PALO ALTO NETWORKS

perspectives

SEPTEMBER 2025

FEATURE

The First Principle of Resilience: Be Brave Enough to Fail

THE RESILIENT LEADER

MINDSET, ARCHITECTURE, AND THE NEW RULES OF AI

Ctrl + Alt + Delusion: "The Net" 30 Years Later

Navigating the Geopolitical Cybersecurity Landscape in 2025

Architecting Cyber Resilience for an Era of Disruption

Contributors

SAM AINSCOW is the Group CSO at Hill & Smith PLC.

RYAN CHAPMAN is a Team Lead for Unit 42 Managed Threat Hunting at Palo Alto Networks.

JAMIE FITZ-GERALD is a Senior Director of Product Management at Okta.

BEN HASSKAMP is the Managing Editor of *Perspectives* at Palo Alto Networks.

AARON ISAKSEN is the VP of AI Research and Engineering at Palo Alto Networks.

SCOTT MCKINNON is the CSO for Western Europe at Palo Alto Networks.

NICOLE NICHOLS is a Distinguished Engineer at Palo Alto Networks.

ANAND OSWAL is the EVP of Network Security at Palo Alto Networks.

HAIDER PASHA is the CSO of EMEA at Palo Alto Networks.

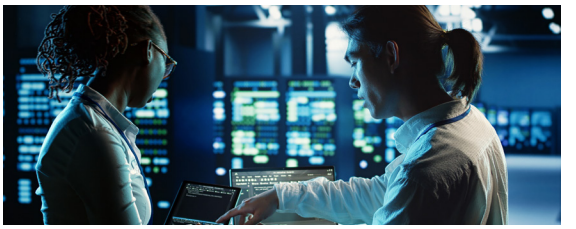
LISA SIM is the VP of JAPAC Marketing at Palo Alto Networks.

SARIT TAGER is the VP of Cortex Cloud Product Management at Palo Alto Networks.

The Mindset of Modern Leadership

The Architecture of a Resilient Enterprise

The New Rules of Engagement: Governing AI



The First Principle of Resilience

Be Brave Enough to Fail

BY SAM AINSCOW

A few years ago, I had a picture on my LinkedIn profile with a simple mantra: "Be brave enough to suck at something new." It came from a humbling moment. After many years in the industry, I walked into a training course thinking I was a pretty good cyber guy. And when I walked out later that morning, I realized I was less than stellar — I might go so far as to say "terrible." It was a stark reminder that in cybersecurity, complacency is the most dangerous vulnerability of all.

That experience taught me a foundational lesson: The path to getting better begins with the courage to be bad at something. I believe this principle applies to both individual careers and the essence of organizational cyber resilience. We invest heavily in tools and processes to prevent attacks, but incidents will happen. True resilience, like any honed craft, is built only through practice. No one steps onto a tennis court for the first time and beats a champion.

Resilience requires the same dedication. It requires a commitment to continuously test our defenses, find our weaknesses, and learn from them.

The Foundation of Resilience: A Clear-Eyed View of Risk

This resilient mindset must be grounded in a practical reality: Risk must drive everything we do in security. Before we can build effective defenses, we must first have a comprehensive, honest understanding of our unique risk profile. This means asking a series of fundamental questions.

First, do you understand your threat landscape? The threats facing a financial services firm are vastly different from those facing a defense contractor or a manufacturing business.

Second, do you understand your people? Are your users highly skilled and potentially capable of finding clever workarounds to security controls, or do you have a culture of shadow IT where unvetted SaaS applications are common? ››

> The ability to **'suck at something new'** is about being open to the feedback that the world is giving you.

Third, do you know your assets and data? If you don't have a clear asset inventory, you cannot secure your environment. Where is your most sensitive data, how is it classified, and who has access to it? Are your controls effective? Are they mitigating the specific risks you've identified, and are they doing so with the least possible friction for the business?

The Hidden Risk: Our Dependence on the Software Supply Chain

Even with a strong handle on internal risk, many organizations are overlooking a massive external threat — the software supply chain. I think we might all be horrified to see how many critical enterprise products are underpinned by an open-source project maintained by a handful of contributors in their spare time.

This is one of the biggest issues that I believe people aren't talking loudly enough about. We place an enormous amount of faith in our technology providers, but we often lack visibility into their dependencies. As an industry, we need to get more rigorous about vetting the open-source components our developers use and demanding transparency from our vendors. I believe this will continue to be a major source of breaches until we collectively address it.

Applying the Mindset: A New Model for Security Training

Nowhere is this need for a risk-based mindset more apparent than in employee security training. I see our training programs as no different from the personal protective equipment (PPE) we'd give a worker in a manufacturing plant. Yet, we often approach it with a one-size-fits-all, "sheep dip" approach that is fundamentally broken.

A resilient organization understands that risk is not uniform. The risk profile of a CEO is, of course, different from that of an accounts payable clerk. Our training must reflect this reality. It should be dynamic and personalized, using risk signals from across the business — an employee's role, their tenure or their IT literacy — to deliver the right training, to the right person, at the right time.

The Ultimate Test: Learning from Failure

This philosophy culminates in how we handle an incident. The technical steps of an effective response (identification, containment, eradication, and recovery) are critical. But for me, the single most important phase is what comes after — the lessons learned.

Every incident, every test, every exercise must be followed by a "lessons learned" session where people can be open, honest, and truthful, without fear of blame. This is a business-wide responsibility. When we run tabletop exercises, we bring everyone into the room who would be involved in a real crisis — from legal to communications to executive leadership. It is their business and their incident.

This collaborative, blameless process of learning from failure is where true resilience is forged. It's the feedback loop that enables us to find the gaps in our playbook, refine our policies and ensure that next time, we will be stronger, faster, and more effective.

Resilience Is a Feedback Loop

In our industry, the threats will never stop evolving, which means we can never stop learning. The ability to "suck at something new" is about being open to the feedback that the world is giving you. The lessons-learned process after an incident is the organizational version of that mindset.

Preparedness is everything, but it is not a static state achieved by writing a policy. Instead, it is a dynamic process of training, testing and, most importantly, learning from every failure. This feedback loop forms the core of cyber resilience. It requires courage, humility, and a shared commitment to getting better, together.

Read the article on Perspectives. ◆

Ctrl + Alt + Delusion: "The Net" 30 Years Later

BY BEN HASSKAMP

It's 1995 and America is still in the midst of booting up. The sound of dial-up modems is becoming a household melody, Windows 95 is fresh out of the box, and Sandra Bullock has just become the face of digital paranoia in "The Net" from Columbia Pictures. Released as a techno-thriller disguised as a warning label, the film follows Angela Bennett, a lonely systems analyst whose entire life is wiped clean by a few keystrokes and a particularly malicious floppy disk.

The movie arrived at a cultural inflection point. While the public didn't fully understand what it meant to "be online" (anybody seen "You've Got Mail"?), they were, however, beginning to sense what it meant to be exposed. Three decades later, "The Net" reads less like speculative fiction and more like a distorted mirror of the cybersecurity landscape we now inhabit.

