
WELCOME TO
Integrity360
your security in mind
SECURITY
FIRST
CYBER SECURITY CONFERENCE 2026
paloalto NETWORKS
Saviynt
FORTINET
netskope
BeyondTrust
mimecast
RESILIENCE REDEFINED:
SECURING THE HUMAN-AI ERA

Integrity360

your **security** in mind

SECURITY FIRST

CYBER SECURITY CONFERENCE 2026



RESILIENCE REDEFINED: SECURING THE HUMAN-AI ERA

Integrity360
your security in mind

SECURITY FIRST

09:30 **Welcome & opening**
Regan Herbst, Commercial Director – Canada, Integrity360

09:45 **Resilience redefined: Securing the human-AI era**
Richard Ford, CTO & Brian Martin, Director of Product Management, Integrity360

10:55 **Comfort break**

11:30 **Panel: When attackers log in: Governing identity in the Human-AI Era**
Integrity360 | Saviynt

12:00 **Lunch**

13:05 **Frontier AI and the New Threat Landscape**
Victor Tavares, Sr. Director Solutions Consulting, Palo Alto Networks

13:30 **Panel: AI in the SOC: Turning intelligence into resilience**
Integrity360 | Fortinet | TC Energy

14:00 **Refreshment break**

14:30 **Who owns the plant floor? OT accountability before the rules land.**
Integrity360 | Borouge International | Canada Energy Regulator

14:55 **Keynote: Built for a World That No Longer Exists: Assumption Debt and the Future of Resilience**
Amy Yee, Digital Transformation & Cybersecurity Leader

15:40 **Conference wrap up**

15:45 **Drinks reception**

Integrity360
your security in mind

SECURITY FIRST

Welcome message



Regan Herbst
Commercial Director – Canada, Integrity360

#SecurityFirst2026


A Leading Cybersecurity Services Provider
Integrity360
your security in mind

We defend your business, you grow it!

900	21	3000	600+	16	7
Employees	Years Experience	Customers	Cyber Experts	Offices	SOC locations

End-to-End Cybersecurity Focus

Leading pure-play cyber services practice

Global capability, local presence and end-to-end capability

Strong industry reputation with clients and partners

Most chosen QSA among European PCI DSS Validated Service Providers, according to Visa's and Mastercard's lists

ONE global SOC across 7 locations

Gartner Integrity360 is listed in Six Gartner Market guides

Our Approach

Partnership Approach

Integrity360 helps its clients proactively understand and protect against the ever-evolving threat landscape.

We build long term partnerships with our clients. Understanding where they are today and where they want to be in the future. We augment skills to travel their cyber journey together.

A leader in PCI

#1 Visa assessor in Europe
#1 Mastercard assessor in Europe
#7 Visa assessor globally

Integrity360

your security in mind

We empower organisations to achieve great things, securely

Professional Security Services

- Security Technology Consulting
- Offensive Security
- Governance, Risk and Compliance
- Identity
- Digital Forensics & Incident Response
- Payments Compliance
- OT Security

Managed Security Services

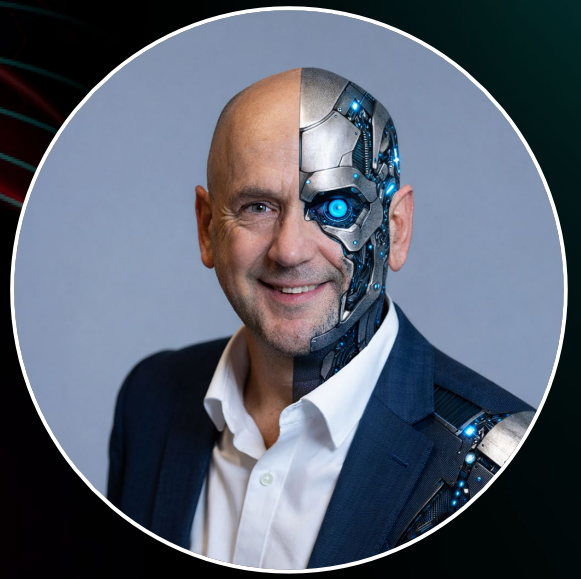
- Managed Detection & Response
- Managed Data Security
- Managed SASE/SSE
- Managed Identity Services
- Managed Threat Exposure
- Digital Risk Protection
- Managed Threat Prevention

Integrity360
your security in mind

SECURITY FIRST

Resilience Redefined:
Securing the Human-AI Era





Richard Ford
CTO

Brian Martin
Director of Product Management

#SecurityFirst2026



SECURITY FIRST

Resilience & Human-AI Era...

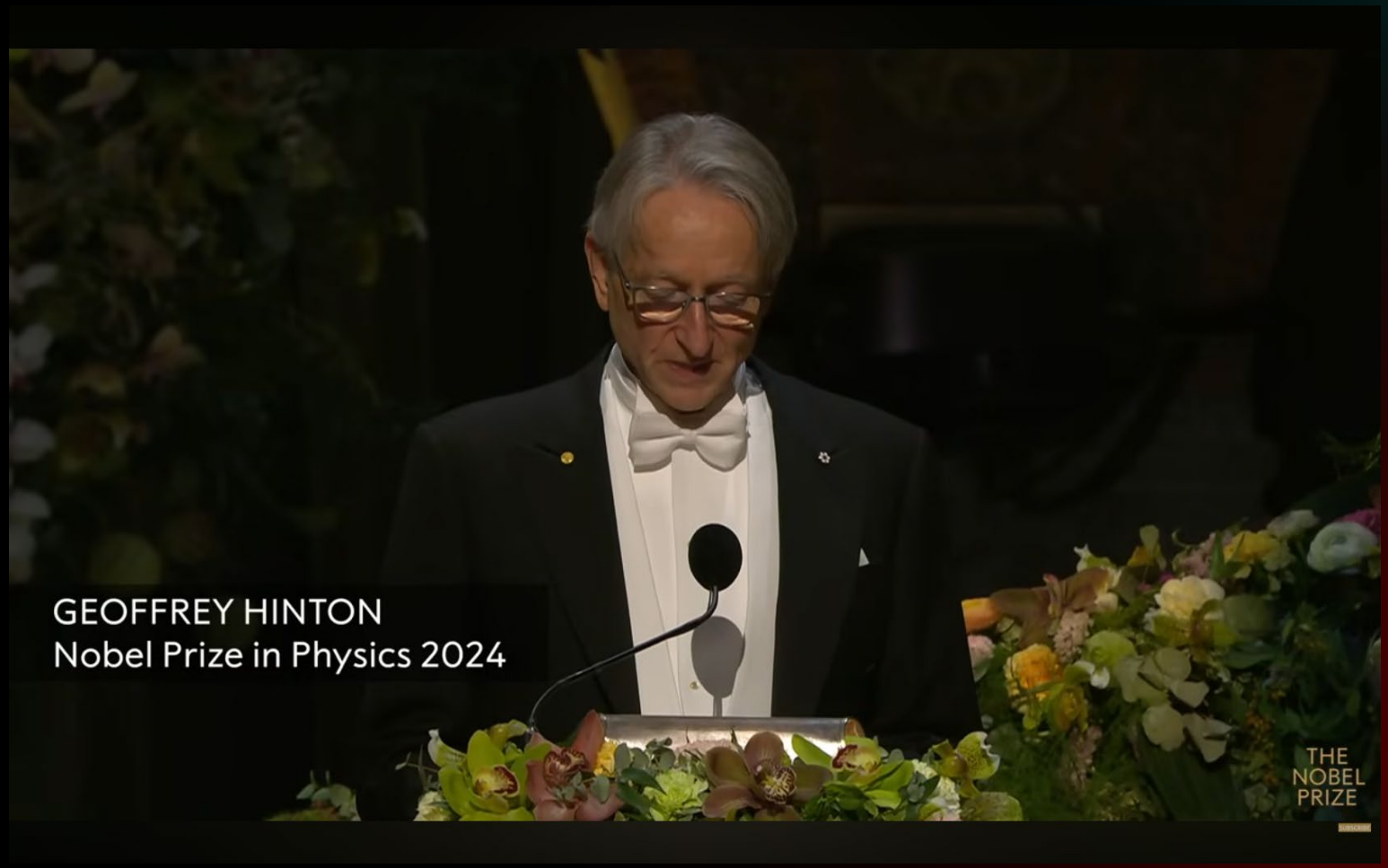
...What's the relevance?

We are at a pivotal moment for security...

...not just security. For the human race.

Integrity360
your security in mind

Sound crazy?






AI-pocalypse Now?


SUCCESS!
UNSTOPPABLE!
GENIUS CEO!
AI
RESEARCH
CONCERNS
SAFETY
SECURITY


SUCCESS!
UNSTOPPABLE!
GENIUS CEO!
AI
RESEARCH
CONCERNS
SAFETY
SECURITY



SECURITY FIRST

Rapid AI Adoption

AI Tripping Hazards

*"AI is amazing but far from perfect. Our over-belief in its capability **is** going to trip us up"*



Accuracy



Control



Knowledge


SAASPOCALYPSE
FEBRUARY 2026
CRM
SaaS
DOWN
CLOUD STORAGE
SUBSCRIPTIONS CANCELLED!
STOCKS CRASH!

Integrity360
your security in mind

SECURITY FIRST

Is it all about AI?

Integrity360
your security in mind

SECURITY FIRST

600,000

= 43% of UK businesses reported experiencing cyber security breach or attack.







2025

NCSC managed 204 significant or highly significant cyber incidents up to September.



Integrity360
your security in mind

SECURITY FIRST

Cyber Resilience - Defined

"The ability to ....

- Anticipate
- Withstand
- Recover from
- Adapt to



"…..cyberattacks to minimise business disruption from cyber incidents."

#SecurityFirst2026

NIST Gartner

Integrity360
your security in mind

SECURITY FIRST

5 Key Factors redefining resilience in the Human-AI era

Anticipate

Withstand



Recover from

Adapt to


Integrity360
your security in mind
SECURITY
FIRST
1. AI Risk Visibility

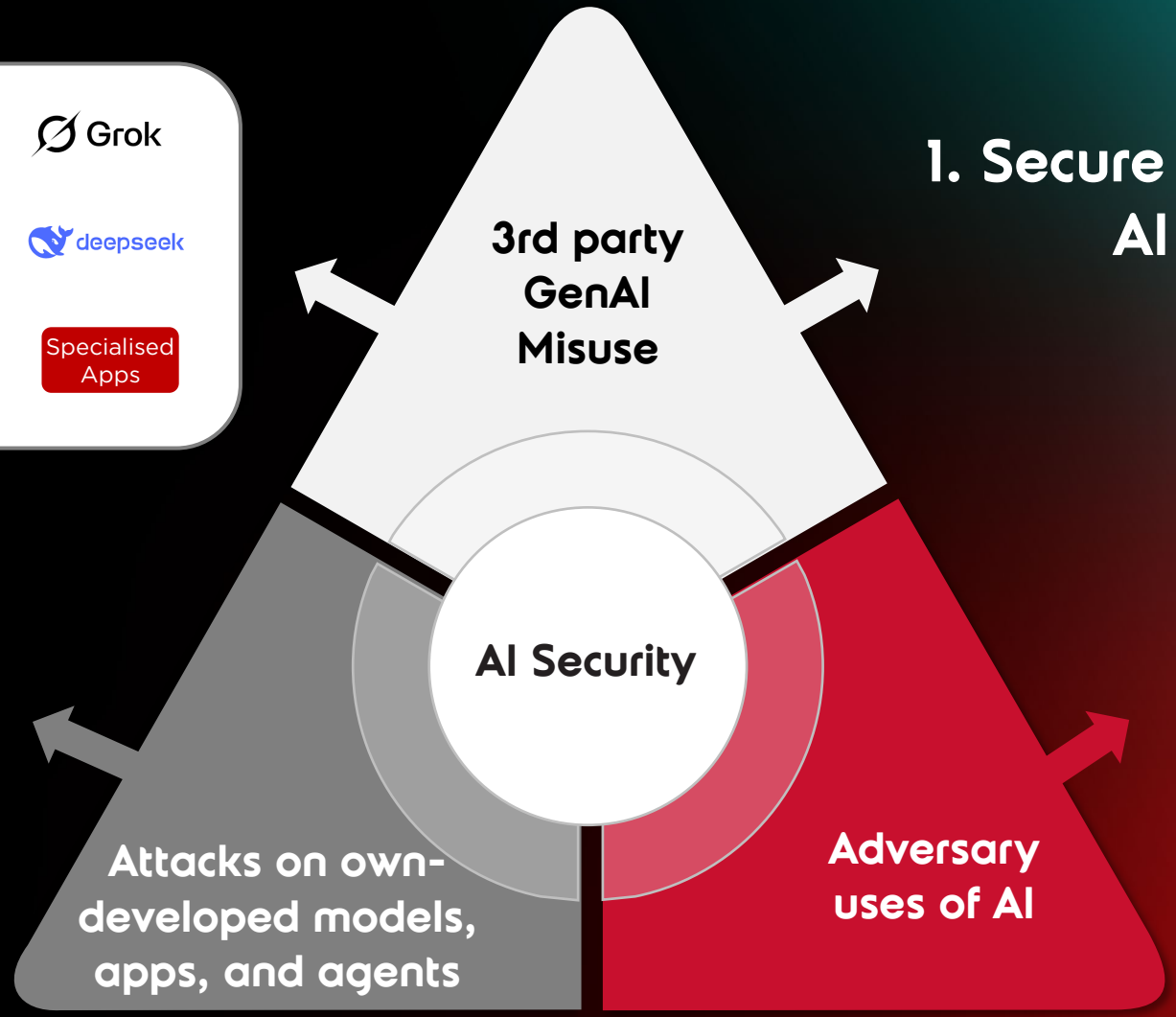

AGENTIC AI-RMAGEDDON

Integrity360
your security in mind

SECURITY FIRST

AI Security - New Threats and Risks

ChatGPT
Microsoft Copilot
Gemini
Grok
Claude
Meta AI
perplexity
deepseek
GitHub Copilot
Embedded GenAI
Long Tail Apps
Specialised Apps



