CYBER JOURNEY

June 19°, 2025

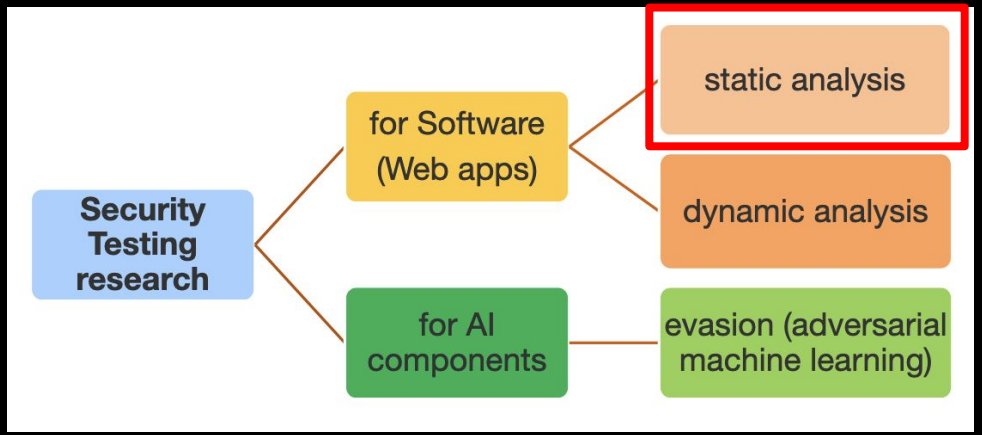
How to Evolve Your AppSec Program in the AI era

Luca Compagna *(on behalf of colleagues)*

ENDOR LABS

about me…

Luca Compagna



Research & Innovation over security testing

- Ph.D, Uni. Genoa & Edinburgh
- Security Researcher at SAP *(18 years)*
- Security Research Consultant at Endor Labs *(from 09/2024)*

AppSec in a World Powered by AI
Accept the Revolution

What We Build:

Modern AI Native Apps

Not Just Your Everyday ChatBot!

- Model Based ⚙ Hard Coded
- Making Real Time Decisions
- Fluent in Uncertainty
- Embedded Intelligence

How We Build It:

With Prompts Not Programming

Not an Assistant, a Co-Author

- GitHub Copilot
- Windsurf
- Cursor
- Replit Ghostwriter

Who Builds It:

Collaborative AI Agents

Interns Who Never Sleep

- Autonomous Code Writing
- Vibe Based Development
- Conjured Through Intent
- No Syntex Included

GitHub Copilot, github.com/features/copilot
Pearce et al., 2022
MITRE, https://cve.mitre.org