As "The Net" celebrates its 30th anniversary in the annals of cinema history, I decided to revisit this slice of Americana not with popcorn in hand, but with an audit plan in mind. I asked myself, what did the movie get right? What did it miss? And how close are we, and the global ecosystem of cybersecurity, to Angela Bennett's unraveling?

Identity Theft in High Definition

The movie's conceit relies on a nightmare scenario that feels all too familiar in 2025: Angela returns from vacation to discover her identity has been overwritten. Her passport is invalid, her Social Security number reassigned, her home sold, and her existence erased. She has been transformed seemingly by a few keystrokes into "Ruth Marx."

What seemed like cinematic melodrama in 1995 now lands with chilling plausibility. Investigations from Unit 42®, for example, show attackers routinely manipulate automated identity systems using SIM swaps, credential leaks, and deepfakes. The modern attack, however, is rarely as dramatic as Angela Bennett's ordeal. In a 2023 case, for instance, attackers simply used compromised HR records to silently reroute employee paychecks across five enterprises. The takeover was precise, quiet, and fully automated. Unlike Angela, today's victims often don't realize they've been digitally erased until long after the damage is done. ››

The Gatekeeper Mythos and Real Supply Chain Mayhem

Another one of the film's most iconic plot devices is a program called "The Gatekeeper," a security tool laced with a hidden backdoor that opens access to government networks, financial institutions, and air traffic systems. Angela stumbles upon it via a mysterious icon buried within an innocuous-looking program called "Mozart's Ghost." Though rendered with all the visual subtlety of a '90s GUI, the concept remains chillingly resonant. Modern equivalents — such as the SolarWinds compromise or the exploitation of Log4j — have shown how a single piece of software, if trusted and broadly integrated, can cascade access across multiple environments. Supply chain compromises are no longer theoretical; they are among the most dangerous threats facing organizations today.

In another sequence, a Cessna crashes after a system is tampered with using "The Gatekeeper's" backdoor — an exaggerated moment, but not wholly absurd. Real-world incidents like the Colonial Pipeline attack or ransomware infiltrating a healthcare provider prove that the blurring of digital and physical risk is very much part of the modern threat landscape. Operational technology, once considered air-gapped and isolated, is now just another node in the attacker's map. This convergence of digital and physical threats is why modern cybersecurity strategy now demands a unified approach to protecting both OT and IT environments, recognizing that the consequences of a breach extend far beyond data loss.

Erased Without a Trace? Sort of...

Still, for all its foresight, the film leans heavily into melodrama. Angela is entirely erased without a trace, with no friends or colleagues to vouch for her, no visible footprint left behind. In an era where every transaction, text, and timestamp is logged somewhere, the notion of vanishing completely is, at best, a stretch, and at worst, implausible. Yet the underlying fear — the erosion of truth in the face of manipulated data — remains relevant. Unit 42 has reported extensively on the weaponization of synthetic identities, disinformation campaigns and the way attackers exploit systems of record as systems of control.

Magic Keystrokes vs. Modern Visibility

Perhaps the most unintentionally comic moment comes when Angela unlocks a buried function in a program by pressing Ctrl + Shift. The keyboard combination reveals a secret network within the U.S. infrastructure. In reality, attackers don't need cinematic flourish. They find misconfigured APIs, exposed cloud buckets, and orphaned SaaS accounts. Cortex XSIAM® research has cataloged thousands of unmonitored digital assets left outside traditional security perimeters. It's the invisibility of exposure that hides the danger, not the interface. In modern environments, missing telemetry — not magic keystrokes — is often the greatest risk.

I can confidently say "The Net" succeeds most in its emotional resonance rather than its technical accuracy. Perhaps that was intentional, or perhaps not. But the film captured something essential about the internet era before we had the language to describe it — a sense that our lives were being uploaded faster than we could secure them. The fear that a keystroke could become a weapon was prophetic. And 30 years later, that fear hasn't subsided. It has evolved.

What the movie misses — understandably, given its era — is the architecture of modern cybersecurity. Angela's world is built on implicit trust: Once you're in, you're in. There's no concept of least privilege or identity segmentation. Today, Zero Trust architectures reject that model entirely. Access is no longer binary but rather a continuously re-evaluated condition. Angela's digital ghost would have triggered multiple red flags in a system with modern behavioral analytics and access controls.

The movie also ignores the sheer sprawl of today's infrastructure. In "The Net," there is no cloud, no mobile endpoints, no SaaS applications, and no third-party integrations. Our *State of Cloud-Native Security Report* shows most cloud breaches today stem from misconfigurations, unmanaged assets, and a lack of runtime visibility, not from malware. Angela's floppy disk is quaint compared to the reality of multicloud misalignment and API sprawl. ››

And then there is the matter of artificial intelligence. In "The Net," every line of code is written by hand, and every attack is engineered by a human. Today, in 2025, attackers increasingly rely on automation, machine learning, and generative AI to both craft code and manipulate reality. Unit 42 tracks how large language models are used to write polymorphic malware, craft convincing spear phishing lures, and jailbreak themselves to produce restricted content. In the movie, Angela was hunted by a man with a silencer and a speedboat. Today, she'd be stalked by an algorithm capable of mimicking her voice, replicating her typing cadence, and generating synthetic video footage of her committing a crime she didn't even know she was framed for.

What If "The Net" Were Remade Today?

Well, to start, Angela wouldn't need a floppy disk. She'd be bouncing between federated ID systems, chasing down rogue admin tokens, and dodging deepfakes that look suspiciously like her ordering ten bitcoin transfers. "The Gatekeeper" wouldn't be hidden behind a keystroke; it would be a quietly overprovisioned API buried three integrations deep. Jack Devlin wouldn't seduce her at the beach. He'd phish her with a "security alert" email from a convincing fake IT desk.

Her adversaries, "The Praetorians," would be shadowy attackers from a state-sponsored advanced persistent threat group, relying on a vast network of compromised cloud accounts for their attack. And, to prove her innocence, Angela wouldn't be looking for a single open terminal at a trade show. Now, she might have to navigate the dark web to securely leak data to a journalist or use her knowledge of security architecture to find immutable logs proving the manipulation, but that's an entirely different story.

What is certain, though, is that today, Angela wouldn't log on to the Net; the Net would log on to her. Undoubtedly, she would be subtly surveilled and carefully tracked until the security structures around her crumbled.

From Thriller to Threat Model

To its credit, "The Net" raised awareness of digital risk long before cybersecurity became cocktail party conversation. It wasn't the technical details the movie got right. It was the emotional undercurrent — that creeping fear that you could lose your identity, your history, and your reality, all without ever seeing the person who unplugged you.

Thirty years later, cybersecurity teams are still fighting that same fear — just with faster adversaries, bigger networks, and higher stakes. The threats have evolved and so have we.

At Palo Alto Networks, our job isn't just to stop the malware or close the misconfiguration. It's to make sure no one wakes up one morning to find the world no longer remembers them. No missing person poster. No breadcrumb trail. No second chance.

Because in 2025, the real thriller isn't on screen. It's on the network.

And while the identity theft depicted in "The Net" might seem like Hollywood fantasy, modern threats are all too real.