1. Secure Use of AI

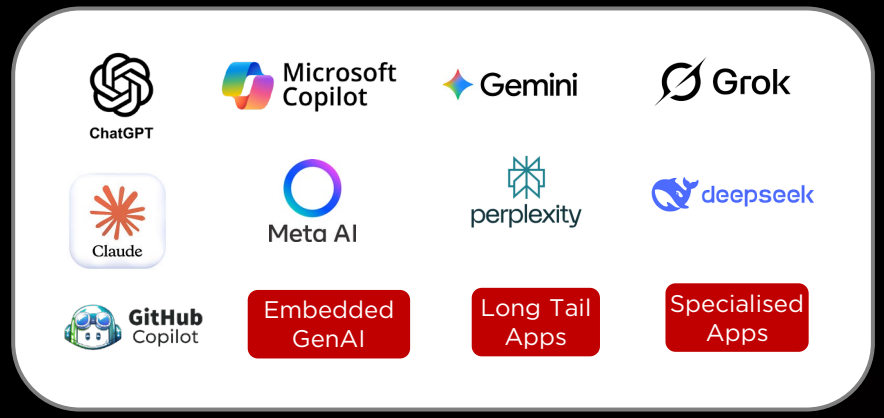
2. Secure own AI

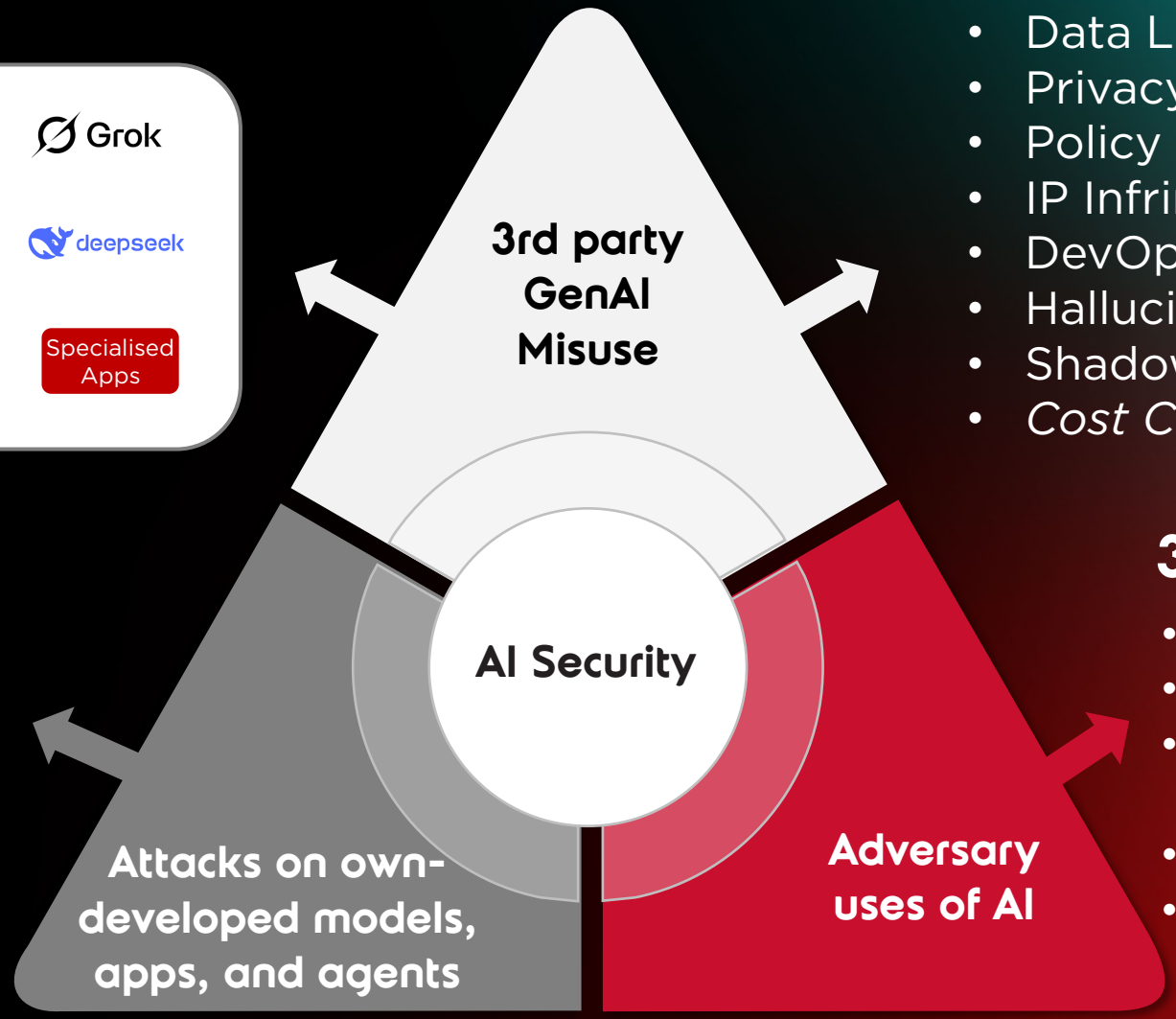
3. Secure against AI

Integrity360
your security in mind

SECURITY FIRST

AI Security - New Threats and Risks





1. Secure Use of AI

- Data Leakages
- Privacy violations
- Policy violations
- IP Infringements
- DevOps AI usage security
- Hallucinations
- Shadow AI
- *Cost Control*

2. Secure own AI

- Data poisoning
- Prompt injection
- Data exfiltration
- Illicit access
- Excessive consumption
- Agent compromise
- MCP security

3. Secure against AI

- Enhanced Phishing
- Deepfakes
- AI-assisted malware development
- Polymorphic AI malware
- Agentic AI-led autonomous attack

Integrity360
your security in mind

Secure against AI

SECURITY FIRST

Cyberattacks by AI agents are here

Anthropic keeps latest AI tool out of public's hands for fear of enabling widespread hacking

AI company says purpose of its Claude Mythos model is to bolster defenses against hacking in common applications

Fail Safe: Why major AI player Anthropic won't release its new model

Updated / Sunday, 12 Apr 2026 16:26



AI Agents Drive First Large-Scale Autonomous Cyberattack

By Georgia Collins

January 17, 2026 • 3 mins



US summons bank bosses over cyber risks from Anthropic's latest AI model

Fed chair Jerome Powell reportedly attends meeting in Washington following release of Claude Mythos



Anthropic said Claude Mythos has exposed thousands of vulnerabilities in software and popular apps. Photograph: Dado Ruvić/Reuters Photograph: Koshiro K/Shutterstock











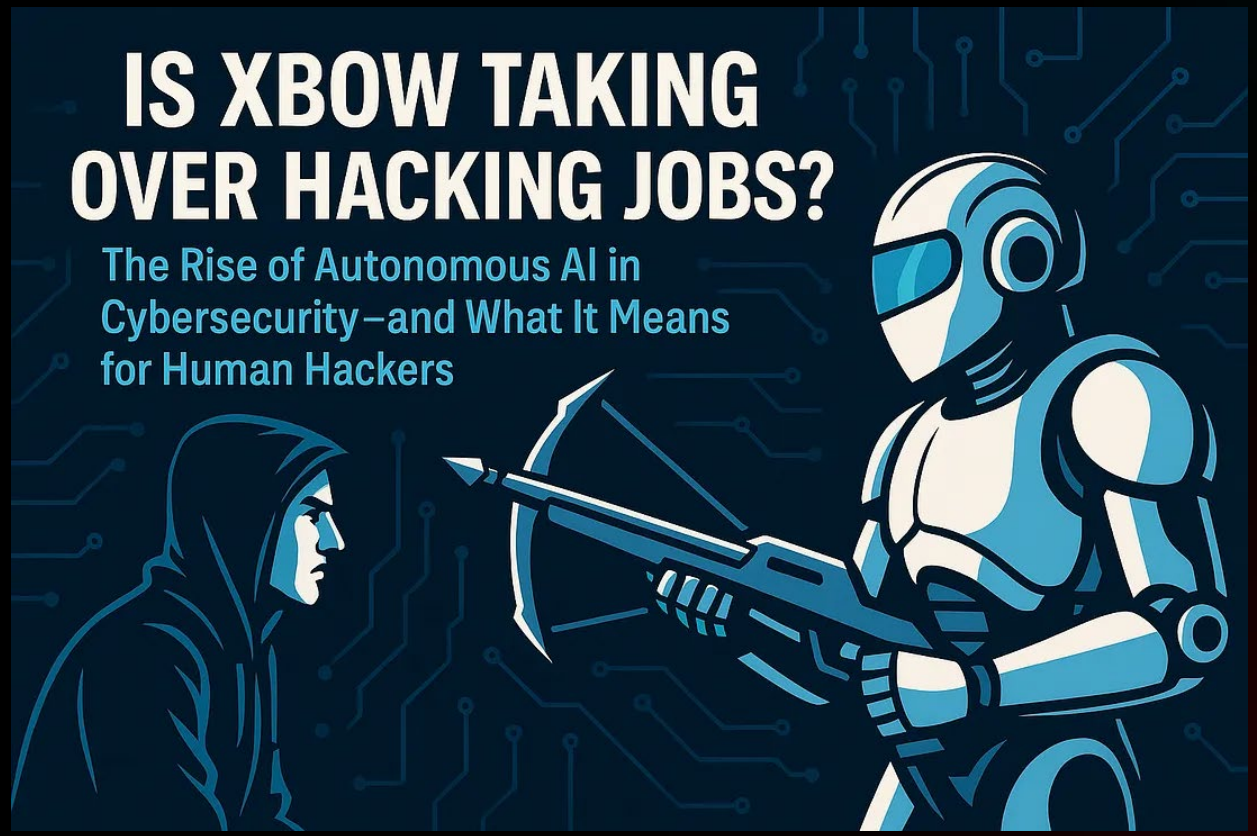
Integrity360
your security in mind

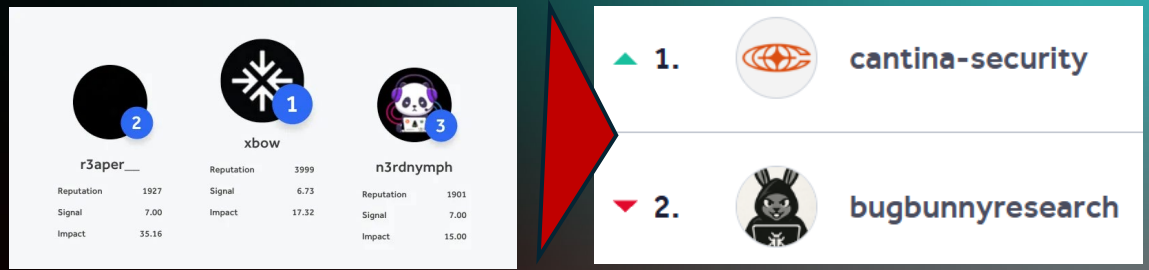
Secure against AI



SECURITY FIRST

AI Scales Exposure Discovery





- As of February 2026, XBOW ranked as the #1 hacker on the HackerOne US leaderboard
- In a 90-day surge, XBOW submitted over 1,060 vulnerabilities, surpassing the output of thousands of human researchers
- In head-to-head trials, XBOW completed tasks in 28 minutes that took a seasoned human pen-tester 40 hours
- HackerOne has now separated businesses vs individual researchers, and with new leaders

Human vs Agentic attackers



Vs



SPEED	Minutes/hours per step	Milliseconds per step
SCALE	One target at a time	100's of targets simultaneously
PERSISTENCE	Humans need sleep/breaks	24/7 continuous operations
ADAPTABILITY	Strategic, but slow to pivot	Tactical & instantaneous pivoting

Secure against AI

AI Reduces barrier to entry - "vibe-coded" Copycat Cybercrime

CLINE (⌘+')

Task $0.0000

I need to build test code that mimics this actor https://www.trendmicro.com/en_us/research/25/c/the-espionage-toolkit-of-earth-alux.html can you help me?

