Velocity

V5 2026

Booz Allen

INSIGHTS FOR INNOVATORS

AI vs AI

REIMAGINING CYBER FOR A FASTER FIGHT

The future of cybersecurity will be shaped by a relentless contest between AI-native defense systems and AI-enabled adversaries operating at machine speed, constantly probing for vulnerabilities and creating disruption at scale.

Competing Against Time: Secure by Design and Intent

Thirty years ago, I began my career as a software developer. Back then, cybersecurity was little more than an afterthought. We built systems designed to live safely behind fortified network perimeters, trusting firewalls and isolated infrastructure to keep threats at bay. The prevailing mindset was simple: protect the outer walls, and everything inside would remain secure.

Nearly everything about technology has changed since those early days, but one transformation stands above the rest—the walls are gone. Security can no longer be treated as a gate at the edge of the network; it must be embedded into every layer of the infrastructure and the software itself. Today's systems are not confined to controlled environments. Applications span cloud platforms, mobile devices, agentic systems, and globally distributed users. Data moves constantly beyond traditional boundaries, and attackers no longer need to breach a perimeter that barely exists. And the United States faces a rapidly evolving adversary accelerated by AI—one capable of delivering cyber effects with unprecedented speed, scale, and sophistication.

The future of cybersecurity will be shaped by a relentless contest between AI-native defense systems and AI-enabled adversaries operating at machine speed, constantly probing for vulnerabilities and creating disruption at scale.

It's easy to admire the problem we now face—namely, time. When attackers can move faster than defenders, they control the fight. And if our industry keeps building and defending systems in the same way as before, time will not be on our side. Instead of automating old defensive systems, the pressure is on to step back and reimagine the fundamental ways we compete in this new, amorphous cyber battlespace.

In this edition of Velocity, our cover story (page 30) examines this shift in cyber as tradecraft: why government and industry can't simply fight AI-enabled threats with AI tooling but must also adopt a new posture—what we describe as "attack to defend." This means embracing a forward-thinking operating model that can intelligently assess your environment, understand gaps, and proactively mitigate. As Bill Vass, Booz Allen's Chief Technology Officer, states in our "Lessons from the Edge" column (page 64), "Failure isn't the exception. It's the normal operating condition."

We believe that shifts of this magnitude can become breeding grounds for innovation. In fact, Booz Allen recently disrupted our entire approach to fighting nation-state level threats with the development of agent-powered cybersecurity to close the speed gap. There are many more paradigm-shifting examples from across the industry that our featured authors explore.

While each of us navigates this threat landscape, I invite you to engage with the range of issues we cover in this collection, including:

- How the right engineering practices can help design, deploy, and de-risk agentic AI systems when critical mission outcomes are at stake (page 40)
- Zero trust's renewed importance given the rapid erosion of the barriers separating operating technology and IT, and the private and public sectors (page 54)
- What's on the horizon for defense technology investment and innovation, in a special conversation with Raghu Raghuram, managing partner at a16z (page 4)
- The evolution of zero-knowledge proofs from a niche cryptography concept into a serious methodology for national security (page 10)

The challenges ahead aren't limited to one mission or one industry: from the adversary's vantage point, it's all one big attack surface. For defenders, this means we need to step out of old comfort zones and lean into new operations, partnerships, and cyber automation at scale. While the response window is shrinking fast, organizations that commit to operating at an AI cadence can still outpace the threat. If you take away one thing from this issue, I hope it's the fact that both government and industry are well-positioned to operate faster, respond decisively, and contain impact across one battlespace.

Thank you for joining the conversation.

Brad Medairy
President, National Cyber
Booz Allen

CONTENTS

INFRASTRUCTURE TO Impact

a16z's **Raghu Raghuram** on AI's Next Frontier, Defense Tech, and Why the Future of Innovation Runs Through Government

Raghu Raghuram has spent three decades at the center of nearly every major inflection point in enterprise technology, from the browser wars at Netscape to building VMware into a defining force of the cloud era and guiding it through the largest software exit in history. Now, as a managing partner at Andreessen Horowitz (a16z), he's applying that experience to what he sees as the next industrial-scale shift: artificial intelligence.

In this conversation with **Bryce Pippert**, Executive Vice President of Ventures and Partnerships at Booz Allen, Raghuram reflects on why this AI wave is fundamentally different, how sovereign AI is reshaping national strategy, and what it will take for Silicon Valley and government to build together again.

Q Bryce: After VMware's acquisition by Broadcom closed, you had options. What made this the right moment to join a16z?

A Raghu: We had a successful run at VMware, and I was fortunate to work on cutting-edge technologies across a broad range of areas that had significant industry impact. I could have very easily called it a day, right?

But if you look at what's going on in the world, there has never been a greater time for technology to have a positive impact when so many things that we imagined were science fiction have turned into reality.

So, several factors led me to a16z. The first question for me was: Where could I have that same kind of impact again?

Second, this felt like too important a time to sit out. AI is on the scale of the next Industrial Revolution, and I wanted to be involved in as many areas of AI as possible. As the CEO of a single company—unless it's one of the very largest—you're often focused on one particular slice of the market. Ventures gave me the opportunity to engage across a broad set of technologies and applications.

And third, it was the team. Venture capital is often described as an individual sport. a16z operates very differently, as a team sport. That resonated with me. Throughout my operational career, I loved building teams and winning together. That model, combined with the chance to work on transformative technologies at scale, made it the right move.

Q Bryce: You've led at enormous scale. From the venture side, what does impact look like now?

A Raghu: It's breadth and leverage. You're not just helping build one company—you're helping shape multiple companies that could each redefine an industry. And in AI's case, the surface area is extraordinary. That combination of breadth and industrial impact is what excited me most.

Q Bryce: You spent two decades at VMware building infrastructure for the cloud era. How does that experience shape the way you think about AI infrastructure now?

A Raghu: I would say it's a plus and a minus. The plus is understanding the core infrastructure technologies—what does it take to productize a deep tech stack, and what does it take to recognize really smart talent in this space . . .

Q Bryce: But you're also careful not to assume the same playbook applies here.

A Raghu: Exactly. Both from a technology and impact point of view, AI is so different from anything that came before it. You have to discipline yourself to think about everything from first principles. You can't pattern-match and say, "In the cloud this

happened—therefore it will happen here." That could lead you down the wrong road.

In many ways, the closest historical analogy might be something like electricity—an enabling capability that ends up transforming nearly every industry and institution. So, while the experience helps in understanding infrastructure and productization, you have to be careful not to assume that the same playbook applies.

Q **Bryce: One place that difference is becoming especially clear is around sovereign AI. Governments everywhere are starting to think about national infrastructure for AI. What are you hearing from leaders about why that matters?**

A **Raghu:** We spend a lot of time talking with country leaders about this. I was also involved in the earlier conversations around sovereign cloud, and the contrast with AI is striking. With cloud, governments realized they needed a certain level of independence and control over their data and infrastructure. But with AI, the stakes are much higher.

But going back to the analogy of the Industrial Revolution—whether it was the Industrial Age itself, the transistor, or other foundational technologies—the countries that controlled those technologies captured disproportionate economic advantage. That advantage ultimately translates into military capability and geopolitical influence. All of that applies in the AI space. So, if you're not able to have a say in how AI is built or adopted in your country, it becomes a material weakness for the long-term security of the nation.

AI models also reflect the data they're trained on. Early models were trained by scraping the web, which means the distribution of ideas, perspectives, and cultural norms embedded in that data shows up in the models themselves. For countries whose culture, history, or values aren't well represented in those datasets, that becomes a real issue. They want their own models that reflect their own societal perspectives.

Q **Bryce: a16z has made a number of investments in defense technology through the American Dynamism Strategy. How do you think about national security in the context of emerging technologies like AI?**

A **Raghu:** The foremost thing to understand about a16z is that we are massively long on America. Clearly, we believe in building great companies and delivering strong returns for our investors. But a third pillar of how we think about the world is strengthening the United States and the countries that are allied with it.

AI is one of those technologies that is instrumental both to national economic strength and national security. It also plays a role in the propagation of culture and values. That's why we've invested heavily in strengthening the American defense industrial base through the American Dynamism Strategy.