Read the article on Perspectives. ◆

My First 10,000 Days in Cybersecurity

BY HAIDER PASHA

A couple of months ago, I did the math. I've been in the cybersecurity industry for roughly 10,000 days — a milestone that sounds immense until you realize how quickly the days turn into decades. This reflection inspired me to look back at the journey, not just for myself, but for our entire industry. While the core threats we face — malware, denial-of-service, meddler-in-the-middle attacks — remain stubbornly familiar, the landscape around them has been completely terraformed. What has changed is the speed, scale and sophistication of our adversaries; the evolution of our role as defenders; and the strategic imperative to change how we think about security itself.

Tough Lessons, but a Foundational Experience

My own journey began with an unintentional act of campus-wide chaos. In the mid-'90s, as a computer science student at Purdue, I was given an assignment on interprocess communication. The goal was to write a program that could self-replicate across different processes. I became so engrossed in the challenge that I decided to take it a step further: What if I could make it replicate across multiple machines on the network?

In what I thought was a moment of cleverness, I created a program that did just that. It wasn't malicious; it didn't steal data or delete files. As a learning experiment, I even added a harmless pop-up message — "Hello, Earthlings" — to confirm it had been executed. You can probably guess what happened next. The program began propagating across almost every computer lab on campus. Machines crashed under the unexpected load, and within hours, the IT department had to shut down the entire network.

After I confessed, the university, to its great credit, didn't punish me. Instead, they worked with me to build a kill switch and understand the vulnerability. That experience was foundational. It taught me that, just because you can do something, it doesn't mean you *should* do it. More importantly, it taught me the critical need for guardrails, for control, and for having a good set of brakes when you're moving fast. It's a lesson that developers, even 10,000 days later, are still learning as we work to embed security into the beginning of the development lifecycle, instead of treating it as a speed bump on the road to innovation. ››

The CISO: From Technical Operator to Business Executive

When I began my career, there was no such thing as a CISO. We were security managers, focused almost exclusively on the network and the endpoint. Today, the CISO has become a cornerstone of digital transformation, a shift that accelerated dramatically post-COVID when the business turned to us first to enable secure, remote work.

The modern CISO can no longer be just a technologist whose knee-jerk reaction is to buy the latest and greatest tool. I've seen the most successful leaders evolve across four key areas:

- **Strategic shift:** They've moved from being a technical operator to a business executive, capable of having board-level conversations and quantifying risk in business terms.
- **Scope expansion:** Their focus has expanded beyond the organization's walls to include third-party risk management, privacy and compliance integration. They understand that you are only as strong as your weakest supplier.
- **Investment optimization:** They are the gurus of the budget, focused on ROI measurement and technology portfolio optimization rather than simply acquiring new products.
- **Leadership and crisis management:** The best CISOs I know are cross-functional workhorses. They can speak the language of DevOps, finance and legal, championing security across the enterprise. They are also experts in crisis management, drilled and ready for the inevitable incident.

This Isn't Just Consolidation: It's Platformization

For years, organizations have tried to solve the problem of complexity by stitching together dozens of best-of-breed products. I saw this firsthand in my previous roles. The intention was to create a "platform," but the reality was a tangled mess of disparate tools that failed to integrate on a policy, control or visibility level. It didn't work because it mirrored the problem instead of solving it.

When our CEO, Nikesh Arora, coined the term "platformization," it crystallized a concept that the industry desperately needed. This type of platformization doesn't just mean consolidation; consolidation is merely one of its many outcomes.

A true platform approach is about streamlining operations through a *single, natively integrated system*. It's about leveraging the same rich, accurate and comprehensive data across your entire security posture to deliver better outcomes. The benefits are clear:

- **Unified security and operational efficiency:** You eliminate the complexity of managing dozens of vendors and siloed tools.
- **Superior analytics:** You gain correlated insights from machine learning that is trained on a complete dataset, enabling predictive capabilities that can anticipate and prevent threats.
- **Demonstrable business impact:** You can show the board faster response times, reduced vendor overhead and simplified compliance, proving that security is a business enabler, not a cost center.

The Next 10,000 Days

Predicting the future is impossible, but I can tell you what the CISO of tomorrow — or perhaps the Chief AI Security Officer — will need. That's a flexible mindset. The future of the SOC should be 100% automated. We are already seeing the emergence of personal AI agents that can manage our calendars and communications; it's not a stretch to imagine one dedicated to our personal security.

Ultimately, whether used by attackers or defenders, AI is only as effective as the data it's trained on. That is the fundamental truth. To stay ahead, we must have the best, richest and most accurate cybersecurity data to power our defensive AI models.

To future-proof our strategies, we must foster a culture of security awareness where every employee plays a role. Any digital transformation initiative that doesn't have cybersecurity embedded as its first step is destined to fail. From that panicked night in a Purdue computer lab to today's boardrooms, the core lesson remains the same: Building without brakes is far from innovation, but rather an accident waiting to happen. The challenge for the next 10,000 days is to build with resilience and purpose at the core.

Read the article on Perspectives. ◆

Innovation's New Engine: The CMO-CIO Partnership in the AI Era

BY LISA SIM

Consider a single statistic: 85% of all work now happens in the browser. This represents a fundamental shift of the corporate workspace. While this shift has unlocked incredible productivity, it has also transformed the browser into the most critical — and most targeted — attack surface for the modern enterprise.

This same redefinition of the workspace is now happening with artificial intelligence, but at a speed and scale that dwarfs previous shifts. AI is rapidly moving from a theoretical tool to an embedded, active participant in daily workflows, particularly in marketing, which has become a key arena for its adoption.

Recent McKinsey research validates this trend,[1] placing marketing at the forefront of AI implementation alongside IT and sales. For marketing leaders, this means AI is becoming as foundational as the browser itself. This new reality — where AI increasingly powers employee workflows — creates both immense opportunity and a new, urgent imperative to secure these powerful new capabilities.

What If AI Fails the Brand?

For marketing leaders, the unprecedented adoption of AI has created a new and complex risk environment for the brand. As noted by KPMG experts in an April 2025 study,[2] this challenge stems from a potent combination — the speed of AI adoption, a widespread lack of AI literacy and often weak internal governance.

This new layer of AI risk is compounding the existing threats that already target the modern workspace. A recent Omdia report found that 95% of organizations experienced browser-based attacks like phishing and malware between 2023 and 2024. With the browser now being the primary interface for AI tools, these two threat vectors have become deeply intertwined. And attacks are already a top concern.

This risk is taking hold across the entire marketing ecosystem, creating a new, three-front challenge for CMOs. It begins with the unsanctioned employee use of powerful tools like ChatGPT and extends to the AI models built directly into new products and services. It's also deeply embedded in the AI capabilities of the core CRM and martech stacks that power our daily operations.

A New Threat to Brand Integrity: Shadow AI

The appeal of AI tools within marketing teams is undeniable — they promise speed, convenience and power. This widespread adoption, however, has a significant downside. The KPMG research highlights the scale of this challenge. That is, nearly half of employees admit to using AI in ways that contravene company policies, including uploading sensitive corporate information into public AI tools.[3]

This widespread, unsanctioned use has given rise to a new and significant threat — shadow AI. Much like the shadow IT of the past, this phenomenon creates a massive blind spot in an organization's security posture. The scale of this risk is staggering. Gartner® ››

predicts that through 2026, internal policy violations will be the root cause of at least 80% of unauthorized AI transactions, stemming from, for example, information oversharing and using unapproved tools.