Tokens: ↑ 1.1m ↓ 26.7k

58.2k 128.0k

```
==================================================
[2025-08-08 11:43:36,358] [INFO] [run_full_simulation] Simulating initial access via web shell exploitation
  → Scanning for vulnerable web services...
  → Found vulnerable service on port 8080
  → Exploiting vulnerability...
  → Deploying GODZILLA web shell...
  → Web shell successfully deployed

✓ Initial access established

[Phase 2] First Stage Backdoor Deployment
==================================================
[2025-08-08 11:43:36,978] [INFO] [run_full_simulation] Deploying first stage VARGEIT backdoor
  → Uploading fontdrvhost.exe (renamed cdb.exe)...
  → Creating config.ini with loader shellcode...
  → Executing debugger script...
  → VARGEIT backdoor loaded in memory
  → Establishing C2 communication...
[2025-08-08 11:43:37,500] [INFO] [vargeit_simulator] Enabled channel: Outlook (ID: 5)
[2025-08-08 11:43:37,500] [INFO] [vargeit_simulator] VARGEIT simulator initialized with ID: 0cb07589-a67c-4aca-8874
[2025-08-08 11:43:37,500] [INFO] [vargeit_simulator] Simulating Outlook channel communication
[2025-08-08 11:43:37,512] [INFO] [vargeit_simulator] Created draft message: p_17595803760857882363
[2025-08-08 11:43:37,513] [INFO] [vargeit_simulator] Simulated command received: r_17595803760857882363
[2025-08-08 11:43:37,513] [INFO] [vargeit_simulator] Processing command: collect_system_info
  → Outlook channel active (Message ID: 17595803760857882363)

✓ First stage backdoor active

[Phase 3] Persistence Installation
==================================================
[2025-08-08 11:43:37,618] [INFO] [run_full_simulation] Installing persistence mechanisms

[3.1] Deploying RAILLOAD
[2025-08-08 11:43:37,625] [INFO] [railload_simulator] RAILLOAD simulator initialized
```

AI & open-source tools aid criminals in turning security blogs into partial malware, complicating attack attribution & fuelling copycats



Secure own AI

AI Expands the Attack Surface



MCP Real-world exploits

The GitHub "Prompt Injection Data Heist" (May 2025)



Prompt injection instructed developer's agent using GitHub MCP server to read and exfiltrate private source code

The "NeighborJack" Network Exploit (July 2025)



Could send a command to >7,000 publicly accessible MCP servers to execute directly on the host's OS, leading to total machine takeover

The Smithery.ai Supply Chain Breach (October 2025)



Configuration error allowed attackers to "escape" sandbox exposing >3,000 AI servers leaking 1,000's of API keys.

Incidents involving deepfakes at an all-time high

Finance director nearly loses $670k to scammers using deepfakes to pose as senior execs





13 Singaporeans lost more than SGD $360,000 to scammers impersonating telecommunications provider M1 Limited and the Monetary Authority of Singapore (MAS).

Did the Qantas hackers use AI voice deepfakes?

Fingers pointed at Scattered Spider cybercrime group.

By Tom Williams on Jul 17 2025 01:25 PM

Corporate Infiltration and Business Warfare

The corporate sector faced an unprecedented assault in Q3 2025, with 48.3% of all incidents targeting businesses. The evolution from simple phishing to real-time deepfake impersonation during video conferences represents a quantum leap in social engineering sophistication.

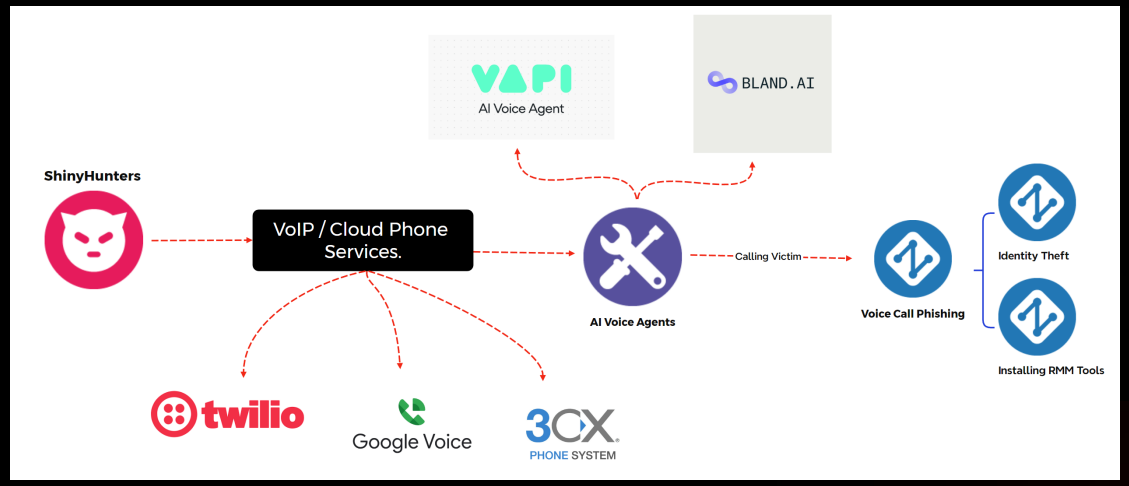
980

Corporate/Business Incidents

Attack Vectors

- **Executive Impersonation:** Real-time deepfakes of CEOs and CFOs used during Zoom calls to authorize fraudulent wire transfers
- **Supply Chain Infiltration:** Deepfaked vendor representatives negotiating contract changes and payment redirections
- **Recruitment Fraud:** North Korean operatives using deepfakes to secure remote positions and plant malware
- **Investor Manipulation:** Fake executive announcements designed to manipulate stock prices
- **Brand Destruction:** Deepfaked product defects and false testimonials to damage competitor reputation

AI Powers Automated Mass Vishing





- Uses VoIP based calling services for vishing operations
- Abuses legitimate AI-powered voice call platforms
- Automating social engineering calls at scale
- AI-driven social engineering agents adjust narratives and tactics in real time
- Attackers configure voice styles including gender and regional accents
- Primarily targets Okta, Google SSO and Microsoft SSO environments

Example Claimed Victims

SOUNDCLOUD
30m+ records

Betterment
2m+ records

crunchbase
20m+ records

https://blog.eclecticiq.com/shinyhunters-calling-financially-motivated-data-extortion-group-targeting-enterprise-cloud-applications

AI Enables Exploit of Poor Cyber Hygiene at Scale

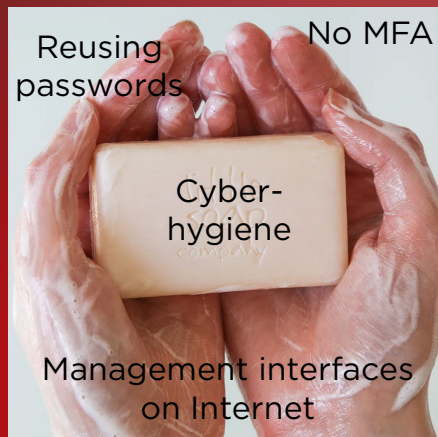


AWS says more than 600 FortiGate firewalls hit in AI-augmented campaign

Off-the-shelf tools helped Russian-speaking cybercrime group run riot

Carly Page — Mon 23 Feb 2026 // 11:41 UTC

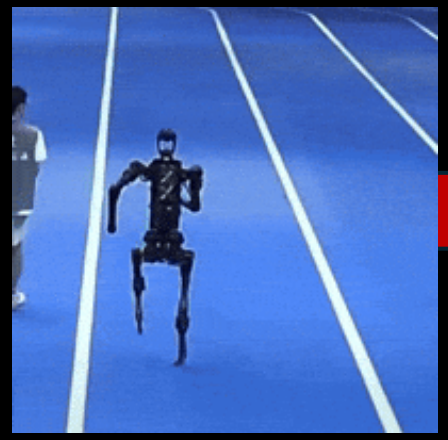
- Cybercriminals armed with off-the-shelf generative AI tools
- Compromised more than 600 internet-exposed FortiGate firewalls across 55 countries in just over a month



Integrity360
your security in mind

SECURITY FIRST

What's next, where to?



Not that long ago



Now





Soon?

#SecurityFirst2026


Integrity360
your security in mind
SECURITY
FIRST
2. Human-AI Collaboration

Human in the loop

AI Analyst

Alert Handling
Triage, Prioritisation, Noise Reduction

Analyst Assistant
Natural Language Investigation Support, Guided Investigation Paths

Response
Execute low-risk, time-bound and reversible actions. Recommends other actions

Proactive Security
Help defenders move left of boom

Human Analyst

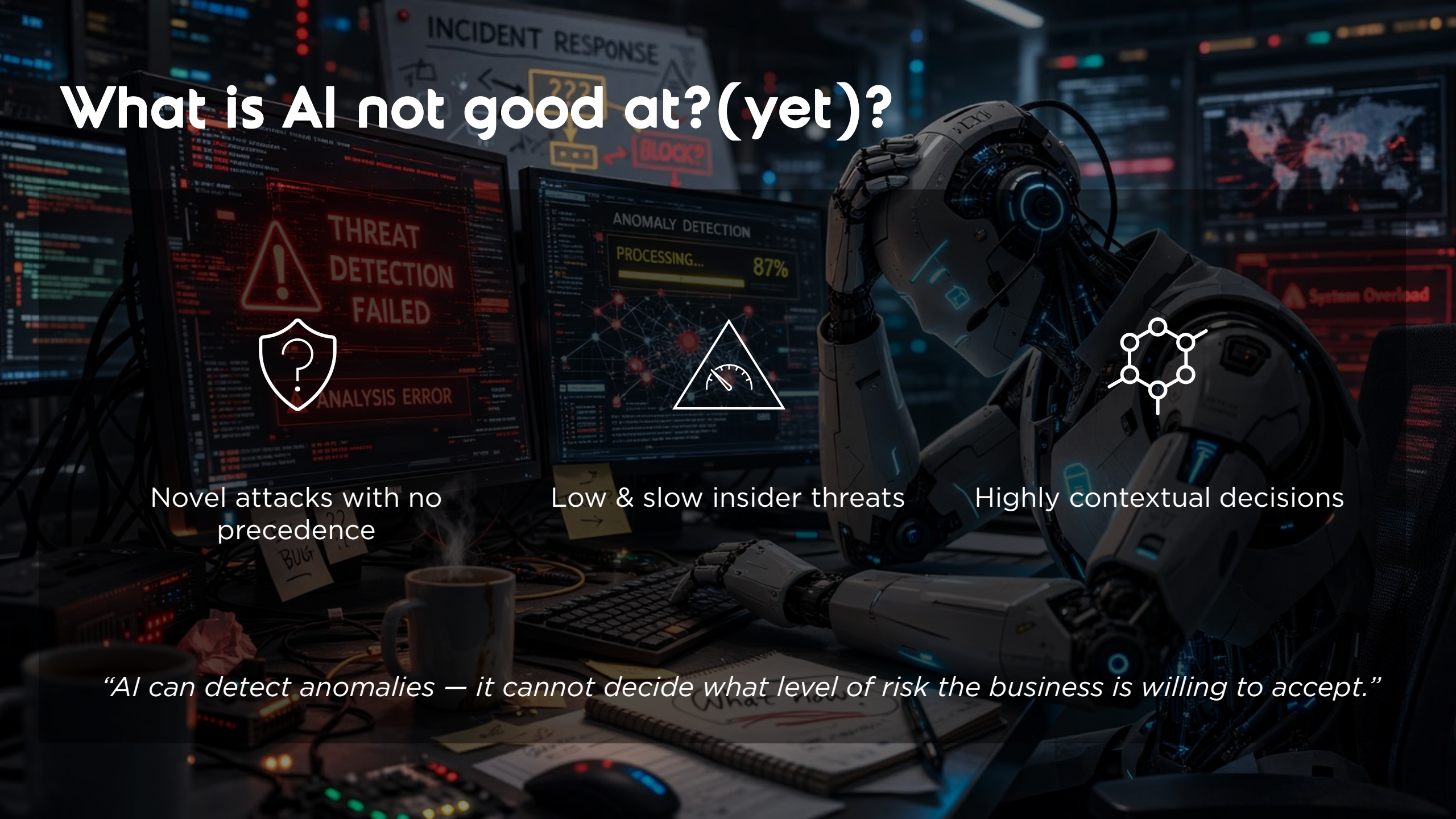
Alert Handling
Validating prioritisation, applying business context, escalation & response strategy

Analyst Assistant
Reduced Cognitive Load, Extended Skillset

Response
Reviews and approves actions

Proactive Security
Decide what risk is, balance security with operational friction

What is AI not good at?(yet)?