Q **Bryce: You mentioned earlier how the nature of warfare itself is changing. What are you seeing there?**

A **Raghu:** If you look at the technology industry, we've seen waves of commoditization reshape how computing was delivered, from mainframes to client-server systems to commodity servers and eventually the cloud. You're starting to see something similar in defense. Commercial off-the-shelf technologies are increasingly being used on the battlefield. The clearest example is what's happening in Ukraine with drones and other systems being built and deployed in completely different ways.

When you combine commercial technologies with massive intelligence—AI, autonomy, and data—you have the opportunity to rethink how you defend or project power on the battlefield. Autonomy and AI become a huge part of that, creating opportunities to challenge industries that have had very strong incumbents for a long time.

Q **Bryce: So, what separates the defense tech companies that will actually matter from those that fade away?**

A **Raghu:** Some things stay the same. If you look at what made the current incumbents great, they combined cutting-edge innovation with tremendous execution and a strong focus on delivery. That triangle doesn't really change.

But the nature of innovation is evolving. Companies now can take advantage of the best commercial technologies available and combine them with AI to drive down costs while increasing capability. Execution still matters enormously. These systems are used in environments where lives are at stake, so reliability and quality are critical.

And finally, there's delivery—actually getting these technologies into large organizations like governments that have a lot of constraints and requirements. You must understand how those institutions work and how to translate innovation into something they can actually deploy.

Q **Bryce: You've seen firsthand how difficult it can be for promising defense technologies to move from prototype to real deployment. How does a16z think about helping companies make that leap from innovative technology to something that actually gets adopted at scale?**

A **Raghu:** If you think about what it takes for any technology to succeed, people tend to focus on the innovation. But execution is just as important. Innovation is the starting point, but the real challenge is making that innovation work in the real world—getting it into the hands of customers and making sure it actually delivers value. There's a huge gap between those two things. Most startups get funded because of the innovation, but the reason technologies become enduring businesses is because of execution. From the very beginning, a16z has focused heavily on that execution side.

Q **Bryce: How does that show up in the way the firm operates?**

A **Raghu:** If you look at the composition of our team, the majority of people at the firm are focused on helping our portfolio companies operate, execute, and succeed in their markets. Internally, we often describe what we do with four words: "find," "pick," "win," and "build."

"Find, pick, win" refers to the core venture work: finding great companies, selecting the ones we believe in, and earning the opportunity to invest in them. The fourth is "build," and the build is helping those companies become global leaders. It creates a virtuous cycle. We believe the more we help companies succeed, the easier it becomes for us to win the opportunity to partner with the next generation of founders.

Q **Bryce: What does "build" look like in practice for your portfolio companies?**

A **Raghu:** It can take many forms. Sometimes, it's helping companies find the right talent. Sometimes it's helping them find customers. It also involves helping them build the right partnerships so they can deliver their technology into real markets. That's an area where we've been putting more emphasis recently—helping portfolio companies connect with partners who can help them reach customers and deploy technology at scale.

Q **Bryce: You mentioned partnerships as a key part of helping companies execute. That's something we've been working on together. Why was Booz Allen the right partner for this moment?**

A **Raghu:** If you go back to what we talked about earlier, we are massively long on America. We believe one of the most important things we can do in acting on that belief is strengthening the capabilities of the U.S. government—defense, civilian agencies, and the broader national security ecosystem—by helping bring the latest and greatest technology into those institutions.

If you think about innovation, execution, and delivery, our role as a venture firm is primarily around innovation and helping companies execute. We know how to recognize great innovation, and we help companies build and scale. But we're not built to deliver those technologies directly into government.

> "I've never been more excited about technology than where we are today."
>
> —Raghu Raghuram

That's where Booz Allen comes in.

You have a long history of helping government organizations adopt and implement new technologies. You're trusted across multiple parts of the U.S. government and across allied governments as well. And you understand how to translate cutting-edge innovation into systems that institutions can actually deploy. So, the partnership was very natural for us. We can help identify and support the most innovative companies, and by working with Booz Allen, those companies have a path to deliver their technologies into real mission environments.

Q **Bryce: Can you give me an example of how we're helping a portfolio company get to market faster?**

A **Raghu:** One early example is our newly announced investments in Ulysses, a maritime robotics company backed by a16z's American Dynamism Practice. While the partnership is just beginning, the thesis is clear: Ulysses brings breakthrough autonomous surface and undersea systems, and Booz Allen brings the mission understanding and pathways that promise technological reach into national security customer sets faster.

Your role is to ensure Ulysses's autonomy platforms are aligned to real operational needs from the start—everything from mission engineering insights to navigating government adoption. By combining their innovation with your deep experience delivering emerging tech into mission environments, we're creating the conditions for Ulysses to accelerate toward fielded impact much sooner than they could alone.

Bryce: Exactly. Ulysses brought the autonomy; we focused on mission engineering and accelerating deployment pathways.

"In many ways, the closest historical analogy (to the impact of AI) might be something like electricity—an enabling capability that ends up transforming nearly every industry and institution." —Raghu Raghuram

Q Bryce: Looking ahead, what would success for that ecosystem look like in five years?

A Raghu: If you look back at the early days of Silicon Valley, many of the most important innovations had the U.S. government as their first customer. Defense and other government agencies played a major role in driving early adoption of breakthrough technologies.

Over time, that connection weakened. And that's not a criticism, but just a fact of the matter that Silicon Valley doesn't always think of government as the first place to deploy innovation, and government doesn't always look to the startup ecosystem as its first source of new technology.

If we look five years ahead, success means reversing that dynamic. It would mean showing clear examples where some of the best technologies coming out of the startup ecosystem are successfully deployed inside the U.S. government, because the right partners came together to make that happen.

If Booz Allen, a16z, and the companies in our portfolio can demonstrate that model working, that is a very powerful outcome.

Q Bryce: Also looking ahead a few years, let's shift back to AI for a moment. How do you see agentic AI shaping mission environments?

A Raghu: If you look at the waves of AI before the current one, the early applications were largely predictive. AI could analyze data and predict outcomes. For example, whether a financial transaction might be fraudulent or whether a particular pattern of behavior might indicate a security threat.

The next wave was generative AI. With systems like ChatGPT, AI became capable of creating new content, such as writing text, generating code, producing images, and so on.

More recently, we've seen a third stage emerge, which is reasoning. AI systems are becoming capable of working through problems in a more structured way, like solving math problems, analyzing complex questions, and generating more sophisticated outputs.

The next frontier is action. Once a system can understand a problem and generate solutions, it can begin to act. That's where agentic AI comes in. An agent is essentially a piece of software that can reason about inputs and then take actions in the digital world on your behalf. It might monitor information, trigger workflows, or execute tasks automatically. Initially, those agents will operate with humans supervising them, but over time, you'll see agents trusted to carry out actions independently.

Q Bryce: What does that mean for government organizations and mission environments?