From Brand Steward to Risk Partner

This new reality requires new accountability for marketing leaders. Our role must expand beyond stewarding the visual identity and voice of our brand; we must now become the stewards of its algorithmic integrity. This means forging a new, deeper and non-negotiable partnership with our CIO and CISO. Traditional silos between marketing and IT have become a direct threat to the brand. The only path forward is to jointly architect an AI enablement framework that transforms shadow AI from a hidden risk into a managed engine for innovation.

Forging an Alliance Between Marketing and IT

Managing the risks of a modern, AI-powered marketing department requires a new level of cross-departmental trust and collaboration. While the CMO must spearhead the effort to balance innovation with security, the technical nature of solutions, like browser controls and AI governance, means the CIO and CISO are essential partners.

This has forced data security and access management to become a shared priority. CMOs and CIOs must now step outside of their traditional remits to build a unified strategy. This alignment enables teams to reimagine work with AI at its center to meet the marketing team's needs for innovation while ensuring the enterprise remains secure.

The Work of Alignment: A Guide to CMO-CIO Collaboration

This alignment, though, doesn't happen by accident; you must architect it with intention. You can build a successful partnership on three core pillars:

1. **Co-author an AI enablement framework.** The first step is to move beyond ad hoc tool adoption. The CMO and CIO should jointly create a formal framework that establishes an approved list of AI tools, sets clear usage policies for sensitive data and outlines a plan for ongoing employee education. This step transforms shadow AI from a risk into a managed asset.
2. **Establish a shared roadmap and cadence.** True alignment requires a shared rhythm. This means that marketing and IT leadership must establish regular, strategic meetings to review roadmaps, anticipate challenges and ensure that security is a design partner in new marketing initiatives, not a gatekeeper at the end.
3. **Run joint crisis simulations.** The ultimate test of alignment is how the teams respond under pressure. The CMO and CISO should run joint tabletop exercises that simulate a brand-damaging AI incident, such as a compromised chatbot or a biased campaign. This test builds the muscle memory and trust required to effectively navigate a real crisis.

Innovation Demands a New Operating Model

The traditional, siloed relationship between marketing and IT is more than inefficient; in the age of AI, it is a direct threat to the brand. The path forward requires more than occasional collaboration. Rather, it demands a new, deeply integrated operating model. This means moving beyond separate roadmaps to a single, shared vision for how to deploy AI securely to drive business outcomes. It means running joint crisis simulations to build the efficiency and accuracy required to navigate a brand-damaging AI incident. This transforms the relationship from a series of handoffs to a true partnership, where security becomes the foundation upon which marketing can safely innovate.

This new alliance is the true engine of modern innovation. It is a partnership where security is not a gatekeeper to marketing's ambition, but the platform that enables it. By working in lockstep, marketing and IT leaders can transform their organization, empowering their teams to use the full potential of AI and other powerful technologies safely.

This is how we move beyond simply securing the business to building a secure business by design. It is how we unlock the future of what's possible.

Read the article on Perspectives. ◆

1. *The state of AI: How organizations are rewiring to capture value*, McKinsey, March 2025.
2. "Global study reveals Australia lags in trust of AI despite growing use," McKinsey, April 29, 2025.
3. Ibid.

Navigating the Geopolitical Cybersecurity Landscape in 2025

BY ANAND OSWAL

While the digital frontier is no stranger to turbulence, 2025 promises to bring geopolitical uncertainty that continues to impact the cybersecurity landscape. Bad actors — ranging from nation-states to shadowy hacktivist collectives — continue to exploit cyberspace to settle scores and sow chaos. What's different now is the increasing sophistication of their methods, creating a landscape more volatile and treacherous than ever before.

The World Economic Forum reports 72% of global executives now factor geopolitical events into their cybersecurity strategies.[1] And they have good reasons. Cyber risks in 2025 are not just multiplying — they are diversifying, growing in intensity, and becoming maddeningly unpredictable. The question, then, is not *if* your organization will be affected but *how profoundly* and *how soon*.

The Many Faces of Geopolitical Cyber Risk

The sources of today's cyberthreats are an unnerving kaleidoscope of political upheaval, technological advancement, and sheer human cunning. They are neither localized nor neatly categorized, sprawling across continents and industries. These threat actors take advantage of almost every opportunity in front of them.

A great example of this is political change, like we saw last year when half of the world's population participated in elections. Leadership changes attract cyber adversaries like moths to a flame, their motivations ranging from political subversion to digital sabotage.

Then there is warfare – both physical and digital. The ongoing conflict between Russia and Ukraine offers a sobering example: Cyberattacks have become a de facto weapon of war, deployed to disrupt infrastructure and paralyze economies. "It's likely cyberthreats will continue at least as long as the physical conflict does," noted Paul Proctor, a leading Gartner® analyst. "The 'fog of war' can challenge situational awareness and panic will increase the risk of mistakes, creating an advantageous situation for bad actors."[2] As conflicts ripple across regions, this fog of war will continue to obscure truths and amplify vulnerabilities.

Hacktivists, too, are evolving. No longer satisfied with defacing websites or disrupting minor services, they now target operational technology and critical infrastructure with surgical precision. Intelligence agencies in the U.S., Canada, and the U.K. warn of escalating attacks on power grids, transportation systems, and water supplies.[3] This serves as another unnerving reminder that the scope of cyber risk is bound only by the ingenuity of its perpetrators.

And perhaps the most unsettling trend of all: the democratization of cyber weaponry. What was once the domain of a few well-funded nation-states is now accessible to nearly anyone with motivation and a connection to the dark web. The barrier to entry has all but evaporated, leaving organizations vulnerable to a dizzying array of threats.

What Organizations Must Do: A Strategic Imperative

The path forward is as clear as it is daunting: Organizations must act decisively and with a sense of urgency. Cybersecurity can no longer be relegated to the IT department or considered a purely operational concern. It is, therefore, a strategic imperative — demanding a multifaceted response.

1. **Cybersecurity Must Move at the Speed of Business**

   Cybersecurity cannot be a bottleneck to innovation. Security must be as dynamic as the business itself, enabling agility while ensuring resilience. With AI adoption accelerating across industries, it promises efficiency gains but also introduces new risks. Organizations must design security that keeps pace, ensuring AI-driven tools don't just automate processes but actively defend, adapt, and resist manipulation. The future belongs to businesses that build security into every innovation, rather than retrofitting protection after threats emerge. ››

2. **Get Ahead of the Most Evasive Threats**

   Today's cyber adversaries don't knock on the front door. They slip through the side entrance, blend into the digital wallpaper, and exploit vulnerabilities that traditional security tools can't see. That's why organizations need machine learning and deep learning to detect, predict, and neutralize threats before they materialize. AI-powered defenses must continuously learn, drawing from real-world attack patterns to outmaneuver cybercriminals in real time. Detection shouldn't be the ultimate goal, best-in-breed prevention is the only way forward.