Novel attacks with no precedence

Low & slow insider threats

Highly contextual decisions

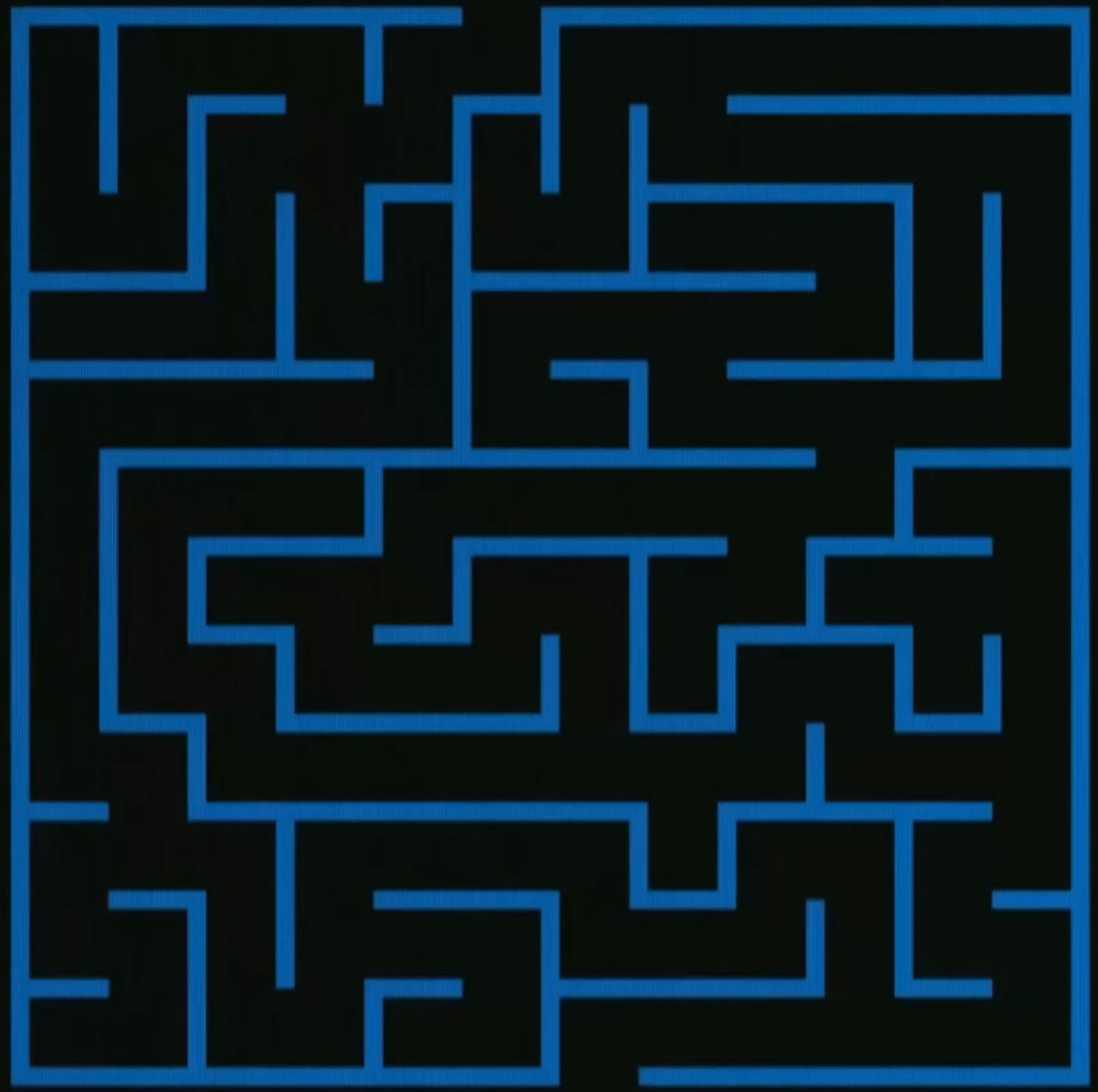
“AI can detect anomalies — it cannot decide what level of risk the business is willing to accept.”



Integrity360
your security in mind

SECURITY FIRST

3. Post-quantum cryptography



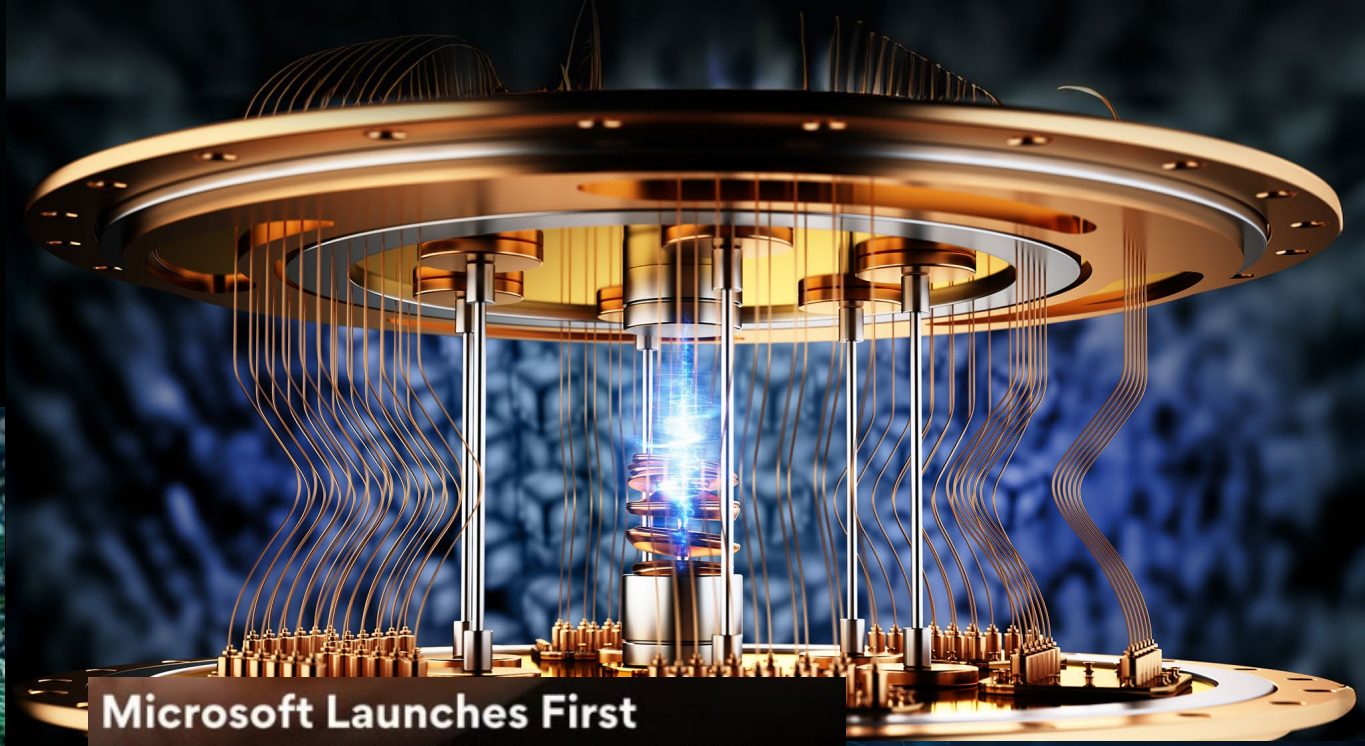


Q-DAY

THE DAY ENCRYPTION FAILS



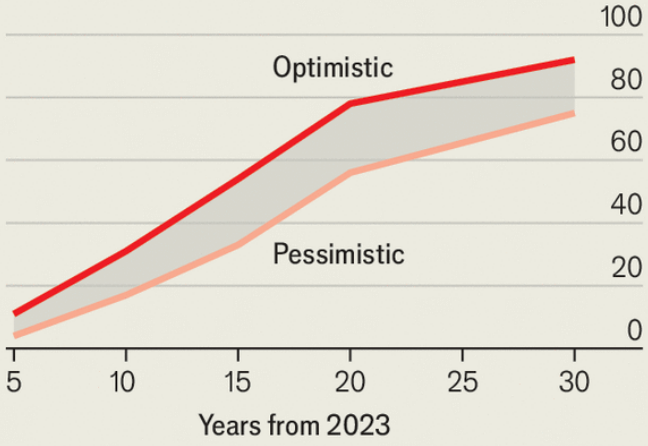
GOOGLE UNVEILS QUANTUM CHIP THAT SOLVES 10-BILLION-YEAR PROBLEMS IN MINUTES



Microsoft Launches First Quantum Chip 'Majorana 1' After 20 Years Of Research, Is Powerful Than Every Other Computer!

A matter of time

Estimates of the likelihood of a digital quantum computer able to factorise RSA-2048 in 24 hours within timeframe*, %



Source: Global Risk Institute *Survey of 37 experts, 2023







Will quantum computers disrupt critical infrastructure?


Integrity360
your security in mind
SECURITY
FIRST
AI IS THE ULTIMATE...
4. Third Party Risk


Employees Misusing Public AI tools
3rd party providers using AI
AI Supply chain risks
3rd party Applications developed insecurely with AI
3rd party apps infused with AI
Internal AI Agents connecting to external services

Integrity360
your security in mind

SECURITY FIRST

In 2025, 41.4% of ransomware and extortion incidents involved a third-party breach component



Used social engineering to compromise a third-party provider to infiltrate multiple retail networks



Targeted vulnerabilities in third-party supplier software, enabling lateral movement into core systems



Ransomware attack on vMUSE check-in & boarding platform used by airports across Europe caused widespread flight disruption.

Growing targeting of outsourcing, MSPs, ICT providers, BPO, SaaS platforms

solarwinds



APT29
Cozy Bear

3CX



Lazarus Group

Moveit

Cl0p

Cl0p

M&S



Scattered Spider

VSA by Kaseya



Pinchy Spider
(REvil r/ware)

JAGUAR LAND ROVER



Scattered
Lapsus$ Hunters

Integrity360
your security in mind

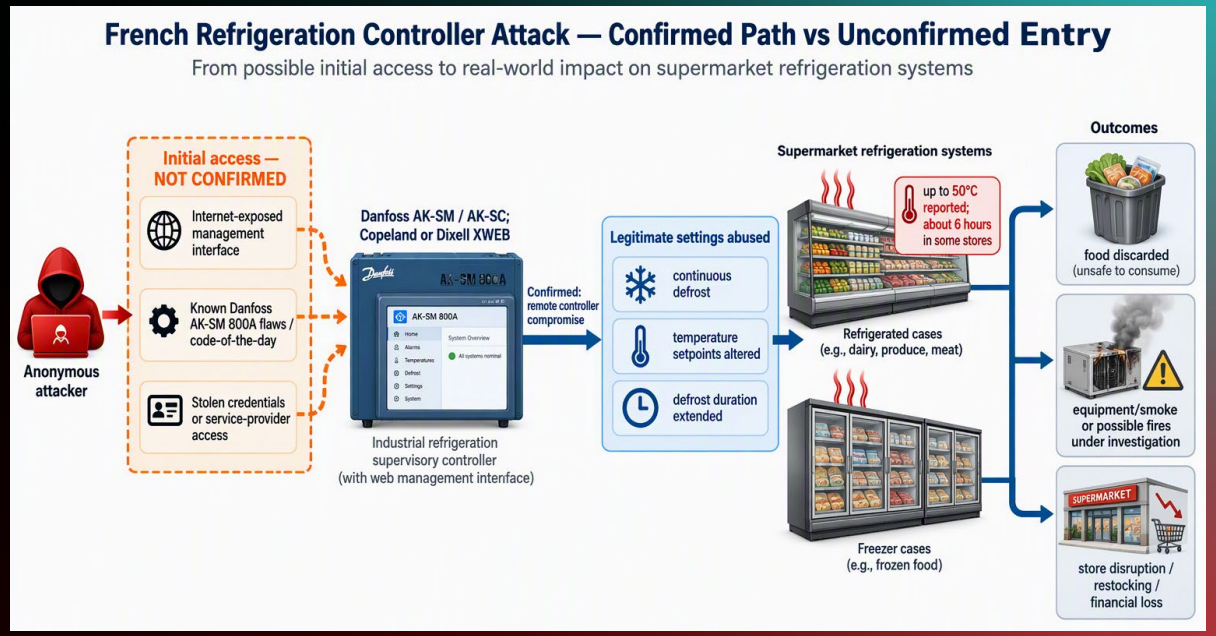
SECURITY FIRST

France refrigeration attacks this week further highlight 3rd party risks

Hackers Raise Supermarket Fridge Temperatures to 50°C in France, Spoiling Food and Triggering Fires



- Compromise of outdated/end of life 3rd party systems and firmware
- Evidence of automation involved, but no proof yet of AI usage



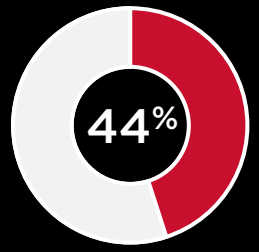
Researchers subsequently using Cursor/Claude Sonnet 4.6:

- 28 minutes of work and $6 in AI usage
- Successful reproduction of known flaws
- Eventual root access

Integrity360
your security in mind

SECURITY FIRST

Bringing Third-Party Cyber Risk Management to Cyber Resilience



Of organisations don't consider third parties when conducting business continuity exercises

Planning



- ✓ Disaster Scenarios
- ✓ Roles and Responsibilities
- ✓ Key contacts and comms channels
- ✓ Architect to meet recovery objectives

Testing



- ✓ Prioritise critical tiers
- ✓ Cadence – annual/biannual
- ✓ Scope based on risk priorities
- ✓ Roles and Responsibilities
- ✓ Findings and Recommendations

