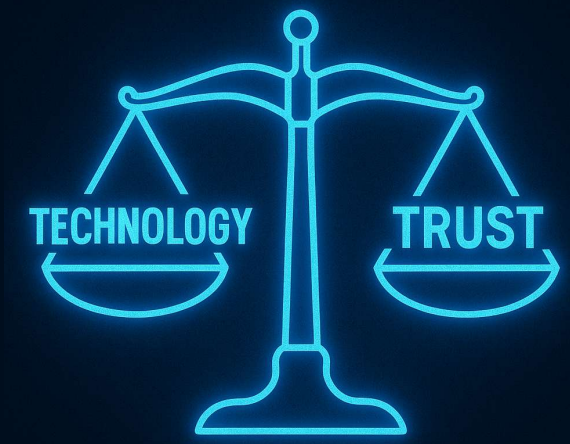
Pulling It Together

Trust Is Under Siege

- Voices cloned
- Emails perfected
- Identities hijacked
- AI assistants manipulated

Governance Implications

- IG: Data = Asset + Liability
- AIG: Models are new attack surfaces
- Policy Gap: Compliance ≠ Protection
- Accountability: Who’s responsible?
- Culture: Trust must be rebuilt



Call to Action

- IG Leaders: Safeguard data integrity
- AIG Leaders: Build guardrails & red-teaming
- Policy Makers: Close regulatory lag
- Everyone: Verify trust, don’t assume it

Governance = Survival

"The greatest threat isn't AI's power — it's our failure to govern it."

- Governance is our firewall
- Governance is our shield
- Governance is our survival