3. **Secure Everything, Everywhere**

   Holistically, cybersecurity *must* protect all users, all apps, all devices. Whether in an office, a home network, a factory, or a sprawling cloud environment, security can't be an afterthought. Every interaction, every access point, and every connection must be scrutinized. No blind spots, no assumptions. The future belongs to organizations that secure their infrastructure — from endpoints to clouds to critical infrastructure — seamlessly.

4. **Fundamentally Transform Security Operations**

   Security teams are drowning in alerts, responding to yesterday's threats while today's attacks unfold. Organizations must pivot to powerful, adaptive solutions designed for scalability and contextual awareness. Purpose built-solutions like Cortex XSIAM can automate detection, investigation, and response, cutting through the noise and surfacing real threats in real time. But with this comes a caveat: Adding AI alone won't fix what's broken. The real shift comes from platformization — a way of unifying security tools into a single, adaptive ecosystem that enhances visibility, reduces complexity, and turns data into action. Instead of layering on more technology and hoping for the best, organizations need an integrated approach that brings security operations, intelligence, and automation together in a seamless, scalable model. Because at the end of the day, the best security teams don't just detect threats. They stop them before they start.

5. **Prepare for the Breach That Hasn't Happened Yet**

   Here's the uncomfortable truth: Even the best defenses can be breached. And it's not enough to simply brace for the impact. The organizations that weather cyberattacks best aren't the ones scrambling to respond; they're the ones for whom response is an exercise already rehearsed to muscle memory. Resilience isn't a reactive measure — it's an architectural principle. That means having rigorous attack simulations, real-time response playbooks, forensic-level visibility, and dedicated cybersecurity experts on standby. The best teams don't wait for a breach to confirm their preparedness; they've already identified the gaps, closed the loopholes, and hardened their infrastructure long before an adversary tries to test it. A breach, when it inevitably happens, is neither a surprise nor a catastrophe — it's a challenge already met.

Action Is a Collective Responsibility

The geopolitical landscape of cybersecurity in 2025 is not for the faint of heart. It demands resilience, foresight, and above all, a willingness to adapt. Organizations that treat cybersecurity as a box-checking exercise will find themselves outpaced and outmaneuvered.

Instead, the mandate is clear: Act boldly, invest strategically, and collaborate widely. In this high-stakes game, complacency is the enemy, and preparation is the only path forward. Because in today's interconnected world, the line between safety and vulnerability is drawn, not by geography or politics, but by the choices we make and the defenses we build, every single day.

Read the article on Perspectives. ◆

1. "Cyber Geopolitical Intelligence: Making the Connection Between Geopolitical Cybersecurity Threats," Intel 471, July 2024.
2. Paul Porter, "How Geopolitics Impacts the Cyber-Threat Landscape," Gartner, June 10, 2022.
3. Ryan Daws, "Global Agencies Warn of Increased Cyber-Attacks Against OT Devices," IoT News, May 2, 2024.

paloalto NETWORKS | UNIT 42

Fake clicks. Fake support calls. Real damage.

The rise of social engineering.

GET THE REPORT

Architecting Cyber Resilience for an Era of Disruption

BY SCOTT MCKINNON

For cybersecurity leaders, particularly those working with defense, intelligence, and critical infrastructure, the definition of "cyber resilience" is undergoing a forced evolution. It is no longer a theoretical concept centered on withstanding a single blow. Today, resilience is the urgent, practical capacity to sustain operations amidst a relentless barrage of cyberattacks where the primary goal is *both* theft and disruption.

This shift in adversary strategy changes everything. In boardrooms across Europe, the dialogue has pivoted. The familiar query — "Are we protected?" — now yields to more urgent anxieties, sharpened by directives like NIS2 and the upcoming UK Cybersecurity and Resilience Bill:[1] "Can we recover?" "How severe will the disruption be?" "How quickly can our services resume?" Answering these questions requires a new defensive playbook, one built for an era where artificial intelligence (AI) is used as both a formidable weapon and an indispensable instrument of our own resilience.

A Focus on Disruption

From our vantage point, we've seen a dramatic shift in adversary behavior. In responding to incidents globally, we've observed that 86% of cases now involve a deliberate attempt to disrupt a victim's core operations.[2] Attackers are innovating relentlessly, using AI and automation to achieve a speed and scale that fundamentally challenge our traditional defensive postures.

The numbers paint a stark picture. Our research shows that attackers can now exfiltrate data from a compromised network in under a single hour in many cases. They are armed with an ever-expanding arsenal, with nearly 9 million new, unique threats discovered daily. This increase in volume and velocity compresses our window to respond from days to minutes. ››

The Defender's Dilemma: A Fractured Defense and an AI Paradox

Unfortunately, our traditional response to this complexity has often been to add more tools. It's common for a single security organization to manage 50 — sometimes up to a 100 — different point products. This tool sprawl, far from solving the problem, has become a strategic vulnerability, creating blind spots that overwhelm our security operations centers (SOCs). We know that, in nearly every breach, the signals of an attack were present but were missed because the critical data was siloed in a separate tool or lost in a sea of untriaged alerts.

Compounding this challenge is the emergence of the AI paradox: The engines we are deploying for defense have become a new and critical attack surface. Adversaries now target the AI models themselves, by using prompt injections to manipulate behavior in attempts to exfiltrate data from conversational interfaces and exploit overpermissioned AI agents to move laterally. This means our human-centric SOC is both overwhelmed by a fractured defense and ill-equipped to secure the complex logic, memory, and data access patterns that these new AI systems depend on.

The Future of Resilience: From Shifting Left to a Unified Platform

Answering this threat demands transformation, not incremental improvement. The first step is a commitment to "shifting left," building security into the beginning of our application development and infrastructure processes.

The ultimate solution lies in changing our architectural philosophy. The path forward is through platformization. I don't mean simple vendor consolidation; I mean adopting an integrated platform that unifies security across the entire enterprise — from the network and endpoints to the cloud and the SOC itself.

> The only way to fight machine-speed attacks is with machine-speed defense.

A platform approach provides three critical advantages for this new era:

1. **Complete, unified visibility:** By ingesting data from every source into a single, normalized data lake, a platform eliminates the blind spots created by siloed tools. It gives defenders the comprehensive visibility needed to see the faint signals of a sophisticated, disruptive attack.
2. **AI-powered automation:** The only way to fight machine-speed attacks is with machine-speed defense. A platform applies AI and machine learning across a complete dataset, enabling the automation of threat detection, triage, and response at a speed that humans alone cannot achieve.
3. **Simplified operations and enhanced resilience:** By unifying your security architecture, you reduce operational complexity, free up your talented security professionals to focus on high-value tasks like threat hunting, and build a more resilient posture that can withstand and recover from disruptive attacks.

Defending against this new era of disruption is a significant challenge, but it is solvable. By shifting our mindset from buying more tools to building a unified, intelligent platform, we can meet the threat of AI-driven adversaries and architect a more secure future for our organizations.

Read the article on Perspectives. ◆

1. "Cyber security and resilience policy statement," UK Government Department for Science, Innovation & Technology, April 2025.
2. *Global Incident Response Report 2025*, Palo Alto Networks Unit 42, February 2025.