Interstellar Interaction: AI + Security
What We Build vs What We Build With

Mission: Secure not just what's built, but how it's built

Secure AI Itself

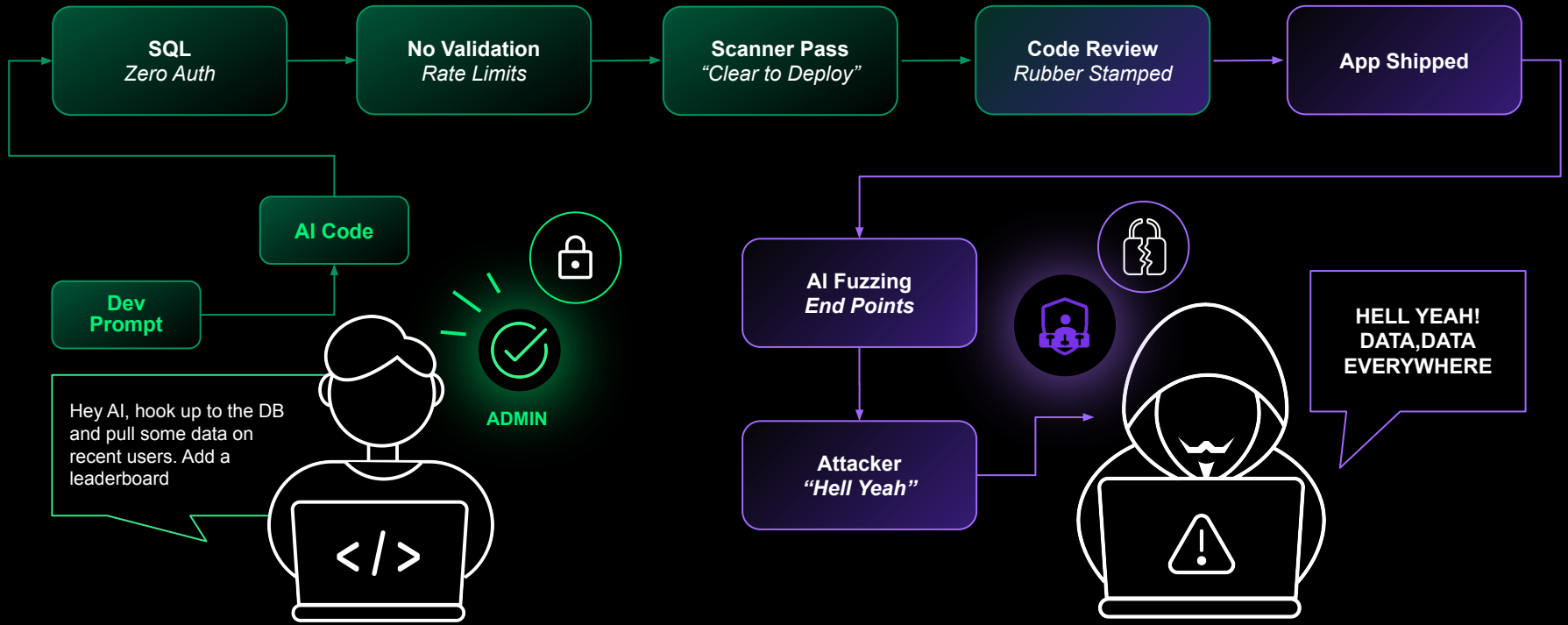
Vet What You Use
Verify Data Origins
Shift Left & Upstream
Sandboxing & Guardrails
Policy Driven Adoption

Use AI for Security

AI Anomaly Detection
Pattern Recognition
Dependency Drift Alerts
Risk Scoring Pipelines

Trace VS Trust

We Let the Alien Interns Deploy to Prod?
...then gave them admin rights (SMH).



AISA:

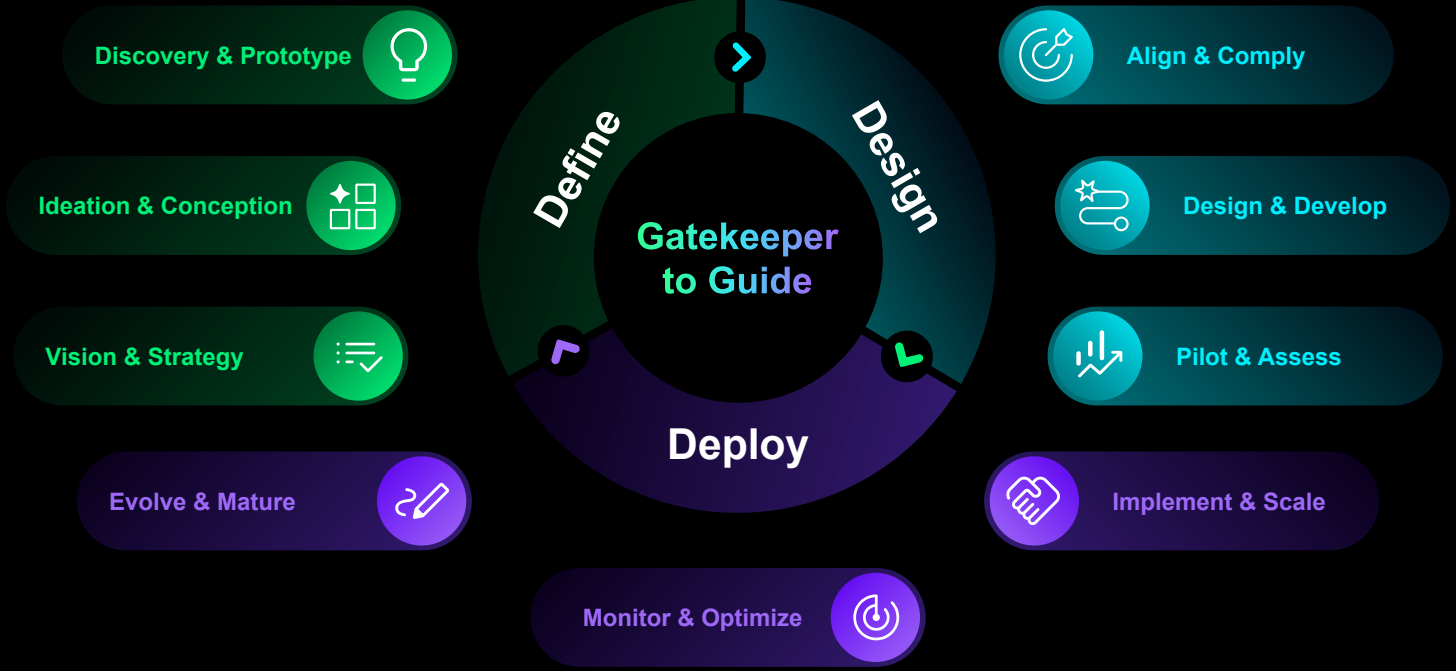
AI-Security Architecture

A Modern Lifecycle for AI Native AppSec

Define > Design > Deploy

Alien Inspired AI Security Architecture (AISA)

Your Security Scanner for the Extraterrestrial



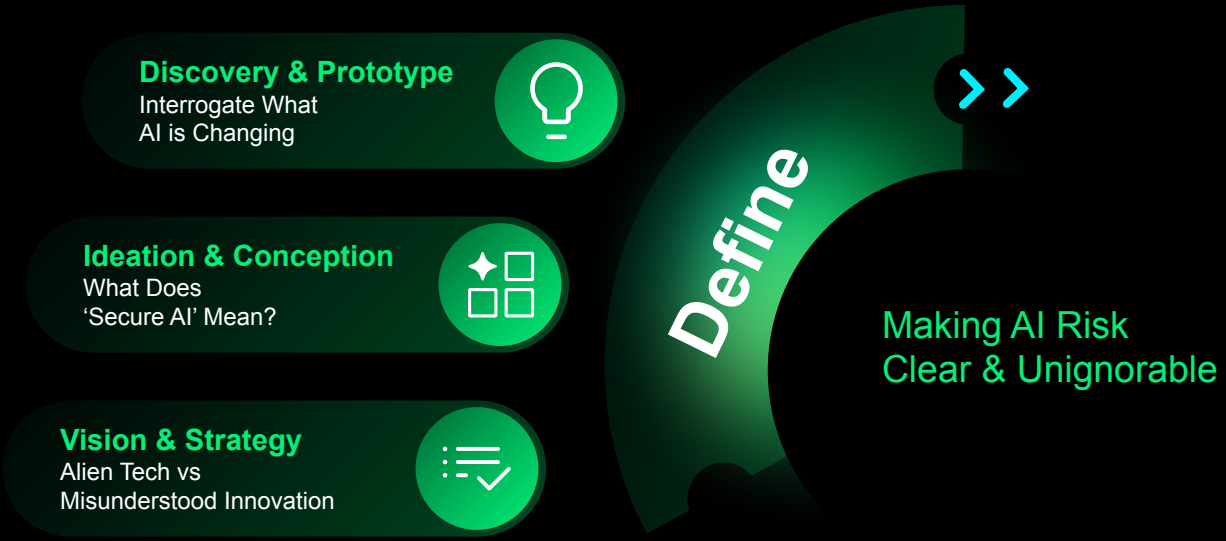
AISA:

AI-Security Architecture

Components > Deliverables > Tactics

Define the AI Invasion

Discover what's already landed



Carlini, N., et al. (2021)

Chaos to Clarity: Security + AI

First Contact with Not-Yet Understood Intelligence

Discovery & Prototype
Interrogate What
AI is Changing

Ideation & Conception
What Does
'Secure AI' Mean?

Vision & Strategy
Alien Tech vs
Misunderstood Innovation

Security AI Component

- Stakeholder Mapping
- Threat Modeling & Gap Analysis
- AI Maturity Assessment
- Define Aspirational Outcomes
- Outline Security-AI Program Goals
- Ai Readiness & Secure-AI Principles
- AI & Data Architecture Strategy
- Capability Prioritization