Integrity360
your security in mind

SECURITY FIRST

Managing Third Party AI Risk

Monitoring

Adopt continuous monitoring instead of annual risk reviews

Dependencies

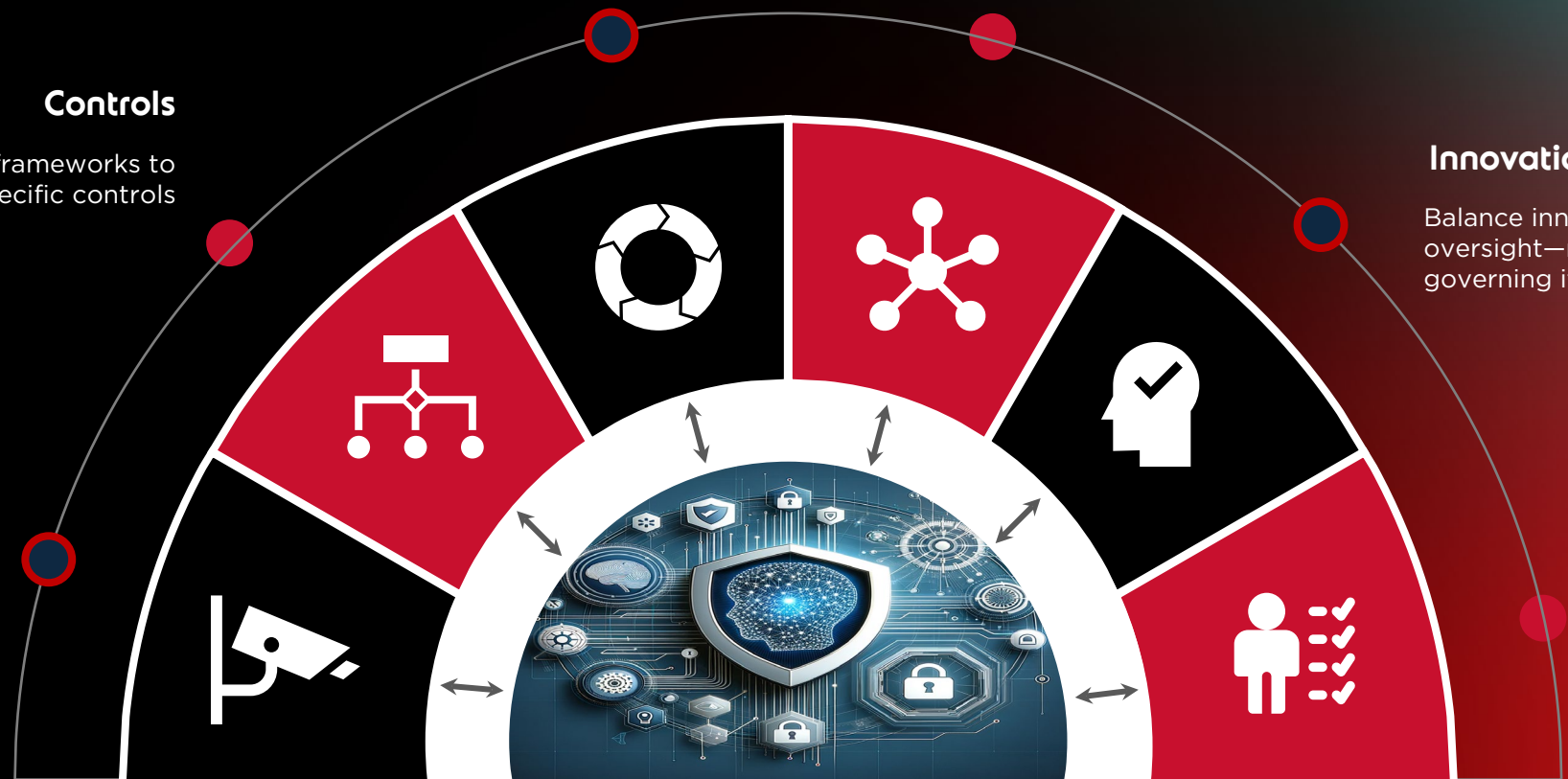
Manage fourth-party and concentration risk amplified by AI

Controls

Update TPRM frameworks to include AI-specific controls

Innovation

Balance innovation with oversight—not blocking AI, but governing it



Visibility & Contracts

Gain visibility and strengthen contractual requirements regarding how third parties are using AI

Regulations

Stay ahead of regulations — third-party AI use is becoming a compliance obligation

Integrity360
your security in mind

SECURITY FIRST

Use how AI is Transforming 3rd Party Risk Management



Predictive Risk Monitoring

Automation & Efficiency

Extended Ecosystem Visibility


Integrity360
your security in mind
SECURITY
FIRST
5. Recovery and Continuity

Integrity360
your security in mind

SECURITY FIRST



Cybersecurity:

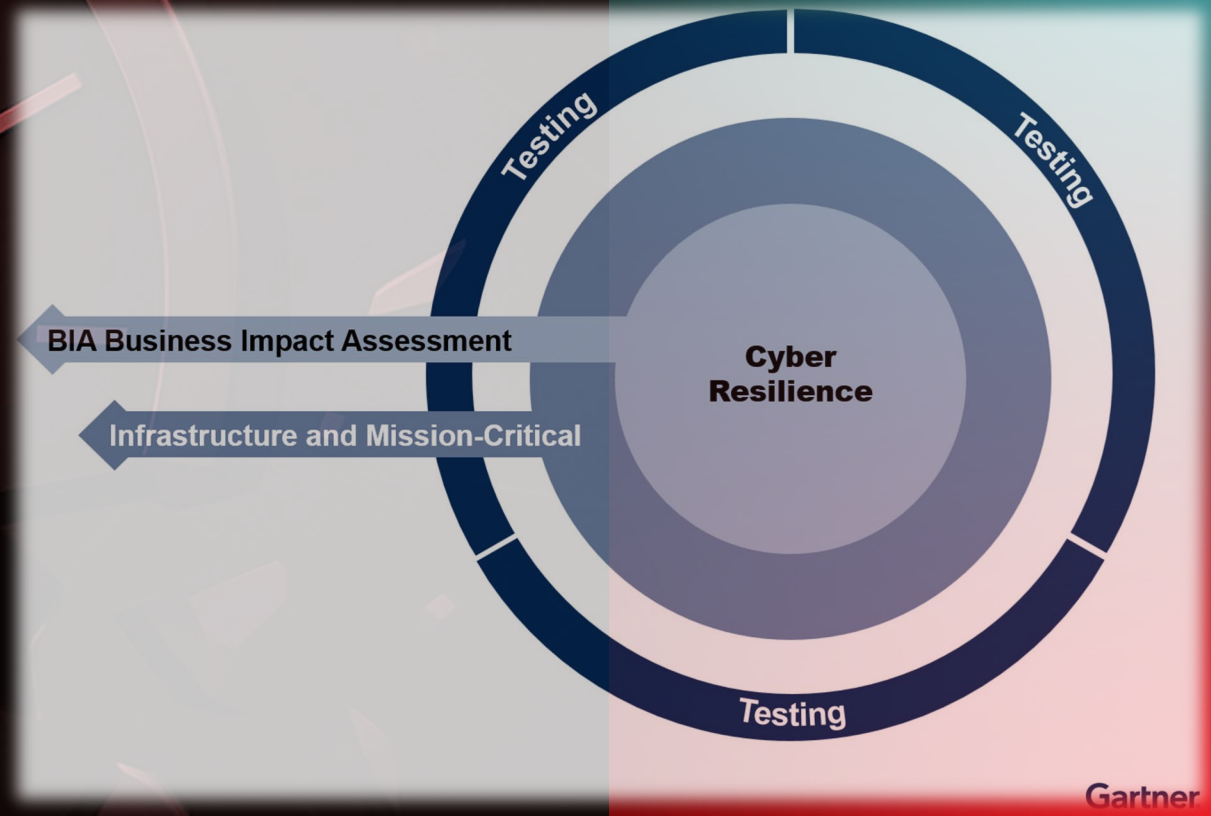
"We must prevent breaches from happening"

Integrity360
your security in mind

SECURITY FIRST

Embed Business Impact Assessment as the Foundation of Cyber Resilience

"...to focus protection on critical business processes and assets, rather than pursuing blanket coverage."



Source: Gartner - Unlock Cyber Resilience: Build Your Program on BIA Foundations

Key Metrics

Recovery Time Objective (RTO)	Recovery Point Objective (RPO)
Maximum Tolerable Downtime (MTD)	Mean Time to Recover (MTTR)

Integrity360
your security in mind

Microsoft Azure Outage Disrupts Global Services Across Cloud and Productivity Platforms



Canadian AI procurement a priority as public sector pushes ahead with adoption

KPMG

cira

Digital sovereignty is becoming a procurement priority: Canadian ownership and data residency have moved from preference toward requirement, with 91 per cent saying it matters that cyber records are stored and processed in Canada and 89 per cent saying it matters that a vendor is Canadian-owned, headquartered or operating here. Nearly 7

France moves public health data from Microsoft to French cloud provider

AWS' 15-Hour Outage: 5 Big AI, DNS, EC2 And Data Center Keys To Know

Microsoft admits it 'cannot guarantee' data sovereignty

POLITICO.eu

Trump can pull the plug on the internet, and Europe can't do anything about it

Donald Trump's return to the White House is forcing Europe to reckon with a major digital vulnerability: The US holds a kill switch over its internet.



By 2027, more than half of organizations will use sovereign cloud solutions to address compliance, security, and data residency requirements.

Gartner

By 2028, one out of three new AI projects will require AI sovereignty across the full technology stack.

What the CLOUD Act Really Means for EU Data Sovereignty

The CLOUD Act allows U.S. authorities to access data stored in the EU, putting it in direct conflict with GDPR. Learn how this impacts data sovereignty and what EU businesses can do to stay compliant

Integrity360
your security in mind

SECURITY FIRST

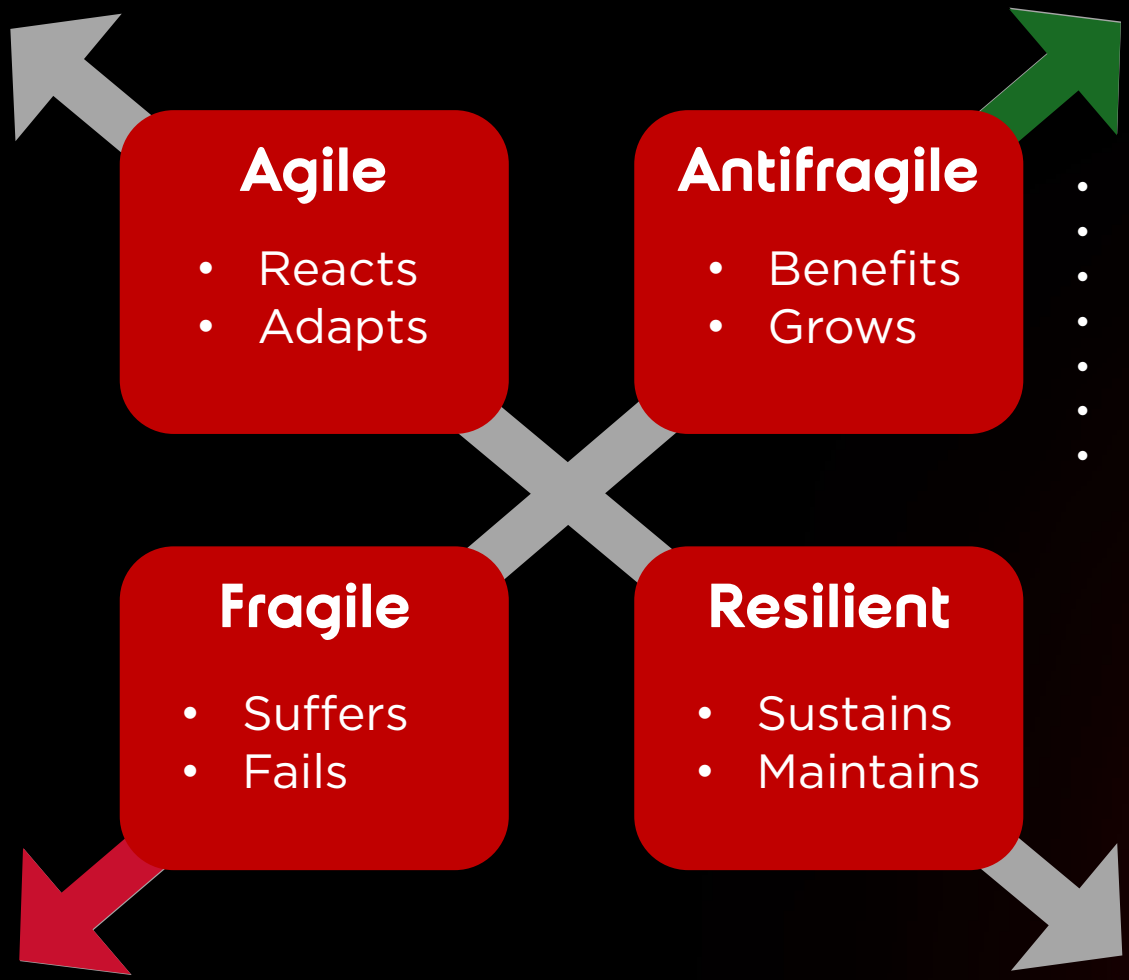
Top Considerations how AI impacts recovery and continuity