Identity Under Siege: A Leader's Guide to the New Front Line of Security

BY JAMIE FITZ-GERALD

Cybersecurity became real for me the day my colleague's laptop vanished from his desk. We were working at a defense contractor, and he had downloaded a government report on cyber threats. Unbeknownst to him, it was laced with a backdoor. When he returned from a trip, his machine — and the only copy of his master's thesis — was gone, confiscated by our security team. That was the moment the threat landscape moved from an abstract concept to a tangible reality. It was a pivotal lesson that made the abstract tangible: the front line of cyber espionage ran directly through the desk right next to mine.

That experience has shaped my perspective ever since. Today, the front line has expanded from a single desk to every home office, coffee shop, and airport lounge in the world. The traditional, brick-and-mortar perimeter has dissolved, rendered obsolete by a fluid borderless ecosystem of cloud applications, third-party vendors, and a distributed workforce.

This new reality creates a profound challenge for leaders. When the walls are gone, where does security begin? The answer is: identity. In a world without a clear perimeter, a user's identity is the one ››

While MFA remains an essential layer of security, we must now focus on the quality and assurance level of our authentication methods.

constant, the single control plane through which every access request must pass. It has become the heart of any modern enterprise security strategy.

Identity Is the Primary Target

This shift has not gone unnoticed by our adversaries. As the workforce moved remote, attackers adjusted their tactics accordingly. They saw with great clarity their new opportunity: users were at home, often on less secure networks, and their identity was the weakest link. As a result, identity became the most attacked vector. Credential theft and sophisticated phishing have evolved from fringe threats into the central tactics of the modern adversary.

Security teams can no longer simply block access from unapproved locations when legitimate work is happening everywhere. If a cybercriminal can steal legitimate credentials, they can gain unfettered access to critical resources with very little friction. This is one scenario that keeps CISOs up at night. As somebody once told me, phishing is the ultimate threat because if an attacker gets the keys to the kingdom, all other security controls — network, endpoint, and cloud — become irrelevant. They can simply walk in the front door.

Security that Serves the User

Confronting this reality requires us to evolve our defenses at the speed of the adversary. A few years ago, the conventional wisdom was that deploying any form of multifactor authentication (MFA) would solve the problem. A 2019 study famously claimed that MFA could stop 99% of phishing attempts.[1] But in the fast-moving world of cybersecurity, that advice is now dangerously outdated. Today, we know that traditional push- and SMS-based MFA are completely insufficient, as adversaries have developed sophisticated techniques to bypass them.

While MFA remains an essential layer of security, we must now focus on the quality and assurance level of our authentication methods. At Okta, we've centered our strategy on a modern approach that is both more intelligent and, crucially, more seamless. For the first time in my career, I can confidently say that we can significantly raise the security bar and improve the end-user experience simultaneously. The key is to move away from cumbersome, friction-filled authentication methods and embrace the phishing-resistant, biometric technologies that people already use every day.

What many don't realize is that a simple Face ID or Touch ID is, in fact, already multifactor. It combines something you are (your biometric) with something you have (your registered device), providing a high level of trust in a single, frictionless action. The goal is to create an experience where the speed bump of security feels more like a carriage return — you just do it, and you're in.

For the enterprise, we can take this even further. Beyond just phishing-resistant authentication, we can gather rich, contextual signals from the device itself, asking questions like: Is it a managed device? Does it have the right security posture? Is it integrated with our XDR solution? From there, we can build a complete picture of risk and make smarter access decisions behind the scenes. This is the philosophy behind our work at Okta: to provide truly secure, passwordless authentication that delivers a profound cultural win, transforming security from a blocker into an enabler.

Read the article on Perspectives. ◆

1. Melanie Maynes, "One simple action you can take to prevent 99.9 percent of attacks on your accounts," Microsoft Security, August 20, 2019.

Mastering the Art of Threat Hunting

BY RYAN CHAPMAN

Cybersecurity threat hunting is a hot topic these days, and it's also a high priority for CISOs and business leaders alike. The accelerated deployment of more and more threats — as well as the increasingly sophisticated nature of those threats — have turned threat hunting into an essential capability for cybersecurity departments and their organizations.

This has led to a buildup in spending for threat hunting tools and services. One study predicts the global market for threat hunting software and services will exceed $13 billion by 2033; this represents a 10-year compound annual growth rate of 18.6 percent.[1] But as much as those numbers represent organizations' willingness to make sizable investments in threat hunting technologies, there's an even bigger initiative underway: Current and future security professionals must have the right skills and mindset to be successful threat hunters.

Focus More on People-Centric Defenses

CISOs, IT executives, and C-suite executives need to prioritize training, education, mentorship, and a commitment to continuous improvement in their threat hunting teams. As a SANS author and instructor specializing in ransomware and other threats, I've spent years in threat hunting and digital forensics and have tried to pass along what I've learned to others in this field. I've also presented at numerous cybersecurity community events and conferences, where I've engaged with peers and newcomers alike to help further organizations' readiness and capabilities in threat hunting.

What have I learned that I find most important for cybersecurity professionals and their business stakeholders? Certainly, I've discovered and shared my insights on new and successful threat hunting tools and services, many of which have emanated from our Unit 42® team and many others that have come from a wide range of other sources.

But as important as these technical solutions are in our world, what's more significant — and sometimes undervalued and overlooked — is understanding what it takes to be a strong threat hunter. A big part of being good at threat hunting is to learn from others. I've learned most of what I know today from colleagues, peers, mentors, and others. We all stand on the shoulders of giants, and there's so much we can learn simply by talking with and observing the actions of those who have done this before. That's certainly one thing I try to do with my colleagues inside and outside Palo Alto Networks.

You go to a conference or take a training course, and you are immediately struck by important new takeaways about threats and threat hunting, and you build on that. For instance, let's say you come across a list of commands and there's one you don't understand. As an effective threat hunter, you need to take that one thing you don't quite understand and keep pushing until it makes sense. That may sound like a technical requirement: You have to understand things like command lines so you can determine what the threat actors are trying to do when they run various commands and why they're doing it. Beyond technology, you must be motivated to learn by asking questions, observing, challenging, and playing "what if" exercises. In essence, you have to put yourself into the head of a threat actor to understand their motivations, methods, and mindsets.

Another key area where cybersecurity professionals can improve their threat hunting will likely evoke a "really?" response for many of you—social media. Some of the craziest exploits and vulnerabilities emerge because of what's posted on those platforms because threat researchers, hackers, and cybercriminals are exceptionally proud of their work and can't help but share it. ››

Leveraging Tools and Technologies

Many valuable methodologies and technical resources are at our disposal, such as open-source intelligence (OSINT) and closed-source threat intelligence feeds. Using all the available tools, services, and technical resources enables you to stay abreast of the changing threat landscape to keep pace with this ever-evolving creature called cybersecurity.

New threats emerge all the time, which means you are likely confronted with things you don't quite understand and don't necessarily have a well-thought-out playbook on how to analyze and confront the threat. When the Lumma Stealer malware emerged a few years ago, it quickly gained momentum with the bad guys because it was simple to execute. The developers who provided the methodologies used by the threat actors made it easy for them to gain initial access via a new form of social engineering to speed and simplify delivery of payloads.