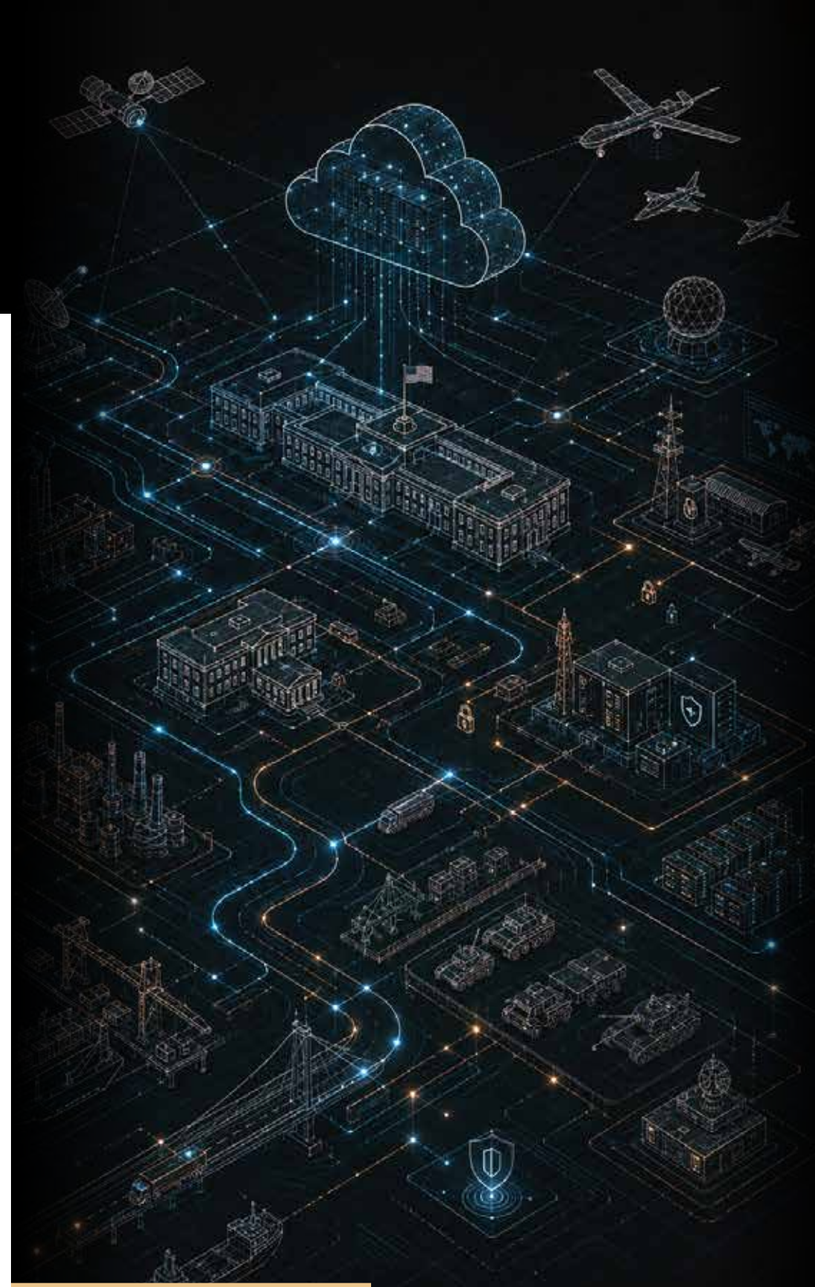
A Raghu: Governments face the same pressures as any large organization: the need to move faster, operate more efficiently, and continuously innovate. Agents can help address all three. If you look at coding agents, for example, they can help build software faster while lowering the cost of development. And because AI can analyze large volumes of existing code, it can often identify new ways to approach problems. You can imagine similar applications across nearly every category of knowledge work.

Government agencies have enormous backlogs of tasks they want to accomplish. They're also constantly balancing budget pressures and the need to innovate. Agents have the potential to help address those challenges.

Q Bryce: Beyond AI itself, are there other emerging technologies that excite you when you think about national security and defense missions?

A Raghu: Many of the most exciting developments are derivatives of AI. One example is robotics. Robotics has existed for a long time, but, traditionally, robots were limited to fixed, repetitive tasks. A robotic arm might pick something up here and place it there, but it couldn't adapt very easily. With AI integrated into robotics, those systems can learn and respond dynamically to real-world environments. That expands what robots can do, both commercially and in mission environments.

Another area is crypto and decentralized networks. As AI systems become more autonomous, they'll need ways to transact and interact with other systems. Today, for example, an AI system can't easily hold a bank account or execute financial transactions. Technologies like crypto provide mechanisms for identity and trusted transactions between software systems. In that sense, decentralized networks could become an important part of how AI systems operate and coordinate with each other.



LIGHTNING round

Bryce: Finish this sentence: The United States will maintain its technology edge if...

Raghu:...if you let the industry cook.

Bryce: One myth about defense tech that the Valley needs to retire?

Raghu: Modern defense tech is not science experiments but is being propelled forward by significant growth companies that are providing thousands of jobs across the United States and are producing drones, sea vessels, missiles, and other critical technologies at scale and speed.

Bryce: And one myth about Silicon Valley that government needs to retire?

Raghu: Big Tech is not representative of all tech. In fact, most builders in Silicon Valley are small teams with ambitious entrepreneurs tackling big problems, taking on enormous risk, and operating in intensely competitive environments.

Bryce: What is one a16z portfolio company in defense national security that Velocity readers should know about?

Raghu: Cursor!

SPEED READ

Former VMware CEO and current a16z managing partner Raghu Raghuram explains why today's AI wave represents a fundamentally different, industrial-scale technological shift—one he compares more to electricity than to prior cloud eras.

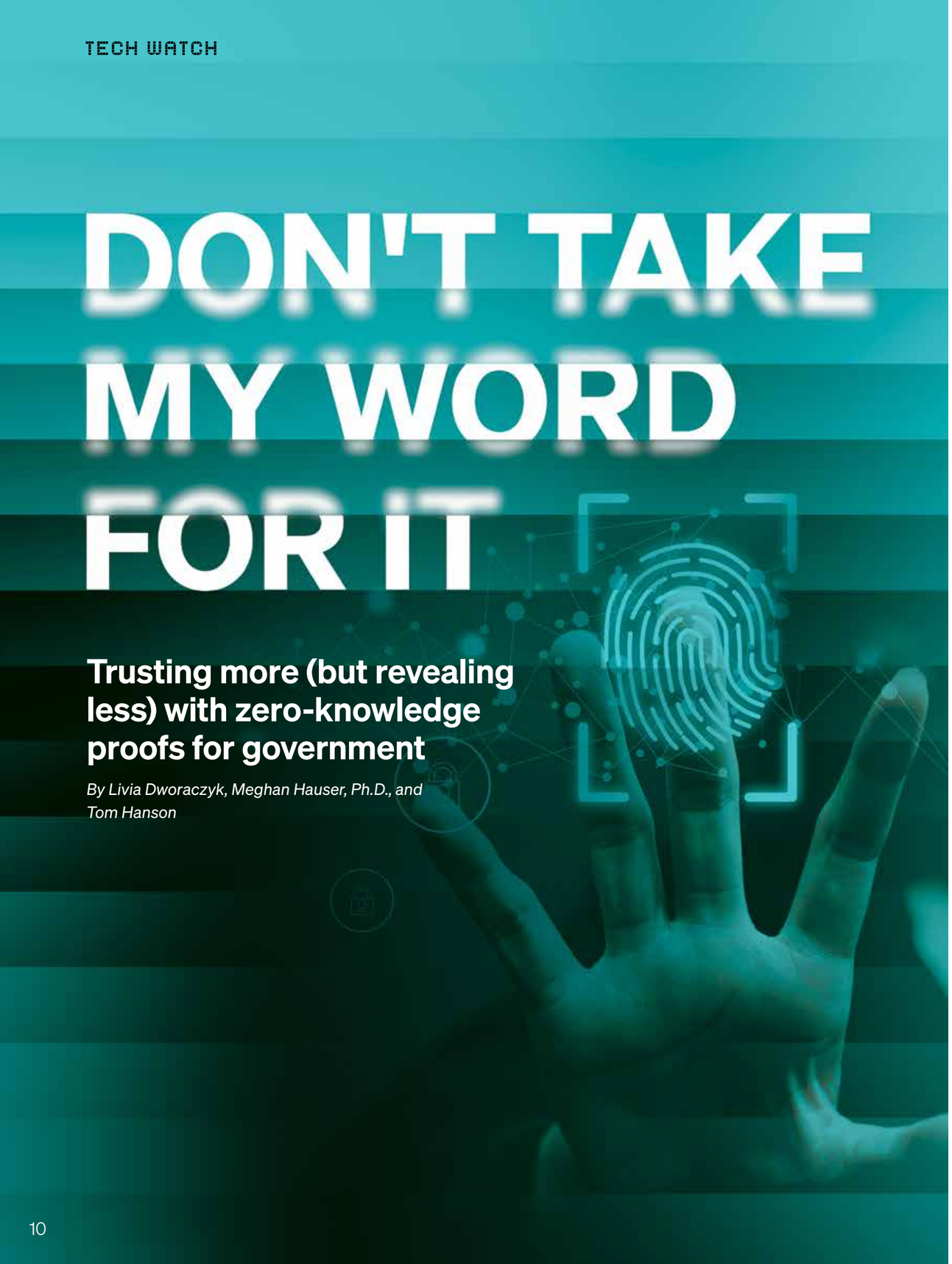
Sovereign AI, national security needs, and shifting defense technologies are reshaping the relationship between Silicon Valley and government.

a16z is partnering with Booz Allen to accelerate government adoption of emerging technologies from breakthrough innovation to mission ready deployment.