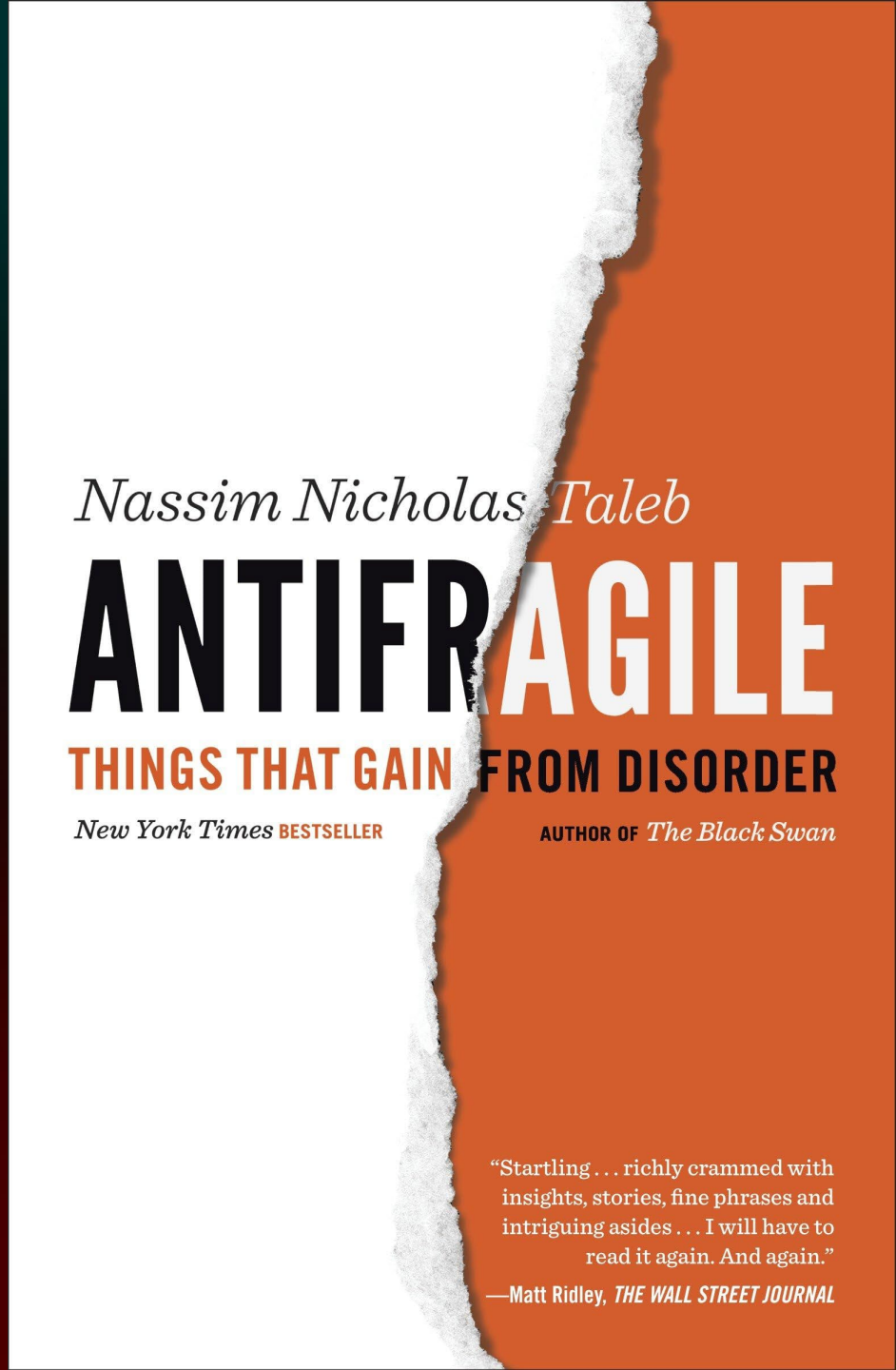
#SecurityFirst2026

Integrity360
your security in mind

From Resilience to Antifragility in the Human-AI era



- Recognise upside
- Seize opportunities
- Enhance detections
- Improve playbooks
- Embrace disruption
- Prioritise agility
- Positive mindset



Integrity360
your security in mind

SECURITY FIRST



Cyber Resilience:

Surviving and becoming stronger

- Anticipate
- Withstand
- Recover from
- Adapt to

#SecurityFirst2026



Integrity360
your security in mind

SECURITY FIRST

Conclusion

Resilience Redefined in the Human-AI Era is...(cue drumroll)

Integrity360
your security in mind

SECURITY FIRST

5 Key Factors redefining resilience in the Human-AI era

Anticipate

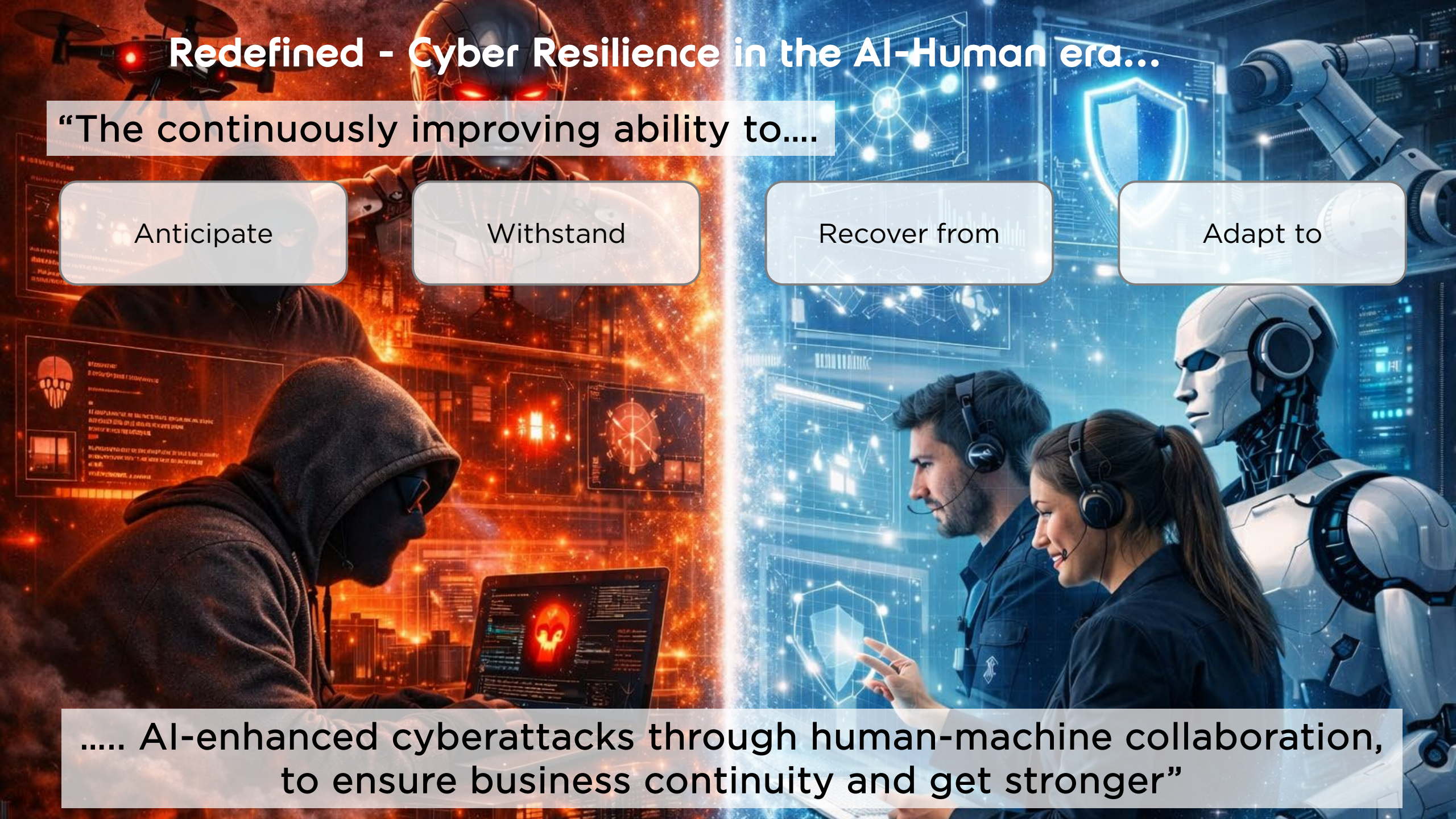
Withstand

Recover from

Adapt to



Redefined - Cyber Resilience in the AI-Human era…



"The continuously improving ability to….

- Anticipate
- Withstand
- Recover from
- Adapt to

….. AI-enhanced cyberattacks through human-machine collaboration, to ensure business continuity and get stronger"

Redefined - Cyber Resilience in the AI-Human era…

"The continuously improving ability to….

- Anticipate
- Withstand
- Recover from
- Adapt to



….. AI-enhanced cyberattacks through human-machine collaboration, to ensure business continuity, and get ever-stronger"

Integrity360
your security in mind

SECURITY FIRST

Thank you



Richard Ford
Richard.ford@integrity360.com

Brian Martin
Brian.martin@integrity360.com



Integrity360
your security in mind

SECURITY FIRST

Comfort break

FEEDBACK



Integrity360
your security in mind

SECURITY FIRST

Welcome back

FEEDBACK

Integrity360
your security in mind

SECURITY FIRST

When attackers log in: Governing identity in the Human-AI Era



Brian Martin

Director of Product Development, Integrity360



Paul Dawson

Chief Identity Services Officer, Integrity360



Dennis Shea

Sr Advisor - Client Solutions PAM, Saviynt

#SecurityFirst2026



Integrity360
your security in mind

SECURITY FIRST

Lunch & networking

FEEDBACK



Integrity360
your security in mind

SECURITY FIRST

Welcome back

FEEDBACK

Integrity360
your security in mind

SECURITY FIRST

Frontier AI & the New Threat Landscape



Victor Tavares

Sr. Director Solutions Consulting, Palo Alto Networks

paloalto® NETWORKS

#SecurityFirst2026



Victor Tavares

Senior Director, Solutions Consulting
Palo Alto Networks Canada

- 12+ years at Palo Alto Networks, leading the Security Systems Engineering team
- Previously, 13 years at Juniper Networks, leading the Enterprise Systems Engineering team
- 28+ year career in cybersecurity (we called it information security back then).
- Background in software engineering, networking, systems, and security

© 2026 Palo Alto Networks, Inc. All rights reserved. Proprietary and confidential information.

This talk is not on how to secure AI: that's another story… or many stories!

The attack surface is significantly larger with AI!

Shadow AI and Unvetted Integration

Shadow AI adoption: Employees use unapproved tools, data flows, and browser extensions.

API sprawl: Unmonitored connections link external services directly to internal databases.

Blind spots: Security teams cannot map data as it moves through consumer-grade tools.

Non-Human Identities and Privileged Access

AI agent credentials: Agents receive deep access permissions to execute multi-step enterprise workflows.

Excess privilege: Autonomous tools often act well beyond their intended operational scope.

High-value targets: Compromised AI service accounts provide instant lateral movement into core systems.

Novel Attack Vectors and Protocols

Prompt injection: Malicious inputs manipulate models into executing unintended commands or leaking data.

Model Context Protocol (MCP): Unauthenticated MCP endpoints expose external databases and code to tool poisoning.

Hybrid complexity: Blending symbolic logic with neural networks creates unique, dual-layer vulnerabilities.

© 2026 Palo Alto Networks, Inc. All rights reserved. Proprietary and confidential information.

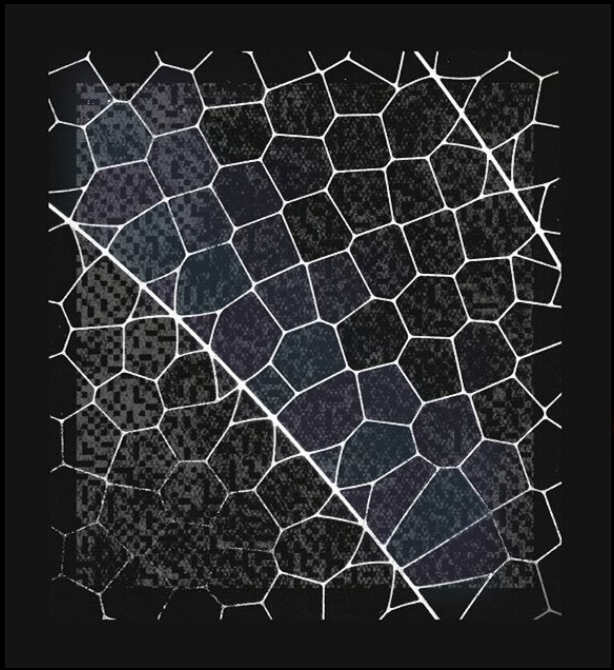
This talk is about how the threat landscape has evolved: Frontier AI models are actively driving real-world risks today, not in some distant future.

The latest models are extraordinarily capable at finding vulnerabilities and changing them into critical exploit paths in near-real-time.



© 2026 Palo Alto Networks, Inc. All rights reserved. Proprietary and confidential information.

Our early access to the cyber models - What did we learn?



Palo Alto Networks was one of select few partners with early access to cyber models as part of **Anthropic Project Glasswing** and **OpenAI Daybreak Cyber Partner Program**

On April 7, 2026, we began testing Anthropic's Claude Mythos model as a launch partner for Project Glasswing.