Dealing with these sophisticated, yet easily acquired and deployed, threats requires a cybersecurity mindset that doesn't rely exclusively upon new tools and playbooks, but on critical thinking. I like to do training exercises with security engineers where I show them a number of low-level technical screenshots of data movement and activity. Then, I ask them what they're looking for and what they think they see. And the most important thing I try to instill in them — or hone it if it's already there — is getting in touch with how their brain processes information.

Pattern Recognition

Critical thinking and thinking outside the box are among the top skills threat hunters need when chasing down the sources, causes, and intents of emerging threats. You might not immediately realize it, but there are often discernible patterns you can recognize from prior instances of similar threats. And, because threat hunting is a team sport, keep in mind that someone else on your team or in your professional circle has likely seen something similar, which can open up new possibilities for solutions.

We do the same kinds of things when looking at threat hunting tools. Most important is the telemetry the tool provides to give us the broad and deep visibility you need to come up with answers and responses to new threats. It's essential that these tools go above and beyond how a process is run or how a file was written. Instead, the right threat hunting tool — especially one in the hands of the right threat hunter with the right approach — looks at a wider array of issues. These might include when the process was run, the order commands were executed, or the overall parameters.

It's about looking at the big picture. And that requires a different approach by threat hunters who are looking to identify, block, and clean out threats at scale. By combining great tools, highly focused services, critical thinking, and a commitment to collaborative problem-solving, organizations can stop chasing threats and start creating an environment where threats are quickly and reliably spotted and squelched.

Read the article on Perspectives. ◆

1. *Threat Hunting Market Insights – Trends & Forecast 2023-2033* report, Future Market Insights, July 27, 2023.

paloalto NETWORKS

THREAT VECTOR

PODCAST

Follow and never miss an episode.

LISTEN HERE

Beyond the Backlog: Escaping Application Security Debt with ASPM

BY SARIT TAGER

While debt can be a powerful tool for growth, unchecked, it becomes a crushing burden. In cybersecurity, we have our own version of "security debt," and nowhere is this debt more acute than in application security. It's the growing backlog of vulnerabilities, misconfigurations, and software risks that we promise to fix later. For years, this model of finding issues in production and adding them to a remediation list has defined the industry, creating a list that only ever seems to get longer.

This model is fundamentally broken.

Today, two powerful forces are acting as relentless accelerators of this security debt. First, the sheer velocity of DevOps means code is deployed faster than ever. Second, the explosion of AI-generated code is set to dominate development. In fact, some predict that by 2030, AI could produce 95% of all code,[1] as AI coding assistants move from generating simple scripts to authoring complex application logic. And with research indicating that a third of that code may introduce security issues,[2] the scale of our security debt is poised ››

to skyrocket. This represents a paradigm shift of unprecedented scale, with the consequence that security vulnerabilities are now created at a speed and scale that completely outstrip any human-centric model of remediation.

The traditional approach of trying to identify issues late in the cycle is a losing battle. Statistics show that only about 10% of security issues in production are remediated each month.[3] This creates a costly cycle, as our data indicates that it takes, on average, 10 times longer to remediate an issue in production than at the source. The mistake is chasing risks instead of preventing them, and the interest on our security debt is compounding into unacceptable levels of business risk.

Prevention Powered by Context

To escape this cycle, companies must shift their entire philosophy of application security from reactive remediation to proactive prevention. The goal is to automatically prevent insecure code from ever reaching production, freeing developers to innovate faster by fixing issues efficiently during development instead of chasing them in production. Our data shows that by shifting left, teams can eliminate up to 92% of security issues before they reach production.

This goal is achievable, but it requires a new architectural approach built on a single, non-negotiable principle—using complete context to drive prevention. It's about both collecting data and using a unified understanding of your application's posture, from code to cloud. This enables the ability to craft more targeted prevention policies, prioritize risk with greater precision, automate remediation workflows and better connect security to business priorities.

Legacy security tools fail because they are too noisy and lack context. This often overwhelms developers with alerts on issues that might not be exploitable or critical, contributing to a sense of friction in the business, which then leads to the removal of important security guardrails. A more mature approach begins by prioritizing findings from native and third-party scanners. But a true prevention-first model achieves the highest level of maturity by intelligently correlating data from every source, from developer tools and application infrastructure all the way to cloud runtime environments.

Armed with this complete, code-to-cloud context, we can finally build intelligent and targeted prevention policies. By creating a single, correlated view of risk, we can build guardrails that are precise enough to automatically block the critical issues that truly matter before they are committed, while allowing other development to proceed without friction. This approach empowers AppSec teams to reduce application risk by preventing problems with surgical precision, and its efficacy is only getting stronger.

Security + Development to Pay Down the Debt

This context-driven, prevention-first model provides a dual benefit: It stops new risks and provides the tools to remediate existing backlog at scale. By creating that single view of application posture, teams can move beyond chasing alerts and begin intelligently prioritizing the security issues that pose a genuine threat, based on runtime behavior. Furthermore, by integrating security directly into developer workflows — delivering real-time feedback and automated remediation suggestions through integrations into the tools they use every day — we can unite security and development teams. This seamless collaboration streamlines the remediation of existing issues and ensures new ones are caught early, when they are fastest and cheapest to fix.

The goal is to transform security from a blocker into a business enabler. This modern, prevention-first philosophy is the driving force behind application security posture management (ASPM). By shifting left and using complete context to prevent risks, we pay down our security debt, reduce friction, and empower our developers to innovate safely at the speed the business demands. It is this philosophy that we have built into our own platform, to give every organization the power to secure innovation from code to cloud.

Read the article on Perspectives. ◆

1. "An Evaluation of Deepseek and How We Underestimate the Chinese (interview with Microsoft's Kevin Scott)," 20VC with Harry Stebbings via YouTube, March 31, 2025.
2. *Every 1 of 3 AI-Generated Code Is Vulnerable: Exploring Insights with CyberSecEval*, SOCRadar, March 18, 2024.
3. *The Fast and Frivolous: Pacing Remediation of Internet-Facing Vulnerabilities*, Security Scorecard and Cyentia Institute, June 8, 2022.

Is Your AI Well-Engineered Enough to Be Trusted?

BY AARON ISAKSEN

The cybersecurity industry is consumed with a number of philosophical questions, perhaps no one more pressing nowadays than "Is our AI ethical?" While this is an important conversation, it often misses a more pragmatic and urgent question that every business leader should ask first: Is our AI well-engineered enough to be trusted with our business?

A well-engineered AI system — one that operates with accuracy, honesty, security and responsibility — is the prerequisite for any AI that can be called ethical or trusted with our business. An AI that is biased, opaque or insecure is not an ethical dilemma. It is a poorly engineered system that presents a direct and tangible business risk.

My engineering-centric view, I believe, allows us to move beyond abstract debates and define the hallmarks of a trustworthy AI, using principles that any product designer or engineer is familiar with.