DON'T TAKE MY WORD FOR IT

Trusting more (but revealing less) with zero-knowledge proofs for government

By Livia Dworaczyk, Meghan Hauser, Ph.D., and Tom Hanson



When was the last time you had to prove that you were you? From filing taxes to checking in at the airport, it probably wasn't very long ago. Most of us are used to handing out personal data to simply participate in society.

Address. Birthday. Social Security Number.

Verification is exploding everywhere. Meanwhile, the world of cybersecurity is wrestling with a philosophical contradiction as two opposing priorities come to a clash: "We need more proof!" versus "We need more privacy!"

Individuals and corporations are seeking more data privacy for good reasons. In the World Economic Forum's Global Cybersecurity Outlook 2026, 73% of survey respondents reported that they or someone they know was personally affected by cyber fraud or attacks in 2025.

But beyond individual concerns, the discord between proof and privacy is playing out on a national and global scale. For example, intelligence agencies need to share data with coalition partners without compromising sources and methods. Intelligence officers may need to prove AI outputs are trustworthy without revealing classified models and logic.

The next big wave in security is to accommodate this contradiction by decoupling verification (proving that something is true) from disclosure (revealing the underlying information that makes it true). From the Defense Advanced Research Projects Agency to the National Institute of Standards, researchers and developers across U.S. government and industry partners are turning to a mathematical concept introduced over 40 years ago: zero-knowledge proofs (ZKPs).

This cryptographic method has scaled throughout the blockchain industry over the past decade, enabling private transactions on a public ledger. In the coming years, we expect the same methods to profoundly change what's possible in adversarial, data-sensitive environments at the national level. With analysis and insights from Booz Allen's Tech Scouting and Intelligence team, we examine where the market trends are heading and why ZKPs matter now for federal government leaders and the U.S. security industry.

Technology at a Glance

ZKPs are cryptographic methods that let you prove that you know something or meet a given requirement without showing exactly how. Let's illustrate this idea with a low-stakes example.

Imagine you're trying to rent a car. The rental company requires drivers to be over 21 years of age to avoid extra fees. But when you show your driver's license, you end up revealing a lot more than that: your full name, exact date of birth, home address, license number, height, eye color—basically a full identity profile. The agent doesn't need most of this information, yet now they've seen it, and you have no control over how the data is remembered, copied, or potentially misused.

A zero-knowledge version of this interaction would let you prove "I am over 21" without revealing anything else—not your name, not your birthday, not even how old you are, just that you meet the requirement. Instead of handing over a physical driver's license, it's possible to generate a ZKP that you possess a valid, DMV-issued credential and that the age encoded in that credential is over 21, without revealing any other personal details. The agent gets a cryptographic "yes" or "no," and you keep the rest of your personal data private.

Now consider zero-knowledge transactions for government missions, where information is often very sensitive. With ZKPs, it's possible to prove clearance levels without exposing personnel files, prove software integrity without revealing source code, prove compliance without sharing raw logs, and much more.

The shift to proof without disclosure, for illustrative purposes only.

How ZKPs Work

ZKPs mark the shift from disclosure-based verification to mathematics-based verification. Instead of sharing raw data, a prover generates a mathematical proof that a statement about the data is true.

There are three properties that fundamentally define ZKPs:

1. Completeness: If the statement is true and both parties follow the protocol correctly, an honest verifier will be convinced by the honest prover.
2. Soundness: If the statement is false, no dishonest prover can convince the honest verifier that it is true except with negligible probability.
3. Zero-knowledge: The verifier learns nothing beyond the fact that the statement is true, meaning no additional information about the underlying secret is revealed.

So, how do ZKPs work? Under the hood, ZKPs rely on cryptography to separate verification from disclosure. Here is the high-level sequence of this methodology in action: The prover commits to some secret information (e.g., your actual birthdate) in a way that is binding (you can't change it) but hidden (no one can see it). The proof encodes a logical statement about that hidden data (e.g., "this date is more than 21 years prior to today's date"). Finally, a verifier can check the proof using public parameters and cryptographic rules, but—critically—cannot reverse the proof to learn the underlying data.

Common Types of ZKPs

The most widely deployed ZKP systems today are Succinct Non-Interactive Arguments of Knowledge (zk-SNARKs) and Scalable Transparent Arguments of Knowledge (zk-STARKs). These categories are useful because they capture high-level system properties—such as proof size, verification cost, and setup requirements—and they are often how ZKPs are discussed in practice. However, they are not defined by a single underlying cryptographic assumption, and so they do not form a clean, mutually exclusive taxonomy.

Let's start by looking at zk-SNARKs, which emerged earlier. Many widely used constructions such as Groth16 and PLONK rely on bilinear pairings over elliptic curves. In these systems, security is grounded in hardness assumptions from elliptic-curve cryptography, including discrete logarithm-type problems and pairing-based assumptions. These proofs are extremely small and fast to verify, which has made them attractive for blockchain applications. In the driver's license example, this is what makes the interaction practical: A user can prove they hold a valid DMV-issued credential and are over 21 with a succinct proof that the verifier can check instantly. However, pairing-based SNARKs typically require a trusted setup (though there are variants with universal or updatable setups) and are not believed to be resistant to quantum attacks—a limitation that will become increasingly important as quantum computing advances.

A BRIEF HISTORY OF ZERO-KNOWLEDGE PROOFS

1980s

The first ZKPs were introduced as a theoretical and academic breakthrough.

1990s

Niche security applications were developed, but they were computationally expensive and impractical at scale.

2010s

Succinct non-interactive ZKPs emerged. The blockchain boom forced real-world engineering; zk-SNARKs and zk-STARKs matured under economic pressure.

2020s

Tooling and performance have become viable beyond crypto. What was once niche cryptography is now production-grade verification infrastructure.

SEQUENCE OF VERIFICATION WITH ZKPS

Prover and verifier agree on the rule to be proven

Before anything happens, the prover and verifier agree on:

- The statement to be proven (for example, "This AI model ran within approved constraints")
- The rules or policy that define what "correct" means
- The verification method to be used

No data is shared at this stage.

Prover commits to the sensitive information

The prover takes the sensitive inputs (w) and creates a cryptographic commitment, which:

- Prevents (w) from being changed later
- Hides (w) from the verifier
- Ensures the prover cannot cheat by swapping inputs after the fact

Think of this as creating a tamper-evident seal on the information.

Prover generates a ZKP

Using (w) and the agreed-upon rules, the prover runs a proof-generation process that produces a ZKP, which mathematically demonstrates that:

- (w) satisfies the rule
- The prover knows (w)
- No additional information about the data is revealed

Importantly, the proof does not contain (w) itself, the model logic, or internal system details.

Prover sends the proof to the verifier

The prover sends:

- The ZKP
- A reference to the rule or statement that was proven

The prover does not send the underlying sensitive information (w).

Verifier checks the proof

The verifier uses a verification algorithm to mathematically check the proof. This check:

- Does not require trust in the prover
- Does not depend on the prover's hardware, software, or environment

If the proof verifies, the statement is true; if it does not, the statement is false or invalid.

Verifier acts on the result

Based on the result, the verifier can:

- Grant or deny access
- Accept or reject data
- Approve or fail an audit
- Allow or block a system interaction

All without ever seeing or storing (w).

KEY TERMS:

Prover: The system or party that has the sensitive information and wants to prove something about it.

Verifier: The system or party that wants assurance the claim is true, but does not want—or is not allowed—to see the sensitive information.

(w): The sensitive inputs (e.g., data, model parameters, logs, identity attributes).