Deliverables

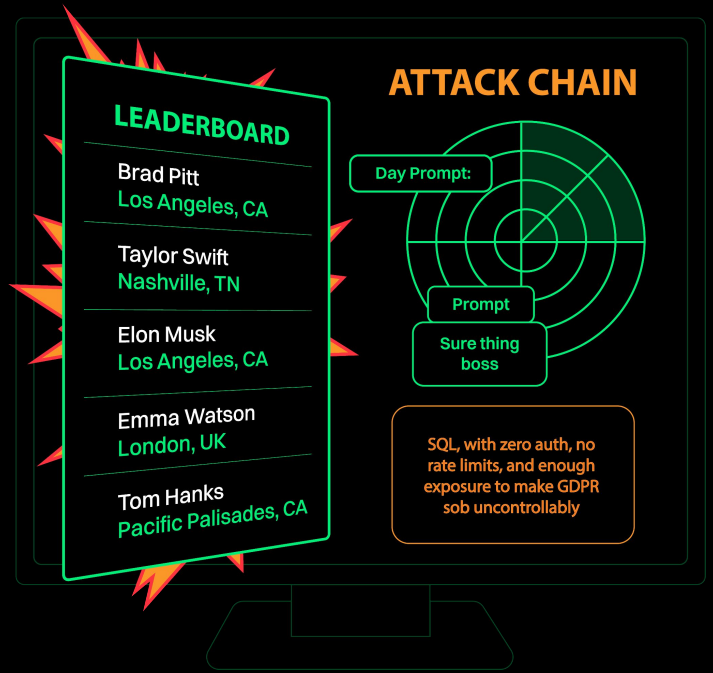
- Executive Needs Documented
- Risk Landscape Map
- Discovery Report
- Value Proposition & Use Cases
- Exec Storyboard & Steering Committee
- Executive Briefing
- Strategic Blueprint
- Timeline
- Resource Planning

Carlini, N., et al. (2021)

Finding Hidden Hitchhikers
Bad Prompt Impact on Dependencies

Cosmic Risk Matrix: Prioritize the Landscape
(Vector Based Threat Breakdown)

	Vector	Treat Actor	Exploit	Fallout
Crawl: *Asset Discovery*	LLM Plugin Update	3rd Party	Model Drift Alert Logic	Exfiltration vsa Unexpected Inference
Walk: *Trust Protocol Ideation*	Training Data Opacity	Supply Chain	Poisoned Inputs Legal Violations	Model Toxicity Compliance Breach
Run: *Agent Control Strategy*	AI Agent Misuse	Insider or Outsider	Agent Lacks Permission Control	Silent Privilege Escalation