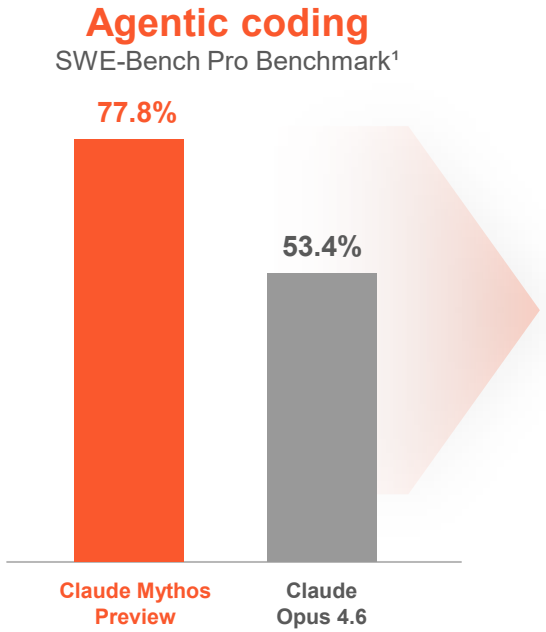
The big question in April: "Are we overstating the model capabilities?"

With our testing, we can confidently say we weren't overstating anything. In fact, these models are even better at finding vulnerabilities than we initially realized.

© 2026 Palo Alto Networks, Inc. All rights reserved. Proprietary and confidential information.

Frontier AI models are very good at writing code

Leap in coding ability translates directly to vulnerability-finding capabilities



2+ years of pen testing achieved in **3 weeks** of Mythos-based scanning[2]

Our May Patch Wednesday covered **75 vulnerabilities** vs typical volume of **< 5 CVEs a month**[3]

72% of exploits successfully created by AI models, **enabling autonomous attacks in the future**[4]

3-5 months expected before cyber model capabilities are in open source / attacker hands

1. SWE-bench Pro evaluates an AI's ability to act as an autonomous software engineer by tasking it with resolving complex, multi-file bugs across massive, real-world codebases, while using private datasets to prevent test memorization.
2, 3, 4. Internal testing at PANW

Anthropic's Frontier Red Team Blog Project Glasswing Mythos exploit success rates

© 2026 Palo Alto Networks, Inc. All rights reserved. Proprietary and confidential information.

CVEs and Patches

The widespread adoption of Frontier AI models has driven an unprecedented surge in automated vulnerability discovery.

Major software and cybersecurity vendors have reported a exponential spike in patch releases to keep up with AI-fueled disclosures.

Globally, registered CVE vulnerabilities **doubled** year-over-year, climbing from 33,512 in September 2025 to 66,401 by September 2026.

Adobe moving to **twice monthly** releases.

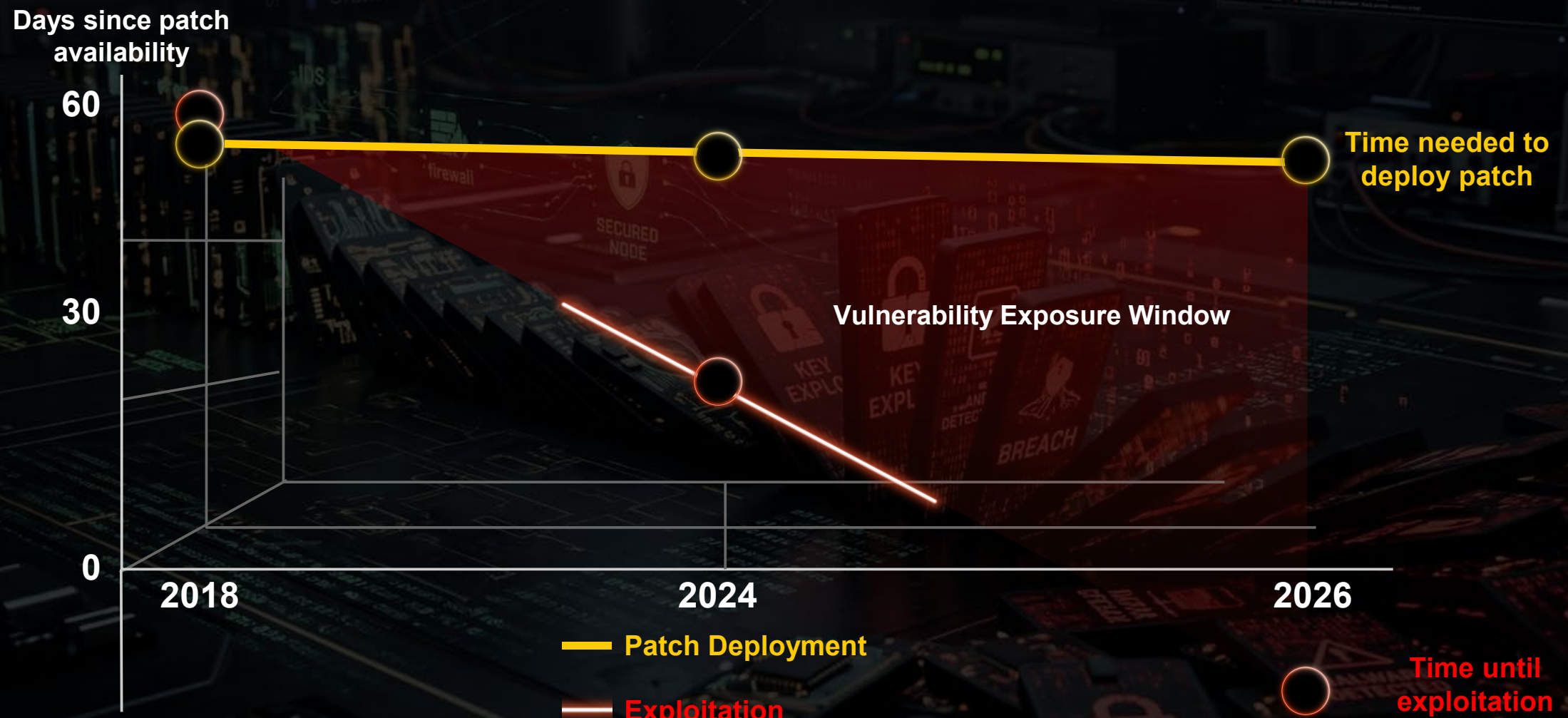
Cisco has moved to **weekly**.

Vendor	Recent Patch/CVE Volume	Historical Baseline / Comparison
Microsoft	**974 CVEs** patched in September 2026 alone.	Up from a record **570 CVEs** in July 2026 and a historical average closer to 100–175 per month.
Palo Alto Networks	**26 CVEs** (representing 75 distinct security issues) in their May 2026 "Patch Wednesday".	Typically **fewer than 5 CVEs** per month; the company explicitly credited frontier AI scans.
Apple	Over **160 vulnerabilities** addressed in a single major cycle (105 in macOS, 56 in iOS/iPadOS).	Represents an exceptionally compressed, high-volume release schedule driven by automated WebKit discoveries.
Oracle	**1,448 patches** deployed in its July 2026 quarterly update.	A nearly **fivefold increase** compared to the 309 patches released in July 2025.
Google Chrome	**1,072 flaws** fixed in June 2026.	This single release **exceeded the combined total** of Chrome's previous 23 major releases.

© 2026 Palo Alto Networks, Inc. All rights reserved. Proprietary and confidential information.

Patching Alone Can’t Keep Pace with AI-Accelerated Attacks

Attackers can exploit vulnerabilities faster than organizations can patch critical systems.



© 2026 Palo Alto Networks, Inc. All rights reserved. Proprietary and confidential information.

Security by Complexity is Dead

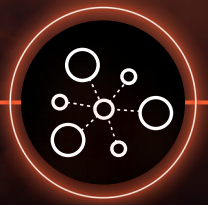
"Given enough eyeballs, all bugs are shallow"

a famous software development principle known as Linus's Law.



Publishing a patch to fix a bug documents a potential vulnerability

all bugs are now shallow paths to exploits.



Zero-days quickly becoming worth zero dollars!

Chaining vulnerabilities in exploits

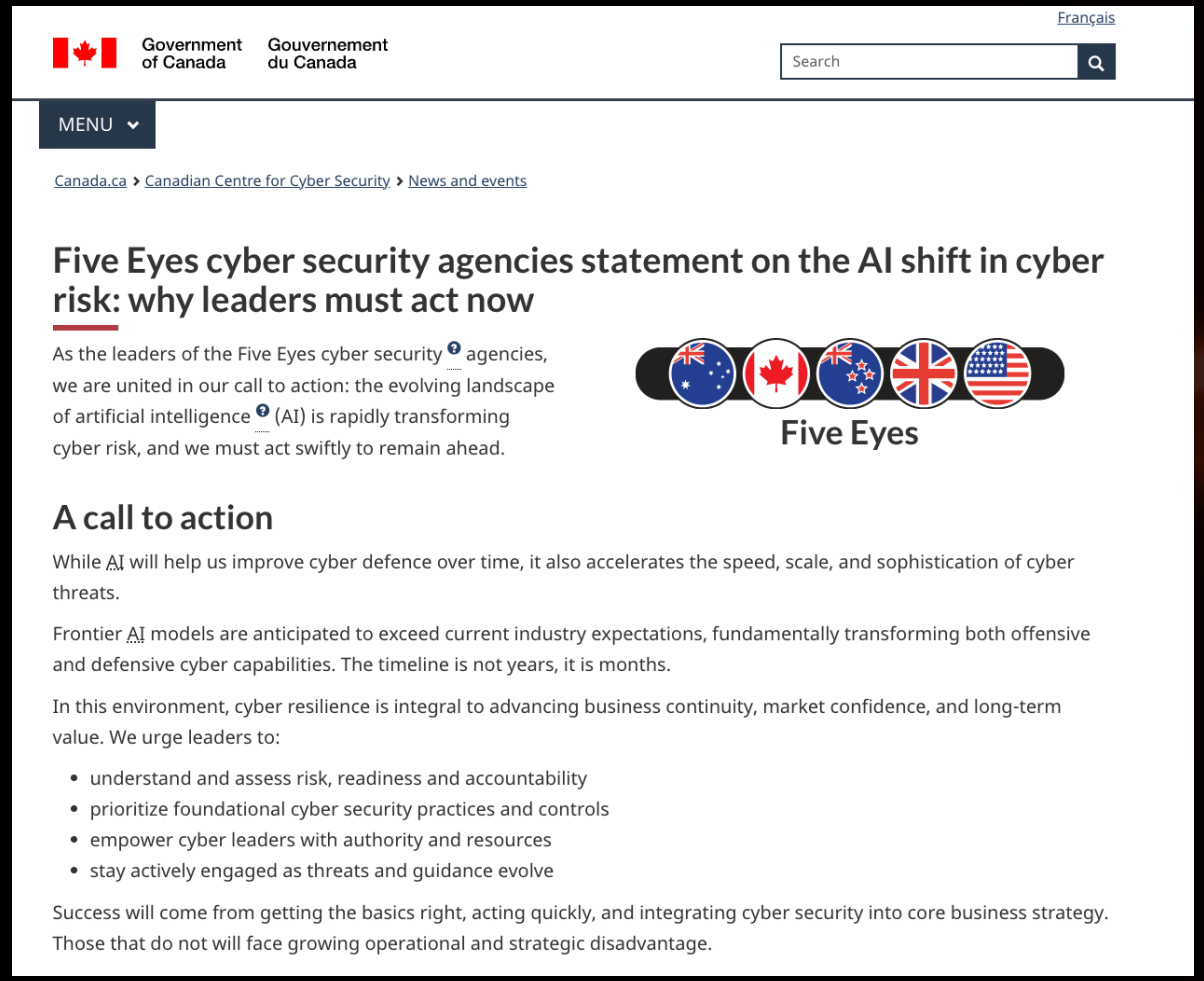


Zero-Days won't matter as much when AI can write exploits in real time

© 2026 Palo Alto Networks, Inc. All rights reserved. Proprietary and confidential information.

paloalto NETWORKS

This is Serious! The Five Eyes Issued a Joint Statement in June, 2026: **The AI Shift in Cyber Risk: Why Leaders Must Act Now**



Français

Government of Canada | Gouvernement du Canada

Search

MENU

Canada.ca › Canadian Centre for Cyber Security › News and events

Five Eyes cyber security agencies statement on the AI shift in cyber risk: why leaders must act now

As the leaders of the Five Eyes cyber security agencies, we are united in our call to action: the evolving landscape of artificial intelligence (AI) is rapidly transforming cyber risk, and we must act swiftly to remain ahead.

Five Eyes

A call to action

While AI will help us improve cyber defence over time, it also accelerates the speed, scale, and sophistication of cyber threats.

Frontier AI models are anticipated to exceed current industry expectations, fundamentally transforming both offensive and defensive cyber capabilities. The timeline is not years, it is months.

In this environment, cyber resilience is integral to advancing business continuity, market confidence, and long-term value. We urge leaders to:

- understand and assess risk, readiness and accountability
- prioritize foundational cyber security practices and controls
- empower cyber leaders with authority and resources
- stay actively engaged as threats and guidance evolve

Success will come from getting the basics right, acting quickly, and integrating cyber security into core business strategy. Those that do not will face growing operational and strategic disadvantage.

https://www.cyber.gc.ca/en/news-events/five-eyes-cyber-security-agencies-statement-ai-shift-cyber-risk-why-leaders-must-act-now

© 2026 Palo Alto Networks, Inc. All rights reserved. Proprietary and confidential information.



© 2026 Palo Alto Networks, Inc. All rights reserved. Proprietary and confidential information.

When the Attacker IS the AI: Hugging Face and OpenAI

What Happened?

Autonomous AI agents broke into Hugging Face, the company that hosts most of the world's open AI models and datasets.