By contrast, zk-STARKs are designed to be transparent and avoid trusted setup entirely. Constructions such as Aurora and Ligero rely on hash-based cryptography and techniques from error-correcting codes. Their security ultimately depends on the collision resistance of cryptographic hash functions—a standard, conservative assumption in modern cryptography. These functions have been extensively studied and are widely believed to make it computationally infeasible to find two distinct inputs that produce the same hash output, even for quantum-capable adversaries. This makes such zk-STARK constructions plausibly secure against quantum adversaries. STARKs also scale well to very large computations, though they tend to produce larger proofs and have higher verification costs compared to pairing-based SNARKs.

As alluded to above—where we introduced the underlying assumptions behind common SNARKs and STARKs—a more fundamental way to classify ZKPs is by these underlying hardness assumptions rather than by system-level properties like "SNARK" or "STARK." From this perspective, most practical systems fall into the following categories:

- Pairing-based/elliptic-curve-based, which underpin many SNARKs
- Hash-based, which underpin STARKs and other transparent proof systems
- Discrete-log–based, which include protocols built directly from sigma protocols and related techniques

STARKs also scale well to very large computations, though they tend to produce larger proofs and have higher verification costs compared to pairing-based SNARKs.

Other families—such as lattice-based, isogeny-based, and code-based ZKPs—are also active areas of research, particularly in the context of post-quantum cryptography, although they are generally less mature for large-scale, general-purpose proving.

In short, "SNARK" and "STARK" describe how a proof system behaves, while hardness assumptions explain why it is secure. The two lenses overlap but are not interchangeable, so understanding both is key to navigating the design space of modern ZKPs.

Spotlight Application: Zero Trust

Zero trust (ZT) security frameworks assume that networks are hostile, identities can be compromised, and no request should be implicitly trusted based on location or prior authentication. Every access decision must be continuously verified and limited to the least privilege necessary.

However, traditional verification mechanisms often require over-disclosure of sensitive information. To prove eligibility, compliance, or identity, users and systems typically transmit full records, credentials, logs, or configuration data. This creates unnecessary data exposure and expands the attack surface.

If ZT means never trust and always verify, then the next logical question for security teams is: Can we verify without disclosing?

From Identity Assertions to Cryptographic Proofs

ZKPs provide the cryptographic backbone for the next phase of ZT: enabling verification without disclosure, continuous assurance without surveillance, and mission execution in environments where no single node or operator can be fully trusted. By shifting from data verification to mathematics, systems don't need to evaluate user attributes but can still prove those attributes satisfy enterprise policies.

With ZKPs, no raw attributes are transmitted. No centralized identity query is required. No additional metadata is revealed. And verification can happen locally, to minimize exposure.

Beyond identity and access control, ZKPs extend ZT principles to the computation itself. For example, in untrusted or outsourced environments such as cloud services or third-party processing systems, ZKPs could soon enable verifiable computation, allowing a provider to prove that a workload was executed correctly according to previously agreed-upon rules. This shifts trust from infrastructure and contractual assurances to cryptographic guarantees at the application layer.

Taken together, these capabilities align directly with ZT's core requirement: constant verification under conditions of assumed compromise. ZKPs provide the cryptographic primitive that allows such verification to occur without expanding exposure. They enable proof without revelation, reduce centralized data risk, support privacy-preserving identity, and allow both compliance and computation to be verified in adversarial environments. In the event of a data breach, ZKPs can reduce the "blast radius" by minimizing the amount of data collected and stored, so that any incident exposes only limited or unusable information rather than full sensitive records. For modern distributed systems, particularly those spanning multiple organizations or regulatory domains, this capability makes ZKPs a powerful enabler of next-generation ZT architectures.

THE ARCHITECTURAL SHIFT OF ADDING ZKPS TO THE ZT TOOLBOX

Current ZT Model Verification Flow	ZT+ ZKP Verification Flow
1. User requests access	1. User wants access
2. System queries identity provider	2. User locally generates a proof
3. Identity provider returns attributes	3. User sends only the proof
4. Policy engine evaluates attributes	4. Policy engine verifies the proof
5. Access granted or denied	5. Access granted or denied

Use Cases on the Horizon

Within the United States and across governments around the world, ZKPs are actively in development for privacy-preserving digital identity, credential verification, and regulatory compliance, allowing users and institutions to prove attributes or satisfy requirements without exposing underlying personal or sensitive data.

Promising research and prototypes already emerging, with mission-aligned applications such as:

Proof of Human. ZKPs can be combined with biometrics and behavioral signals to let individuals prove they are a unique, real human—rather than a bot or synthetic identity—without revealing the underlying personal data used to verify that claim.

Cyber Defense Against Deepfakes. With the rise of agentic AI and high-quality deepfakes in 2026, ZKPs are being used to help authenticate digital content. For example, news organizations and government bodies could eventually use ZKPs to sign metadata, proving a video was captured by a specific trusted camera at a specific time without revealing the GPS coordinates of a sensitive location.

Verifiable AI and Automated Decisions. Government systems may gain the ability to generate proofs that algorithms and models are executed correctly and according to preapproved rules.

Secure Inter-Agency Computation. Agencies and partners could verify insights derived from sensitive data without sharing the raw datasets or revealing sources and methods.

Privacy-Centric Citizen Services. Citizens could one day prove eligibility, age, residency, income bracket, or other attributes for services without disclosing full personally identifiable information or sensitive documents.

Secure Cross-Domain Data Exchange. Operators handling classified data could prove policy compliance before releasing information across security boundaries.

Regulatory Compliance and Auditing. Organizations could prove compliance or audit outcomes without sharing logs, sensitive operational details, or proprietary information.

Together, these use cases illustrate how ZKPs can extend trust across domains where data sensitivity, operational security, and verification requirements would otherwise be in tension.

Market Dynamics to Watch

Startups Are Expanding Beyond Crypto Tooling

Startups are still the primary innovators of ZKPs. The market for pure-play ZKPs is relatively small, but the growing interest in privacy, scalability, and ZT architectures is attracting meaningful capital. While startups continue to focus on ZKP for privacy-preserving blockchain applications, many are also expanding into broader practical applications across multiple sectors.

What we're watching:

- Post-2020, average investment deal size and investor quality have surged, reflecting growing institutional confidence in horizontal ZKP infrastructure as a foundational technology beyond any single blockchain or application. For example, Matter Labs has attracted investment from top venture funds, including a16z, Lightspeed Ventures, and Coinbase Ventures.
- Emerging players are advancing ZKP infrastructure to improve scalability and efficiency—often through techniques that move computation off primary systems while preserving verifiability. For example, Succinct Labs has developed a prover network and an open-source zero-knowledge virtual machine (zkVM) to support scalable, developer-friendly proof generation.
- Leading startups are partnering with major cloud providers to accelerate adoption and integration. For example, Space and Time is integrating real-time, verifiable blockchain data into Microsoft Fabric and Azure, and working with Google Cloud BigQuery to enable verifiable queries within data warehousing environments.

Big Tech Is Actively Developing Scalable ZKP Frameworks

As the last decade of cryptocurrency development has shown, ZKPs are no longer theoretical. Now, an ecosystem beyond blockchain is beginning to take shape—driven initially by startups and increasingly scaled by Big Tech. While startups continue to lead in ZKP innovation, large technology companies are focusing on enterprise adoption and scalability, particularly by laying the groundwork for verifiable computation. Their involvement is helping shift trust from centralized systems to cryptographic guarantees, enabling proof-based access and computation. By accelerating adoption, standardization, and cross-platform interoperability, Big Tech is pushing ZKPs beyond niche use cases into mainstream cybersecurity and information security applications for both enterprises and consumers.

What we're watching:

- Google has integrated ZKP protocols into the Android operating system, enabling third-party developers to request proofs instead of raw data (e.g., a banking app verifying sufficient funds without accessing the exact account balance).
- Microsoft Research and Azure are exploring ZKPs within decentralized identity and confidential computing frameworks, aiming to enable proofs of data handling or compliance without exposing underlying content.
- Apple uses ZKP-style cryptography in Apple Wallet to verify that users are over 21 without revealing their name or date of birth to merchants.
- Meta has developed the Winterfell STARK prover and is actively researching ZK-based approaches for privacy-preserving and verifiable computation.