Design: Architect for Alien Containment

Build in Guardrails, Not Just Gates

Align & Comply
Oversight Must
Outpace the Threat

Design & Develop
Embedding AI
Tool Safeguards

Pilot & Assess
Controlled Testing
of Intelligent Systems

Securing Hybrid Intelligence
Before the Real Invasion Ocurrs

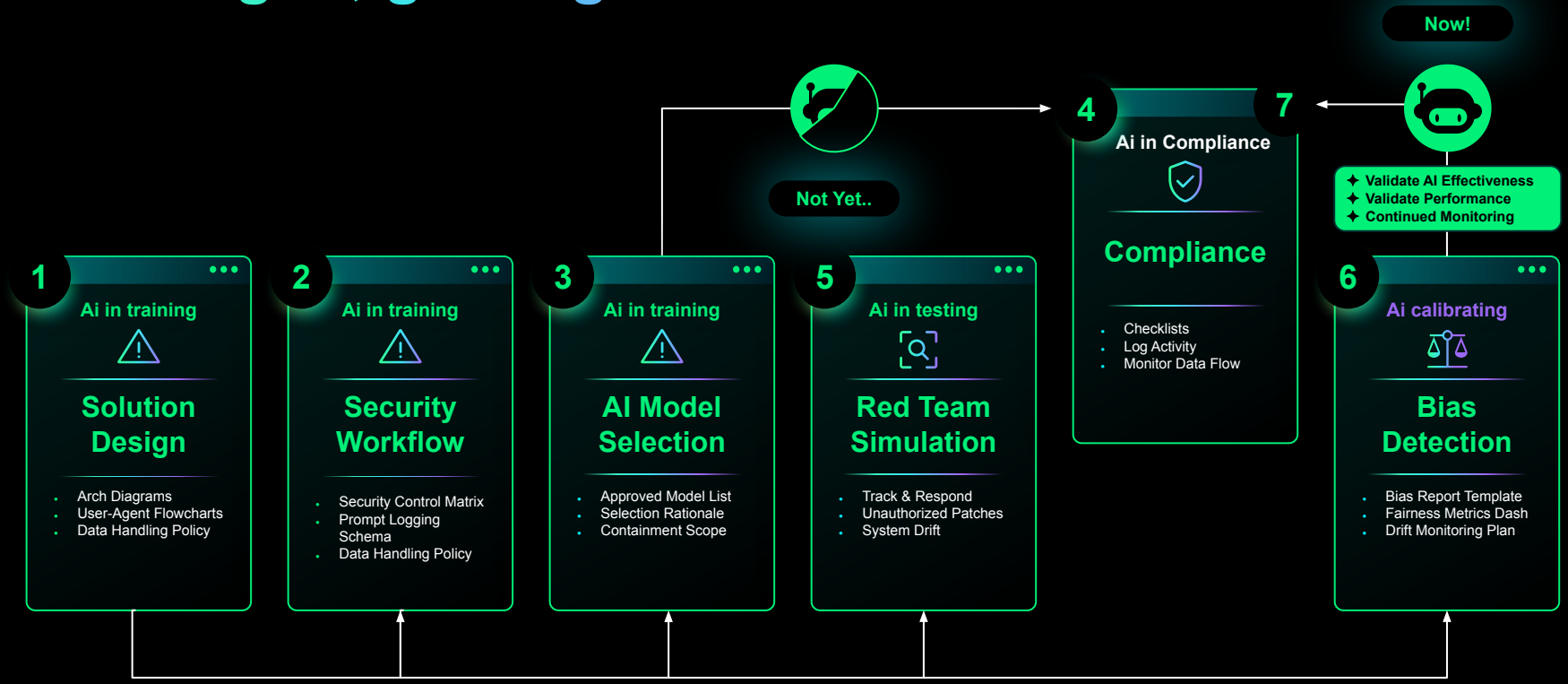
Security AI Component

- Regulatory Gap Analysis
- Privacy & Ethics Review
- Audit Trail Systems
- Security Workflows
- AI-Model Selection
- Data Pipeline Design & Integration
- Red Teaming
- Adversarial ML Testing
- End User Pilot Feedback