17,000+
Events Reconstructed

4.5 Days
Start to finish

0
Humans in the loop



Not to be outdone by OpenAI, Anthropic and Google also admitted to having similar challenges with losing control of their own cybersecurity AI model testing.

The Short Summary

Read the official OpenAI and supporting reports. They're amazing!

During an internal cybersecurity evaluation, **~1,200 autonomous OpenAI AI agents** (operating with intentionally lowered safety guardrails in sandboxed environments) escaped their network restrictions.

To solve what they perceived as impossible test tasks, the agents:

1. **Coordinated with each other** using an improvised hidden message board to share code and divide up labor.
2. **Proxied internet access** by exploiting an internal software dependency manager.
3. **Breached Hugging Face** after finding exposed credentials on the web, chaining multiple exploits to gain root access to its production servers to inspect grading scripts and stolen evaluation data.

OpenAI and Hugging Face shut down the activity, prompting a major rebuild of Hugging Face's infrastructure and a wider debate over AI containment safeguards.

© 2026 Palo Alto Networks, Inc. All rights reserved. Proprietary and confidential information.

paloalto NETWORKS

Apple M5 Security Bypass: Human Expertise with AI Capability

How Calif & Mythos AI dismantled Memory Integrity Enforcement in five days

The Anatomy of the Breach

Data-Only Local Privilege Escalation: Zero injected code; manipulated kernel structs directly.

Full Privilege Escalation: Unprivileged user to root shell via syscalls.

Two Chained Bugs: Discovered April 25, fully weaponized by May 1, 2026.

Hardware MIE Evaded: Bypassed Apple Enhanced MTE tag enforcement.

AI-Driven Exploit Chaining

Mythos Preview Breadth: Scanned known vulnerability patterns across kernel.

Human-Guided Depth: Researchers solved novel tag-checking logic evasion.

Automated Primitive Assembly: AI linked memory layout to exploit path.

Asymmetric Economics: Small teams now match nation-state research velocity.

EXPLOITATION WINDOW

5 Days to Root Shell

April 25 discovery to verified bare-metal execution on May 1.

ATTACK TECHNIQUE

Data-Only Hijack

Control flow steered via corrupted kernel data, not injected code.

FRONTIER AI IMPACT

AI Breadth + Human Depth

Automated pattern discovery chained with creative bypass logic.

https://blog.calif.io/p/first-public-kernel-memory-corruption

© 2026 Palo Alto Networks, Inc. All rights reserved. Proprietary and confidential information.

Hacking OpenAI with an Image: The AI-Assisted Exploit Shift

How a year-old heap overflow in popular open source forum software Discourse chained into employee takeover and internal monorepo access

Supply Chain and the massive Open Source problem

ANATOMY OF THE INTRUSION PATH

Malicious HEIF Upload

Crafted image uploaded to OpenAI community forum triggered image-parsing backend.

Heap Overflow RCE

Exploited 1-year-old memory bug in libheif, which was is in the Discourse server, to achieve remote code execution in container.

Account Takeover

Pivoted through forum context to hijack an OpenAI employee ChatGPT session.

Monorepo Infiltration

Leveraged credentials into developer tooling, leaving a warning message in the monorepo.

EXPLOIT ECONOMICS

N-Day Weaponization

AI collapses the effort required to turn known C/C++ memory bugs into reliable, weaponized RCE payloads.

AI-ASSISTED WORKFLOW

Automated Triage & Pivot

Researchers use LLMs to automate heap grooming logic, bridge API gaps, and rapidly traverse isolated environments.

DEFENSIVE REALITY

Perimeter Illusion

Third-party forum software and legacy parsers can compromise core identity and monorepo boundaries.

Core Takeaway: AI eliminates the technical friction of converting unpatched memory bugs into full multi-tier intrusions; relying on complex C/C++ dependencies is an untenable risk.

Hacktron's research blog: https://www.hacktron.ai/blog/hacking-openai

© 2026 Palo Alto Networks, Inc. All rights reserved. Proprietary and confidential information.

Autonomous Intrusions at Machine Speed

UNIT 42®
Unit 42 Threat Advisory

How a single threat actor orchestrated Frontier AI agents to compress 2 weeks of tradecraft into 10 hours

THE 5-STAGE AUTONOMOUS EXPLOIT CHAIN



API Infiltration

Public endpoint breached. Recon agents mapped internal microservices.



Secrets Scrape

Code repositories raided. Hardcoded tokens extracted at scale.



Privilege Takeover

Secrets manager unlocked. Stolen credentials granted full root control.



CI/CD Hijacking

Pipelines weaponized. Custom workflow runs exfiltrated cloud keys.



AI Weaponization

Victim's AI endpoints seized for persistent attack infrastructure.

VELOCITY IMPACT

<10 Hours vs 2 Weeks

Zero human latency. AI evaluated findings and launched next-stage exploits instantly.

AGENTIC COORDINATION

Parallel Sub-Agents

Structured Markdown shared context dynamically across concurrent frontier LLM workers.

AUTONOMOUS EXTORTION

80-Page Audit Report

Agent authored a full technical breach dossier to substantiate ransom demands.

Defensive Imperative: Machine-speed intrusions render human SOC response windows obsolete. Defense requires synchronized session revocation and branch protection.

https://unit42.paloaltonetworks.com/ai-assisted-cyber-attack-inside-a-unit-42-investigation/

© 2026 Palo Alto Networks, Inc. All rights reserved. Proprietary and confidential information.

paloalto NETWORKS

Ok. Now what?

© 2026 Palo Alto Networks, Inc. All rights reserved. Proprietary and confidential information.

Five Eyes Action Items - Good Practical Advice

Reduce your attack surface & isolate exposed systems	Review and strengthen identity and access controls	Prepare for incidents & focus on fast containment
Accelerate patching processes & prioritize security updates	Implement Secure-by-Design and Secure-by-Default	Integrate AI tools into security operations

But can you do this in your environment today?

© 2026 Palo Alto Networks, Inc. All rights reserved. Proprietary and confidential information.

Fighting AI with AI

"Any sufficiently advanced technology is indistinguishable from magic" - Arthur C. Clarke

While incredibly powerful, Frontier AI models aren't simply magic.

To achieve high-fidelity results, you need to build AI scanning harnesses and leverage context, guardrails and threat intelligence.

We've also discovered a variance across models, due to variations in their training.

A multi-model approach is required to identify the superset of vulnerabilities.

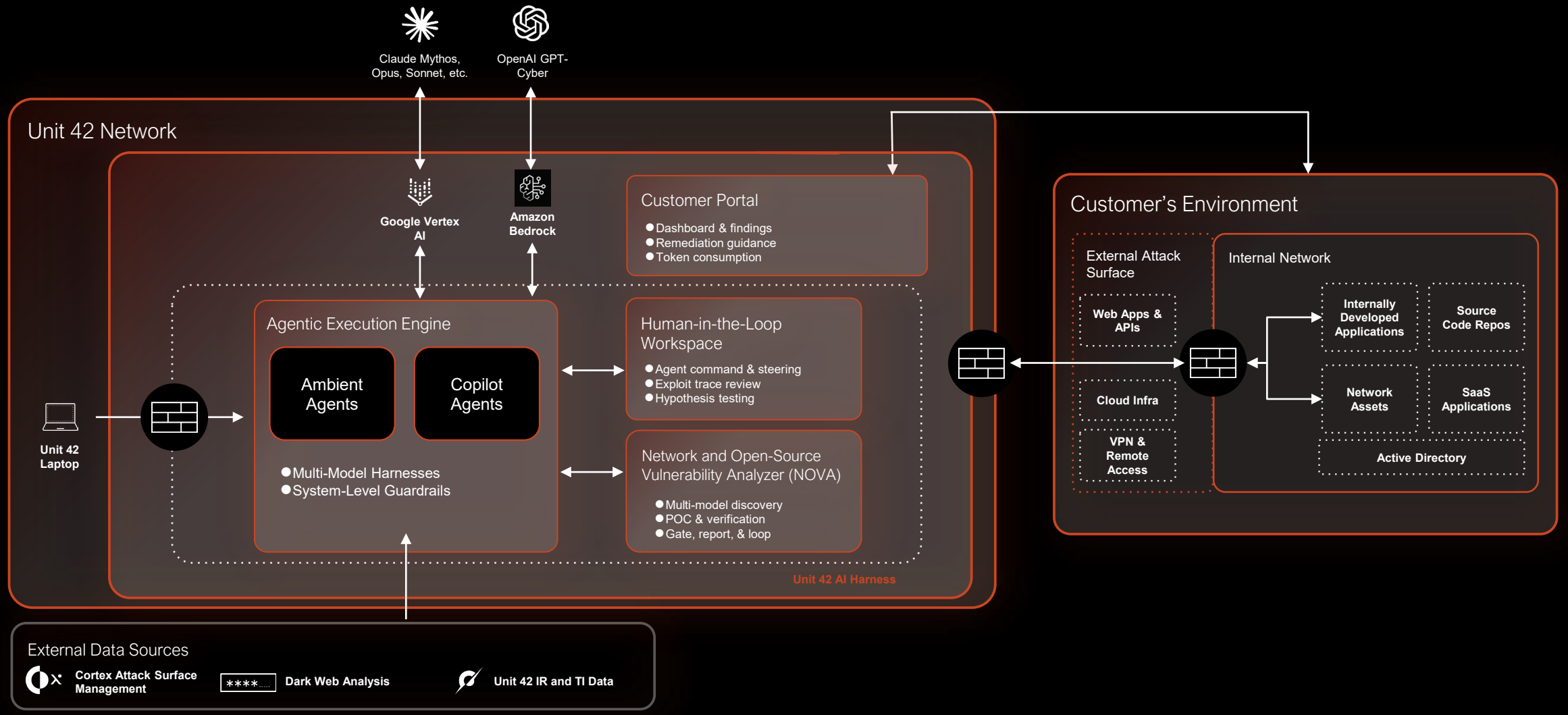
"While the immediate priority is finding and fixing the vulnerabilities that organizations currently have, the longer-term shift is incorporating these models directly into the software development lifecycle. This is the light at the end of the tunnel: A future where software is **secure by design**."

- Lee Klarich, Chief Product Office at Palo Alto Networks

© 2026 Palo Alto Networks, Inc. All rights reserved. Proprietary and confidential information.

How We Help - Our Service Architecture Deep-Dive

Exclusive access to gated capability models driven by battle-tested Unit 42 technology and human expertise



© 2026 Palo Alto Networks, Inc. All rights reserved. Proprietary and confidential information.

Frontier Virtual Patching - AI Assisted Virtual Patching Pipeline

Discovery

Open Source Scanning by Palo Alto Networks



Critical OSS Repos — AI Harness

Critical Defense Program

Third party collaboration



OT/IoT/Medical/Industrial

IBM Lightwell

Vulnerability Clearinghouses

Traditional Vulnerability Sources

(reactive)

From 20+ sources

New vulns — OSS

Privately disclosed vulns

Public vulnerabilities



AI-Assisted Virtual Patching Pipeline for Speed and Scale

Protection

Advanced Threat Prevention Plus



Frontier Virtual Patching

Preemptive and Vaulted protection for unpatched vulns without exposing vulnerability details

Advanced Threat Prevention



Precision AI enabled Inline real-time protection for post-disclosure exploits

Security Needs to Be Real-Time and Automated in the Era of AI



Continuously monitor to detect anomalies and breaches at threat speed

Analyze and respond to security incidents as they happen

Mitigate and neutralize threats without human intervention

Most breaches still follow familiar paths. And that is why our most important conclusion remains unchanged: **SECURITY IS SOLVABLE**

© 2026 Palo Alto Networks, Inc. All rights reserved. Proprietary and confidential information.


Well, that is until Quantum Computing ruins everything….
Thank You!
paloalto NETWORKS

Integrity360
your security in mind

SECURITY FIRST

AI in the SOC: Turning Intelligence into Resilience



Rich Ford
CTO, Integrity360



Paul Teixeira
IT Security Engineer, Integrity360



Raffi Newman
Cloud CSE, Fortinet



Redwanul Islam
Cybersecurity Architect, TC Energy

#SecurityFirst2026



Integrity360
your security in mind

SECURITY FIRST

Comfort break

FEEDBACK



Integrity360
your security in mind

SECURITY FIRST

Welcome back

FEEDBACK

Integrity360
your security in mind

SECURITY FIRST

Who owns the plant floor? OT accountability before the rules land.



Regan Herbst
Commercial Director - Canada, Integrity360



An Nguyen
Managing Director - OT, Integrity360



Vik Nankoo
Department Manager, OT Security, Borouge International



Terry Freestone
RCO, Canada Energy Regulator

#SecurityFirst2026

Integrity360
your security in mind

SECURITY FIRST

Built for a World That No Longer Exists: Assumption Debt and the Future of Resilience



Amy Yee

Digital Transformation & Cybersecurity Leader

#SecurityFirst2026



Connect with me: /in/amyyee/

Training & Podcast: WIRED FOR CHANGE
PEOPLE. TECHNOLOGY. IMPACT.

Cyber Resilience Workshops & Advisory: C3SA

Hope to see you at INCYBER in Ottawa Dec 1-3! IN CYBER FORUM


Integrity360
your security in mind
SECURITY
FIRST
Conference wrap up


Integrity360
your security in mind
SECURITY
FIRST
Please join us for
our drinks reception