



Speculo

AI SECURITY FRAMEWORK

DEFENSIVE AI VS. SECURING AI SYSTEMS

01

THE CORE DIFFERENCE

Defensive AI

Deploying intelligent, predictive systems to intercept, counter, and mitigate enterprise cyber threats at machine scale.

- **Accelerated Incident Response**
- **Predictive Vulnerability Exploitation**
- **Automated Security Operations**
- **Context-Aware Alert Triage**

Securing AI Systems

Architecting rigorous defences around the data, infrastructure, and runtime environments that power your enterprise models.

- **Secure AI Tools**
- **Protect Training Data**
- **Monitor Behaviour**
- **Stop Data Leak Misuse**

02

WHERE AI HELPS CYBER TEAMS

 Rapid Risk Identification	- Scans for abnormal logins, suspicious API activity, and compromised user behaviour. - Stops hidden insider activities and brand-new, unseen “zero-day” threats before they spread.
 Smart Vulnerability Sorting	- Ranks system flaws instantly based on how likely they are to be attacked and their potential damage to the business. Cross-tool implementation: Tenable VPR, Microsoft Defender, and CrowdStrike Falcon.
 Digital Assistant for Analysts	- Condenses complex security alerts into clear summaries and generates step-by-step instructions to fix the issue. Cross-tool implementation: Microsoft Security Copilot, Google SecOps, and SentinelOne Purple AI.

04

MAIN RISKS IN AI ADOPTION

 No Governance	Launching or incorporating AI capabilities without defined policies, accountability matrices, or alignment with corporate risk tolerances.
 Shadow IT	Employees using unapproved public AI platforms or extensions without security visibility or enterprise control.
 Lack of Monitoring	Blind spots across model inputs, runtimes, and downstream executions that allow silent failures or compromises to go completely unnoticed.
 AI Model Attacks	Direct adversarial exploitation of the systems, including complex prompt injection, data poisoning, and model extraction.
 Cloud Misconfigurations	- Public buckets - Weak IAM roles - Exposed AI APIs

06

KEY MODEL TYPES

 Machine Learning	- Predicts which vulnerabilities attackers may exploit. - Useful for risk scoring and prioritisation.
 Deep Learning	Detects complex malware, network traffic patterns, and abnormal behaviour.
 Knowledge Graphs	- Connect users, assets, permissions, and vulnerabilities. - Useful for attack path mapping.
 Generative AI	Explains alerts, summarises incidents, and guides remediation steps.

03

WHAT NEEDS PROTECTION

 AI Tools: ChatGPT, Gemini, Microsoft Copilot, Claude	
 AI Models: GPT, Claude, Llama, Mistral	
 AI Infrastructure: AWS Bedrock, Azure AI Foundry, Google Vertex AI	
 AI Data Layer & Enterprise Context: Vector databases, training data lakes, cloud storage repositories, retrieval pipelines, and secure API gateways.	

05

3 PROTECTION LAYERS

ACTIONS: - Validate prompts and inputs - Monitor model outputs - Detect abnormal AI behaviour	TOOLS: Lakera, Protect AI, Robust Intelligence
ACTIONS: - Scan cloud AI services - Control IAM permissions - Protect APIs and storage	TOOLS: Tenable AI-SPM, Wiz, Prisma Cloud
ACTIONS: - Define acceptable use policies - Monitor public AI tool usage - Apply DLP for sensitive data	TOOLS: Netskope, Zscaler, Microsoft Purview

07

AI RISK MANAGEMENT FRAMEWORKS

 Govern Set AI policies, corporate roles, and risk tolerance.
 Map Inventory AI tools, models, internal data repositories, and cloud integrations.
 Measure Evaluate vulnerabilities, configurations, and data access risks.
 Manage Remediate exposures, enforce active controls, and monitor continuously.
LEADING INDUSTRY STANDARDS: - NIST AI Risk Management Framework (AI RMF) - ISO/IEC 42001 (Artificial Intelligence Management System) - OWASP Top 10 for Large Language Model Applications

08

CISO CHECKLIST FOR AI SECURITY TOOLS

Does it scale to enterprise telemetry limits? The solution must ingest massive data streams without bottlenecking live, mission-critical infrastructure.	Can it explain findings clearly? Technical analysts and compliance regulators alike must be able to trace how a specific risk score or automated block was determined.	Does it integrate cleanly into existing workflow playbooks? It should seamlessly enrich your current SIEM/SOAR setups rather than creating a fragmented, stand-alone management silo.	Is it resilient against adversarial prompt injection? The defence tool itself must be completely hardened against attempts to trick, bypass, or blind its internal detection logic.
--	--	---	---

FINAL TAKEAWAY

AI can heavily strengthen cybersecurity, but only if the protected AI stack itself is secured, governed, and continuously monitored.