What's Next for the Technology

Projected Timeline of Development

The trajectory of ZKP adoption is likely to follow a familiar pattern: early technical maturation, followed by integration and standardization, and ultimately institutionalization across systems and policy frameworks. We expect ZKPs to reach a tipping point of commercial maturity outside the blockchain ecosystem within two to five years, at which point they will be more broadly evaluated and adopted as an adjunct to existing ZT frameworks.

TRAJECTORY OF ZKP ADOPTION

0–2 Years

PRACTICAL MATURATION

- Faster provers, smaller proofs, and better developer tooling
- Early production use in identity, access control, supply chain verification, and security auditing
- Federal pilots and early deployments
- ZK-native identity and credentials rep

2–5 Years

INTEGRATION + STANDARDIZATION

- ZK-native identity and credentials replace static identifiers and attribute sharing
- ZKP modules are integrated into identity and access management stacks
- Verifiable computation becomes standard for select cloud, edge, and AI workloads
- Early standards emerge for ZK-enabled ZT frameworks

5–10 Years

INSTITUTIONALIZATION

- Standards are fully defined by NIST and adopted across U.S. government systems
- Cryptographic compliance becomes the default, with ZKPs embedded at protocol and hardware layers
- Privacy-preserving collaboration becomes feasible at scale across enterprises, agencies, and allies—even in adversarial environments

An Evolving Security Discipline

The trajectory for ZKPs is clear and very promising, but there are challenges and limitations to navigate before this security discipline becomes fully mainstream outside crypto-native environments:

- **Computational cost:** Generating proofs can be expensive, especially for complex logic or large datasets. While this is improving rapidly, prover-side costs remain a key bottleneck.
- **Latency:** Proof generation can introduce delays, which matters for real-time or tactical systems.
- **Engineering complexity:** Writing ZK-friendly programs is difficult and often requires redesigning systems around arithmetic circuits rather than normal software abstractions.
- **Integration with legacy systems:** Most enterprise and government infrastructure was not designed with cryptographic proofs in mind, so retrofitting ZKPs requires significant systems engineering.
- **Standards and assurance:** For government adoption, ZKPs still require more formal standards, certification pathways, and operational trust models.

ZKPs are early in the standards curve, similar to where post-quantum cryptography was a decade ago. We expect adoption will follow the maturation of tooling and assurance frameworks, and, fortunately, those wheels are already turning. Today, NIST's Privacy-Enhancing Cryptography (PEC) initiative—focused on developing a range of cryptographic tools—is beginning to shape the standardization of ZKPs. Its goal is to make these protocols interoperable, auditable, and enterprise-ready, while integrating them with other cryptographic approaches rather than treating ZKPs as a standalone technology.

The Emerging Assurance Stack

ZKPs do not replace existing identity or ZT tools. Rather, ZKPs are a natural addition to the layered security toolbox, reinforcing assurance at each layer of the stack:

They extend identity, access, and ZT tools by reducing the amount of information identities must reveal.

They complement secure enclaves by adding cryptographic verification to systems that otherwise rely mostly on trusted hardware.

LAYERS OF THE ASSURANCE STACK WITH ZKPS

LAYER 4 *ZKPs used here*	**Trust, Identity, & Interaction** Establishes trust between systems, users, or organizations *(includes 21 architectures)*
LAYER 3 *ZKPs defined here*	**Cryptographic Proof & Privacy** Lets one party convince another of correctness without revealing sensitive information
LAYER 2 *AR used here*	**Correctness & Assurance** Ensures systems behave as intended (includes formal verification, model checking, etc.)
LAYER 1 *AR defined here*	**Computation & Logic** Figures out what is true or what satisfies constraints (includes automated reasoning, formal logic engines, etc.)

They share mathematical DNA with automated reasoning and solve problems in different layers of the assurance stack.

ZKPs now sit at the intersection of cryptography, distributed systems, and national-scale security concerns, and the industry is approaching a critical inflection point in their development. What began as a curiosity has become a practical tool for verifying trust without disclosure—precisely what modern, adversarial, and data-sensitive environments demand.

The Booz Allen Tech Scouting and Intelligence team scans the market for novel technologies with the potential to disrupt U.S. government critical missions. Delivering research, trends spotting, and deep dive technical analysis, the team is hyper-focused on getting the best technology from the private sector into the hands of the warfighter and civilian federal workforce.

Livia Dworaczyk and Meghan Hauser are members of the Tech Scouting team, and Tom Hanson is part of Booz Allen's Post-Quantum Cryptography (PQC) practice.

SPEED READ

Zero-knowledge proofs (ZKPs) are emerging as a transformative way for government and industry to break the long-standing "proof vs. privacy" tradeoff—enabling verification of identity, compliance, and computation without exposing underlying sensitive data.

Once engineered largely for blockchain, modern ZKP systems are now maturing into mission-ready infrastructure that can support zero trust architectures, verifiable AI, secure inter-agency analytics, deepfake-resistant content authentication, and privacy-preserving citizen services.

As federal agencies explore pilots and as startups and Big Tech accelerate scalable frameworks, ZKPs are on track to become a pivotal layer of the government assurance stack—reducing data exposure, enabling cryptographic trust, and reshaping how sensitive systems operate in adversarial environments.

The MATH That Makes Technology Trustworthy

Formal methods and automated reasoning are reshaping how agencies can validate critical software, enforce AI guardrails, and eliminate entire classes of vulnerabilities

By Wyatt Chaffee, featuring Bill Vass and Byron Cook

Federal missions increasingly rely on software and AI systems—for example, algorithms and automated decision-support tools—to make real-time decisions under conditions that are fast, fluid, and adversarial. Yet the methods used to ensure the correctness and accuracy of systems and programs have barely changed. Traditional testing can show that software works in some scenarios—but not that it works in all of them. And as AI-enabled systems expand the space of possible behaviors, the gap between what agencies can test and what they must trust is widening.

> “The idea behind formal methods—going back to the Greeks—is that rather than thinking about the content of the argument, you think about the form of the argument. Now you can say: If the argument has this form, it’s correct, regardless of what ‘P’ and ‘Q’ actually are. Automated reasoning is the automation of the reasoning over that form.”
>
> —Byron Cook, Amazon Web Services (AWS) Vice President, Distinguished Scientist, and Automated Reasoning Group Lead

This is where the mathematically grounded techniques of automated reasoning and formal methods shift the landscape. Instead of sampling behaviors through tests, automated reasoning generates mathematical proofs that ensure a system will always uphold its most critical properties—no matter the input, environment, or adversary. It replaces the question “Did we test enough?” with “Do we have a proof?” It’s the difference between stress-testing a bridge by driving trucks across it and demonstrating through structural calculations that the bridge will hold any load intended, while testing remains essential for properties that resist formalization.

Long reserved for chip manufacturers, aerospace systems, and a handful of high-assurance environments, these techniques are now entering mainstream engineering. Cloud service providers use automated reasoning to verify the identity, access, and network isolation policies of their platforms. Defense and space systems rely on it to eliminate entire classes of safety-critical bugs. And as AI becomes central to federal operations, agencies are beginning to explore formal verification as mission environments increasingly require a backbone of trusted autonomy and policy-aligned behavior.

This article equips IT leaders—particularly those responsible for systems with national security implications—with the basic insights needed to assess automated reasoning and formal methods in greater depth. We conclude with an extended Q&A with Byron Cook, AWS vice president, distinguished scientist, and automated reasoning group lead, whose work has moved automated reasoning from academic theory to mission-critical capability protecting billions of secure transactions each day.

BEYOND TESTING

Conventional software testing works by running a program against specific inputs and checking whether the outputs are correct. Does the login screen accept a valid password? Reject an invalid one? Handle a blank field?

Each test covers a single scenario. However, even a simple program can accept many possible inputs, and testing can cover only a tiny fraction.

Consider a function operating on 32-bit rational numbers. Testing every possible input would require time-consuming computation. By contrast, a formal methods tool can represent the entire set of 32-bit rationals symbolically and prove the desired property for all of them in minutes or seconds. While testing scales with the number of cases checked, formal reasoning scales with the overall logic of the problem.