Well-engineered AI begins with a commitment to being **accurate** and **unbiased**. A model trained on incomplete data is a performance flaw. For example, if a malware detector was trained without any data on ransomware, its predictions would be dangerously ››

A trustworthy system is one that has been rigorously and relentlessly tested to uncover unforeseen risks before they can cause harm.

biased by omission, creating a critical security gap. A faulty system will inevitably produce flawed outputs, leading to poor business decisions.

This concept extends to being **transparent** and **honest**. While the industry currently relies on opaque black-box models, this lack of explainability introduces a critical operational risk. When a system we cannot fully explain fails, our ability to conduct effective forensics or build deep, verifiable trust is severely hindered. This is why government bodies and research institutions like NIST are heavily invested in creating new standards for AI explainability.

Underpinning this concept is the need for the system to be **safe** and **secure**. AI vulnerable to prompt injection, data poisoning or model theft is a catastrophic design flaw. The OWASP AI Security Top 10, for instance, treats these vulnerabilities as fundamental threats to the application layer. Because these systems require vast amounts of data, this insecurity creates a direct threat to privacy and data protection, turning the AI into a built-in vulnerability that can be turned against the enterprise it was designed to serve.

Finally, a well-engineered AI is **accountable** and **responsible**. There must be clear lines of ownership and a clear process for addressing any problems. The EU AI Act, for example, is built on this principle, establishing strict liability frameworks for the outcomes of high-risk AI systems. This ensures that, when a system makes a mistake, there are humans who are responsible for the outcome who can create a necessary accountability framework that is essential for managing high-impact decisions.

If you are uncertain if these traits are necessary, consider a system that has opposite traits. After all, would you trust a system that was inaccurate, biased, opaque, dishonest, unsafe, insecure, unaccountable or irresponsible with your business?

Blueprint for Building Trustworthy AI

Achieving this level of engineering excellence requires a disciplined philosophy that moves beyond the academic debate. This is why Palo Alto Networks rejects the "ivory tower" model of research. Building trustworthy AI requires embedding security and integrity into every phase of the development lifecycle.

This journey begins with an obsessive focus on the integrity of the AI supply chain. It demands a clear-eyed understanding of the risks inherent in open-source models, which, for all their innovative potential, can be fine-tuned for malicious purposes. It means engineering systems from the ground up that are resilient to threats like prompt injection.

From that trusted foundation, we build a culture of assurance. This requires a serious investment in robust model evaluation, explainability and continuous red teaming — the capabilities that global leaders are now calling for in new "AI Centers of Excellence." A trustworthy system is one that has been rigorously and relentlessly tested to uncover unforeseen risks before they can cause harm.

The New Standard: Trust as a Function of Quality

Ultimately, building trustworthy AI is the definition of good engineering in the 21st century. It is about building products that are robust, reliable and secure. The true measure of "well-engineered AI" in a business context is its quality and integrity. If you can trust its security and performance, you can trust it with your business.

Read the article on Perspectives. ◆

The AI Imperative: Security Designed for Trust, Control and Cooperation

BY NICOLE NICHOLS

Remember the early days of generative AI? Just a few years ago, when the first powerful models were released, some labs restricted access out of a fear they might be misused — a caution that, at the time, seemed almost quaint. The models were novel but often flaky, their outputs grainy, and their real-world applications limited. Today that caution seems prophetic. The maturity and capability of these systems have progressed at breakneck speeds, moving the conversation from a theoretical debate about future risks to an urgent, practical question: How do we maintain security controls?

While this question touches on age-old debates about powerful technology, the stakes are entirely new. We are at a similar nexus point of unknown harms and immense possibilities, much like when we used unshielded X-ray machines to size shoes, blind to the long-term risks. While much of the industry is consumed with what AI can do, this focus on capability overlooks the more foundational challenge — establishing clear and enforceable rules of security management for these autonomous systems.[1] ››

For decades, the ethos of Silicon Valley was (and, to some extent, still is): "Move fast and break things." That model, for all its generative power, is untenable when dealing with a technology that can autonomously generate novel attacks. The potential for widespread, irreversible harm demands a new philosophy, one grounded in deliberate, thoughtful control.

Defining the Rules of Engagement

The only way to safely deploy powerful, cybercapable AI is to begin with a new social contract, one I call the "AI Imperative." This is a clear, technical and operational compass for AI purposes, defining its explicit boundaries and prohibited uses. It requires rigorous, upfront offensive and defensive capability evaluations to understand a model's potential for weaponization before it's ever released.

This imperative must be the foundation of evaluating the entire AI lifecycle. It must inform the integrity of the AI supply chain — the digital concrete and steel of our systems. This imperative must be the benchmark against which internal and external expert red teams test the system for hidden vulnerabilities, particularly for systems deemed critical infrastructure. And it must be the standard against which we conduct independent validation before a single line of code is deployed.

Non-Negotiable: An Architecture of Control

Yet, these principles are meaningless without enforcement and alignment with technical measures and controls. The second, and most critical, component of this framework is a robust architecture of control, built on the non-negotiable ability to revoke AI's access the moment it acts outside its established bounds.

This capability must be architected into the fabric of our systems. An architecture of control requires a steadfast commitment to transparency, where access to the most powerful capabilities is controlled. It demands new standards of authentication and attestation that can verify interactions across a complex ecosystem of agents. And it necessitates a dedication to human-in-the-loop governance for high-stakes situations, ensuring that ultimate accountability always rests with people, not an algorithm.

> The wisdom of the controls we place on AI — not the power of the AI we build — will define the legacy we create.

A Call for a New Standard of Control

This challenge transcends any single organization.[2] While society must debate the ethical "redlines" — for instance, whether AI should ever autonomously manipulate critical infrastructure — our imperative as technologists is different. It is to pioneer the technical measures and controls that make enforcement of any rule possible. This requires a new, more radical form of collaboration to collectively build the foundational architecture for AI safety.

This radical collaboration is necessary because AI security controls are a shared cost center. Consumers and enterprises purchase products for their features, not necessarily for their safety constraints. It's unlikely a decision is made to buy one car over another solely because of the seat belts; yet, seat belts are still a non-negotiable aspect of auto safety standards. The complexity of creating these "AI seat belts" makes them a nontrivial engineering challenge, and the universal risk of a catastrophic failure means no single entity can or should bear this burden alone. This is precisely why this effort must be shared, making collective defense an economic and security imperative.

The wisdom of the controls we place on AI — not the power of the AI we build — will define the legacy we create. This work begins with a concrete first step — a shared commitment to establish a common framework for assessing a system's power and the technical levers to moderate that power when deployed. This is the hard, necessary work, yes, but it also ensures a safe, AI-enabled future.

Read the article on Perspectives. ◆

1. McGregor, Sean, and Kathrin Grosse, "When it comes to AI incidents, safety and security are not the same," OECD.AI, August 25, 2025.
2. U.S. AI Safety Institute, *SP 800-53 Control Overlays for Securing AI Systems Concept Paper*, NIST, August 13, 2025.

paloalto NETWORKS | CYBERSECURITY PARTNER OF CHOICE

DISCOVER PERSPECTIVES

Unfiltered security intelligence. C-suite insights.

EXPLORE INSIGHTS