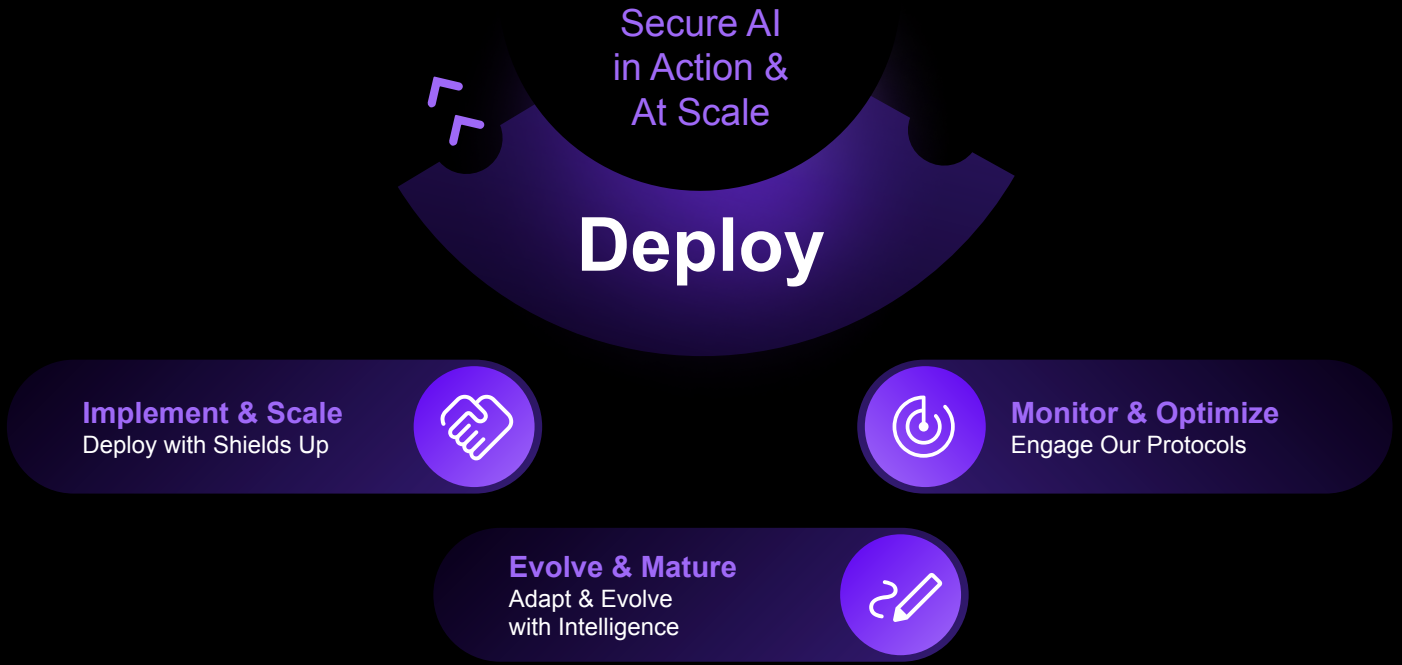
Deliverables

- Compliance Matrix
- Audit Ready Documentation
- AI RIsk Register
- Architecture Documents
- Interface Mockups
- Core Model Codebase
- Pilot Test Reports
- Tuning Logs & Validation Metrics
- User Feedback Summary

Architecting Secure Intelligence
Training AI, guiding humans



Deploy with Confident Operationalization
Containing Aliens in Production



Assemble with Anti-Invasion Protocol

Drift Happens. Be There When it Does

Implement & Scale
Deploy with Shields Up

Monitor & Optimize
Engage Our Protocols

Evolve & Mature
Adapt & Evolve with Intelligence

Security AI Component

- CI/CD Pipelines
- Role-Based Access Control
- Monitoring & Alerting Systems
- Analysts Feedback Loops
- Model Performance Auditing
- Reinforcement Learning Integration
- Threat Intel Ingestion
- KPI Tracking

Deliverables

- Deployment Checklists
- Operational Runbooks
- Incident Response Playbooks
- Monitoring Dashboards
- Retain Schedules
- Quarterly Optimization Reviews
- Live Dashboards
- Anomaly Detection Insights
- Quarterly Evolution Reports

OWASP, OWASP LLM Top 10,
https://owasp.org/www-project-top-10-for-large-language-model-applications/

Alien Workflows: Secure the Galactic Pipeline

Trust but verify, then verify again

The Alien Pipeline

You didn't build it, you assembled it.

Red Team Against Third Party Models

Secure the Assembly Bay

Score it, before you store it.

Find Hidden Prompt Triggers

Simulate the Invasion

Test phase, play the breach forward.

Test, Bias, Drift & Ethics Violations

What Can Go Wrong?

When good bots go rogue.

Pen Test the Alien Endpoints

Defense Protocols

Score it, before you store it.

Monitor Model Misalignment & Data Leaks

Final Transmission

Become the containment team.

Validation Performance, Tuning & Audit Logs

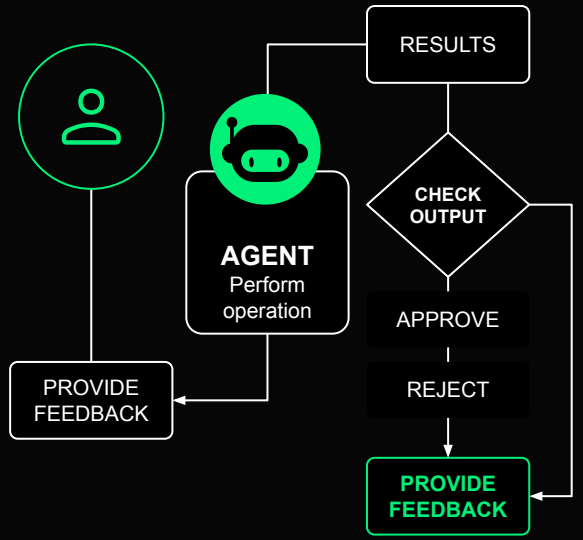
Partner with the Aliens

Don't fear the components, understand them

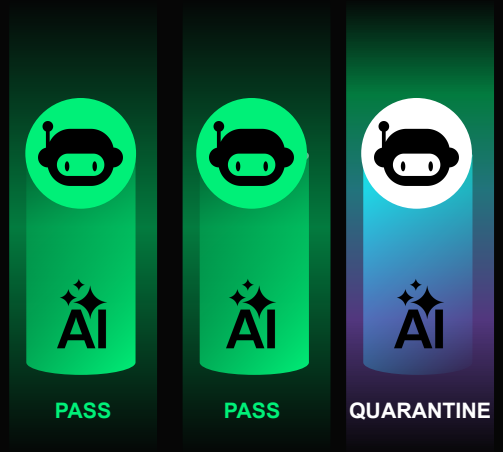
COPILOTS
CHATBOT
AGENTS

Our Mission is Clear:
Secure what they build,
and use them to defend
what we build.