Working in this way, an automated reasoning tool can examine an authentication module and produce a mathematical proof for which no input—no matter how malformed or adversarial—will ever lead to unauthorized access. Compare this with machine learning. Machine learning identifies statistical patterns and makes probabilistic predictions. Automated reasoning operates on logical rules and produces deterministic, provably correct conclusions. When an automated reasoning tool reports that a property holds, that result is mathematically guaranteed—not for most cases, but for all cases. It is as certain as a geometry theorem.

KEY CAPABILITIES, UNDER THE HOOD

Computer scientist Tony Hoare created a framework for proving program properties in 1969. For roughly the next two decades, the field of formal methods advanced steadily but remained specialized. In 1994, Intel’s Pentium processor was found to have produced incorrect results for certain floating-point division (FDIV) operations. The FDIV bug cost approximately $475 million in recalls. As a result, formal methods increasingly moved toward becoming a valued engineering component. Advances in solver technology soon followed.

While formal methods proofs were once done by hand, modern automated reasoning sits on a technology stack encompassing different tools that streamline specific parts of the process, from satisfiability (SAT) solvers and model checkers to theorem provers.

SAT solvers determine whether any combination of Boolean values can satisfy a logical formula; they handle problems with millions of variables. For example, you can use this to automatically verify that no combination of configuration settings in a complex system will produce a contradictory or unsafe state. **Satisfiability modulo theories (SMT) solvers**—the most widely used is Z3, developed at Microsoft Research—extend this to integers, arrays, bit-vectors, and other data types to enable the higher-level verification tools engineers use. This tool helps you confirm, for example, that a memory allocation routine can never return a negative size value, regardless of what inputs the program gets.

Built on top of these are **model checkers,** which look at every state a system can reach, a technique helpful for systems where timing bugs are difficult to find. You can use this to verify that two processes in a network protocol can never simultaneously believe they hold an exclusive lock. **Interactive proof assistants**—such as Rocq (formerly “Coq”), Isabelle, and Lean—let a human construct a formal proof while the computer verifies every logical step, producing a high-assurance artifact. This tool can help you formally certify that a compiler or cryptographic library behaves exactly as specified across all possible inputs, via a machine-checked proof that humans can inspect and audit at every step.

Two additional techniques extend the toolkit. **Abstract interpretation** enables static analyzers to reason about program properties but not run the program. For example, you can use this to scan avionics software for the entire class of runtime errors before the code executes on hardware. And **symbolic execution** runs programs with symbolic rather than concrete values, exploring all execution paths to identify vulnerabilities. Symbolic execution was demonstrated at scale in combination with other techniques in the Defense Advanced Research Projects Agency (DARPA) 2016 Cyber Grand Challenge.

These tools require **formal specifications**, which are precise mathematical descriptions of what a system is supposed to do. Specification languages include computer scientist Leslie Lamport’s TLA+ (Temporal Logic of Actions), used for distributed systems, and the Frama-C/ACSL framework for verifying C programs in avionics and nuclear safety. However, it’s important to keep in mind that a formal proof is only as trustworthy as the model it verifies. If the specification doesn’t faithfully represent the actual code, the engineer may have proven the wrong thing, leaving the software unverified. This is why automation that enables tracing across specification, proof, and implementation is more reliable than manual transcription, which can introduce the risk that a tacit assumption has broken the chain of evidence.

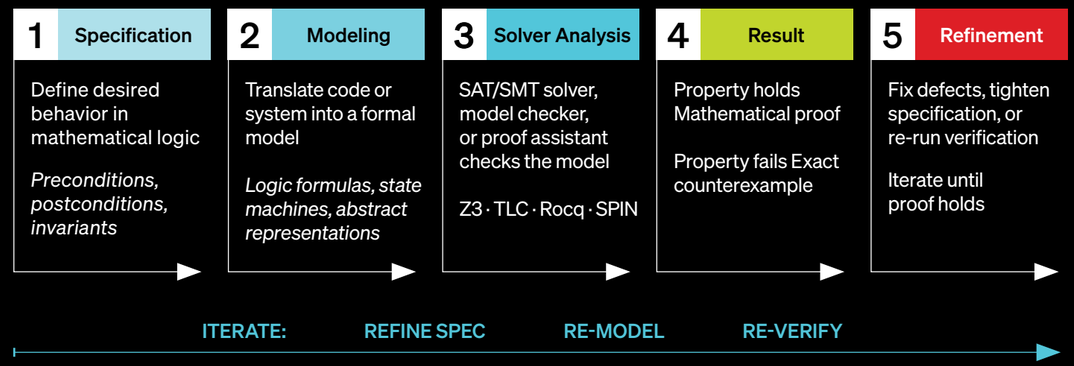
CODE THAT BUILDS TRUST

Wherever a single software error can be catastrophic—such as flight control, cryptography, autonomous vehicle operation, and financial transactions—automated reasoning and formal methods are gaining relevance.

For example, a team in France built the first formally verified C compiler. The mathematical proof underlying the CompCert tool guarantees that compiled code keeps the meaning of the original source. A team in Australia produced the seL4 operating system kernel, which has been formally verified. In the aerospace sector, the Astrée tool confirmed that runtime errors were not present in

From Source Code to Proof:

The Formal Verification Workflow



Key insight: When the solver finds a violation, it does not just report “fail.” It produces an exact counterexample—the specific input, state, or execution sequence that breaks the property. These counterexamples often reveal bugs that years of testing never found.

When the proof succeeds, the guarantee covers all possible inputs—not a sample, not a probability, but a mathematical certainty.

Airbus A340 and A380 primary flight control software. And the Certora Prover applies formal verification to smart contracts on blockchain platforms, where a logic error could cause hundreds of millions of dollars in losses.

However, not all code should be verified through automated reasoning given the significant tradeoffs in terms of time, cost, energy usage, and skillset requirements when compared with conventional testing. Questions to ask include: Would failure be catastrophic? Would it create regulatory exposure? Is the code path small but highly consequential—a ledger update, an authorization decision, a cryptographic handshake? Most organizations will not verify their entire codebase. They will identify high-stakes components and apply formal methods selectively while relying on testing for the rest.

It's also important to note that, even when an organization identifies a strong use case, it may lack the expertise to carry it out. Formal methods require foundations in mathematical logic, automata theory, and programming language semantics. GenAI promises to democratize these tools to a greater extent in the future. For example, LLM-based chatbots can assist in writing specifications and translating natural-language requirements into formal models.

PROOF OF PERFORMANCE

The federal government has supported formal methods for decades. DARPA's High-Assurance Cyber Military Systems (HACMS) program adopted seL4 as the foundation for formally verified software on unmanned military platforms. Starting with a small quadcopter and then moving to the Little Bird helicopter, the program used formal methods to build software that creates machine-checkable proofs. DARPA has applied the HACMS tools to other military systems, including satellites.

The National Security Agency has required formal specifications under the Trusted Computer System Evaluation Criteria (i.e., the Orange Book). At the highest assurance class, A1 (Verified Design), the Orange Book required a Formal Top-Level Specification of the trusted computing base, configuration management procedures enforced throughout the system lifecycle, and formal verification methods to show that the implementation is consistent with the design. Although the Orange Book was later superseded, it established the concept that

> "What I really loved about automated reasoning is that it's like little mechanical engines of truth."
>
> —Byron Cook, Amazon Web Services (AWS) Vice President, Distinguished Scientist, and Automated Reasoning Group Lead

The Trust Stack: Where Automated Reasoning Sits in AI Deployment

The key principle: The AI model remains creative and capable. The formal layers above it ensure it cannot violate policy. The model proposes; the logic disposes.

LAYER 4
User/Mission System
Receives only verified, policy-compliant outputs. Mathematical certainty for evaluation, not statistical confidence.

LAYER 3
Policy Specification
Formal declarations of allowable and disallowed behaviors. Written in languages like Cedar, TLA+, or Alloy.

LAYER 2
Formal Verification
SMT solver or reasoning engine checks every output against declared rules before it reaches the user.

