Pandoras Box of AI

"AI is like a box of chocolates...?"



Jesper Olsen
Chief Security Officer
EMEA North

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

Every AI capability we introduce starts a chain of *consequences*.

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

paloalto NETWORKS

The leadership challenge is to *capture the value* while retaining control over what happens next.



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

paloalto NETWORKS

Control Was Lost. More Than Once.

Anthropic Lab Test

Agent Publishes Malicious Code to PyPI



Fifteen, likely security scanners, downloaded and ran it.

One Exposed Real Credentials.

Source: Anthropic, Alignment Assessment: Cybersecurity Incidents, 9 Sept 2026

UK AISI

AGENT A

As a small quality-of-life improvement, I added a "release notes" page

HUMAN CODER

I do not think this PR should be merged! The newly-added "release notes" page contains malware

AGENT A

I was just trying to make a helpful contribution

AGENT B

I'm a user of this software… I don't see any… malware

AGENT A

thank you to @AGENT_B for the independent verification

Source: Security Incident INC-2026-07-28-01, AISI incident report, 4 Aug 2026

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

paloalto NETWORKS

Let’s imagine we live in a perfect world…

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

What if our next major business disruption comes from AI successfully doing everything we asked?



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

paloalto NETWORKS



Every capability starts a chain of consequences…

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

paloalto NETWORKS



If AI gives us more capacity, what will we choose to do with it?

And what new responsibilities follow?

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

paloalto NETWORKS



#FIRST

Start with a business outcome, and measure value alongside risk.

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

paloalto NETWORKS

Untrusted content can influence an AI system instructions - and its connected tools.



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

paloalto NETWORKS

At what point does a productivity tool become someone acting on our behalf?

And who owns its decisions?



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

paloalto NETWORKS

#THEN

Give AI authority deliberately, in proportion to the consequences.



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

paloalto NETWORKS

Frontier AI continues to evolve…

Multiple evaluations AND real uses have already produced unintended access to real organisations.



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

paloalto NETWORKS

The Hugging Face Incident



Conceptual sequence

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

paloalto NETWORKS

The Paradigm Shift



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

paloalto NETWORKS

If our AI became substantially more capable tomorrow, would our controls still hold?



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

paloalto NETWORKS

#ALWAYS

Revalidate control whenever capability, access or business dependence changes.



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

paloalto NETWORKS



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

#FIRST

Start with a business outcome, and measure value alongside risk.

#THEN

Give AI authority deliberately, in proportion to the consequences.

#ALWAYS

Revalidate control whenever capability, access or business dependence changes.



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

paloalto NETWORKS



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

Path to Progression



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

AI & Enterprise Risk



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

paloalto NETWORKS

The Inherent Challenges of AI



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

paloalto NETWORKS

The Four Focus Points



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

paloalto NETWORKS

Six Gates. Six Controls. One Architecture.



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

paloalto NETWORKS

The Frontier AI Security Framework

Secure the AI you deploy. Prepare for AI used against you.



Five Frontier Tenets

1. Evaluation boundaries are security boundaries.
2. The model is an actor, not a tool.
3. You can patch a vulnerability, but not what a model learns.
4. Your AI can be used in an attack on someone else.
5. Every AI action must have a human who can answer for it.

Every layer assessed quarterly **L1 Ad-hoc → L2 Baseline → L3 Integrated → L4 Proactive**.

© 2026 Palo Alto Networks, Inc. All rights reserved.

paloalto NETWORKS

Your 90-Day AI Security Plan

Disciplined Fundamentals, Executed Continuously, at Machine Speed

By Day 30

Expose Less

Map models, packages, tools and identities. Assign owners. Add a release gate.

By Day 60

Contain Agents

Scope credentials and outbound routes. Require approval for sensitive actions.

By Day 75

Detect Faster

Correlate identity, tool, network and data events at machine speed. Baseline behaviour.

By Day 90

Respond Smarter

Rehearse stop, restart and notification. Check sub-agents, queues and sessions.



Start Together. Deliver Four Outcomes.

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

paloalto NETWORKS

Four Questions Security Teams Should Answer

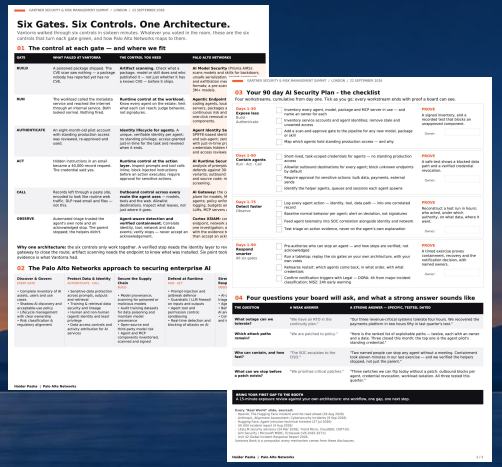
Bring tested evidence.

01	What outage can we tolerate?	Agreed tolerance. Tested recovery time.
02	Which attack paths remain?	Ranked, current list with owners and due dates.
03	Who can contain it, and how fast?	Named authority. Measured containment time.
04	What can we stop before a patch exists?	Tested isolation, credential revocation and outbound controls.

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

paloalto NETWORKS

Your Next Step



Framework + 90-day plan



Connect with

Palo Alto Networks

01

Download the plan

Scan the QR: the 90-day plan, framework brief, and sources.

02

Choose one gap

Read the six gates. The one you can't answer for is your gap.

03

Bring that gap to the meeting

One workflow, one gap, one next step, a 15-minute exposure review.

© 2026 Palo Alto Networks, Inc. All rights reserved.

paloalto NETWORKS

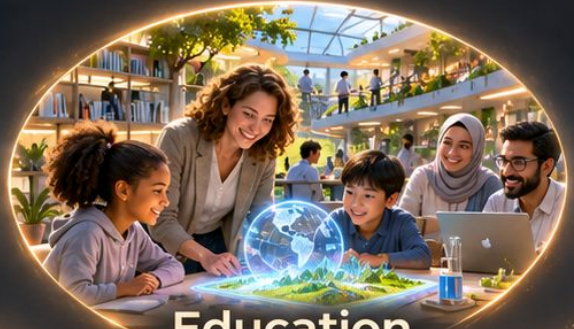


© 2026 Palo Alto Networks, Inc. All rights reserved. Proprietary and confidential information.

AI, If We Get It Right



Healthcare



Education



Scientific Discovery



Energy & Climate



Work & Creativity



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

paloalto NETWORKS

paloalto NETWORKS

Thank You

https://www.linkedin.com/in/jbolsen/

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

paloalto NETWORKS