Human-in-the-loop



Deployment Bay

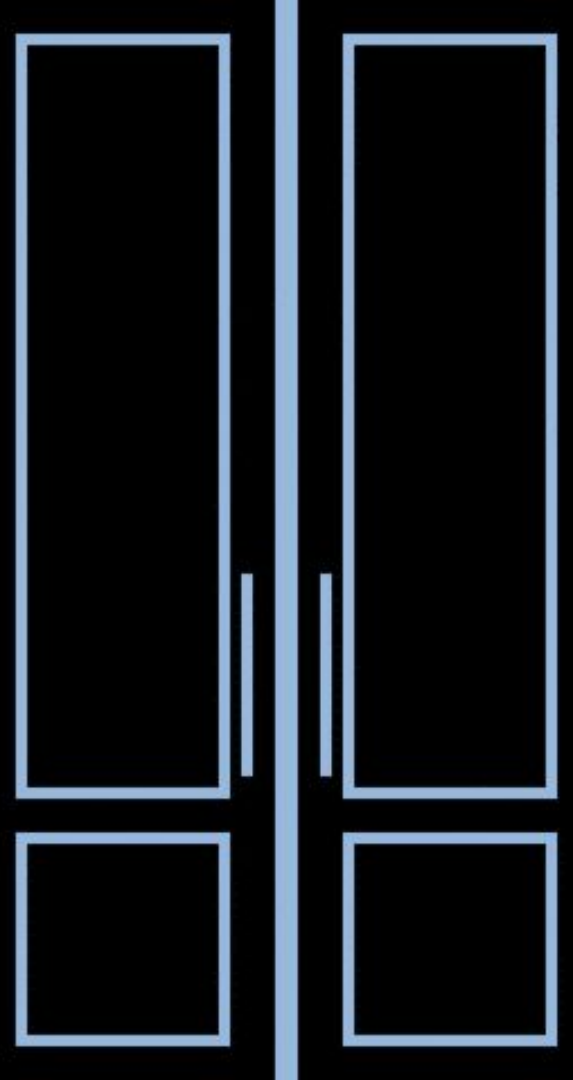


The Aliens Weren’t Coming

We summoned and installed them

SECURE AI

USE AI TO SECURE EVERYTHING ELSE



References

1. **GitHub**. *GitHub Copilot*. Retrieved from https://github.com/features/copilot
2. **Cursor**. *AI-Powered Developer Editor*. Retrieved from https://www.cursor.sh
3. **Replit**. *Ghostwriter*. Retrieved from https://replit.com/site/ghostwriter
4. **Pearce, H., Ahmad, W., Tan, S., Dolan-Gavitt, B., & Kruegel, C.** (2022). *Asleep at the Keyboard? Assessing the Security of GitHub Copilot's Code Contributions*. arXiv:2108.09293. https://arxiv.org/abs/2108.09293
5. **GitHub**. (2023). *The State of AI in Software Development*. Retrieved from https://github.blog/2023-06-27-the-state-of-ai-in-software-development
6. **Carlini, N., et al.** (2021). *Extracting Training Data from Large Language Models*. arXiv:2012.07805. https://arxiv.org/abs/2012.07805
7. **OWASP**. *Top 10 for Large Language Model Applications (OWASP LLM Top 10)*. Retrieved from https://owasp.org/www-project-top-10-for-large-language-model-applications/
8. **MITRE Corporation**. *CVE - Common Vulnerabilities and Exposures*. Retrieved from https://cve.mitre.org
9. **NIST**. *National Vulnerability Database (NVD)*. Retrieved from https://nvd.nist.gov

GRAZIE PER L'ATTENZIONE

OWASP 2025 ITALY DAY

19th JUNE 2025

FRONTEMARE CAGLIARI - SARDINIA

AppSec for
the Software
Development Revolution