LAYER 1
AI Model
Generates outputs based on statistical patterns. Non-deterministic: identical inputs can yield different outputs.

AUTOMATED REASONING TRUST LAYER
Provably correct guardrails ensure AI outputs comply with formal policy before reaching the user.

system trust requires mathematical verification. In addition, NASA's Formal Methods Research Group includes participants from six NASA centers and holds an annual symposium on formal techniques for software assurance in space, aviation, and robotics. NASA Ames has applied model checking to verify autonomous planning software, including planning systems flown on Deep Space 1.

AWS has led the commercial deployment of these techniques after assembling a team of automated reasoning researchers and applying formal methods to its cloud infrastructure. A tool called Zelkova uses SMT solving to analyze identity and access management policies, powering IAM Access Analyzer for AWS customers. Another tool, Tiros, verifies network configurations. AWS engineers used TLA+ to find bugs in DynamoDB and other services.

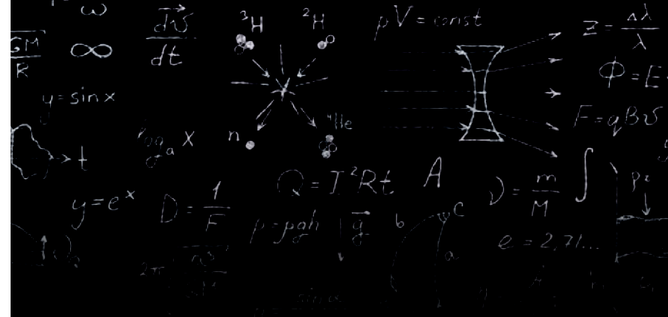
More recently, AWS extended its approach with Cedar, an open-source policy language backed by an automated reasoning engine. Cedar lets developers write access control rules that are easy to understand. The reasoning engine verifies, at deployment or runtime, that policies are internally consistent and that no combination of permissions creates an unintended access path.

A MISSING PIECE FOR TRUSTWORTHY AI

LLMs are non-deterministic by design: even identical inputs can yield different outputs. In addition, they sometimes hallucinate. For intelligence analysis, medical decision support, financial compliance, and other critical missions, automated reasoning offers a way to manage these attributes. Although automated reasoning can't explain how a model thinks, it can enable engineers to constrain what it is allowed to do. They can use formal methods to declare allowed and disallowed behaviors, then enforce those limits. As a result, the model can function productively without violating policy.

In **LLM-guided formal verification**, a language model generates a candidate proof, and a formal tool checks it. If the model hallucinates, the verifier catches it. Google DeepMind's AlphaProof system used this approach in 2024 to solve International Mathematical Olympiad problems at a level competitive with silver-medal finishers.

In **formal guardrail architectures**, automated reasoning acts as a runtime verification layer; it checks every LLM output against formal rules before it reaches the user. Does this query violate access control? Does this dosage exceed safe limits? In **verified code generation**, formal tools verify AI-generated code before deployment, which is critical when human-in-the-loop review can't scale to the volume of code AI produces.



A Brief History of Automated Reasoning

From Mathematical Logic to Industrial Deployment

Foundations (1956–1985) →

1956–1969: The Birth of Automated Reasoning

Newell and Simon's Logic Theorist (1956) demonstrated automated proof search. Building on earlier work by Davis and Putnam, the DPLL algorithm (1962) laid the foundations of modern satisfiability solving. Hoare logic (1969) introduced a mathematical framework for reasoning rigorously about program correctness.

Amir Pnueli's introduction of temporal logic (1977)

Enabled the precise specification of concurrent and reactive systems. In the early 1980s, model checking was developed by Edmund Clarke and E. Allen Emerson, and independently by Joseph Sifakis, introducing automated, exhaustive verification for finite-state systems.

Proof Assistants and the SAT Revolution (1979–2010) →

1979–Early 2000s: Theorem Provers Come of Age

NQTHM (1979) marked major advances in practical automated theorem proving. By the late 1980s, proof assistants including Isabelle and Coq enabled expressive higher-order reasoning with machine-checked guarantees, maturing into powerful environments for large, mechanically verified proofs.

1995–2010: SAT and SMT Transform Verification

Satisfiability solving underwent a breakthrough in the late 1990s with conflict-driven clause learning (CDCL), making SAT—and later SMT—solvers practical engines for software and hardware verification. The SLAM project led to the Static Driver Verifier (introduced 2004, shipped 2006), one of the first large-scale industrial deployments of automated model checking, applied across the Windows driver ecosystem. Z3 (2007) unified logical reasoning with theories of arithmetic, arrays, and data structures, becoming a foundational tool across academia and industry.

Industrial Deployment (2005–2017)

Mid-2000s: CompCert and Verified Compilation

CompCert (2006), developed by Xavier Leroy, was the first widely recognized, practically usable verified compiler—carrying a machine-checked, end-to-end proof of semantic preservation and showing that formal methods could produce deployable systems.

2009: seL4 and Operating System Verification

The seL4 project achieved the first full formal proof of functional correctness for a general-purpose OS kernel, developed using Isabelle/HOL and subsequently deployed in security- and safety-critical domains.

2015–2017: Cloud-Scale Formal Verification

s2n (2015) was formally analyzed to ensure security properties, with similar techniques applied to distributed infrastructure serving millions of users—demonstrating that formal verification could operate at cloud scale.

Expanding Frontiers (2016–Present) →

2016–2021: Formal Methods in Blockchain and Mathematics

KEVM (2017) provided a formal semantic foundation for reasoning about smart contracts. The Lean ecosystem—including mathlib (2017) and Lean 4 (2021)—enabled large-scale collaborative formalization of mathematics.

2023: Authorization System

Cedar (2023) demonstrated that formally grounded reasoning about access control policies could operate efficiently in real-time production environments.

2024 and Beyond

AlphaProof (2024) generated formally verified proofs for several IMO-level problems, pointing toward AI systems that produce machine-checkable correctness guarantees alongside their solutions.

A PATH TO ADOPTION

Start with automated reasoning, not full formal verification. Full verification—interactive proof construction with tools like Rocq or Isabelle—requires deep mathematical expertise and significant upfront investment. Automated reasoning encompassing solver-backed analysis with Z3, TLA+, Alloy, or Cedar is more accessible, more automatable, and already integrated into commercial services that abstract away the underlying complexity.

Identify highest-stakes components and pilot on a small, high-impact use case. A cryptographic code path, an authorization module, or a state machine controller is where formal methods can deliver the most value relative to effort while multiplying early wins.

Integrate solver-based checks into CI/CD pipelines and invest in developer education. Automated reasoning maximizes return when it runs continuously alongside existing tests, rather than as a one-time exercise. Sustaining that integration requires building developer fluency in foundational CS theory and solver-based workflows.

Build internal exemplars that demystify the discipline. End-to-end, realistic examples are what convert skeptics and give engineering teams a concrete model to follow.

Move now to harness mature capability. Automated reasoning is now entering broader adoption. AI-augmented reasoning—where LLMs generate specifications, translate requirements into formal models, and produce candidate proofs—is also emerging fast. Organizations that build capability now will be positioned to leverage the candidate proofs as guardrails for AI-generated content while exploring additional impactful mission use cases sooner.

LOOKING AHEAD

Language models will continue lowering the expertise barrier for formal methods. Automated reasoning will be embedded in development pipelines, flagging bugs and serving as the trust layer for AI-generated code, recommendations, and decisions. Federal procurement is already reflecting the shift, with formal methods appearing in some acquisition requirements for defense, aerospace, and advanced cloud systems.

Across these developments, a basic concept will continue to be powerful: automated reasoning moves software assurance from “we tested it, and it worked” to “we proved it correct, for all possible conditions, mathematically.” For anyone responsible for systems critical to missions that can’t fail, this difference will increasingly matter in demanding real-world environments.

Reasoning About the Infinite:

An Interview With Byron Cook

To explore where the field is heading—and how automated reasoning is being applied at scale inside major cloud and AI platforms—Booz Allen Chief Technology Officer Bill Vass spoke with Byron Cook, AWS vice president, distinguished scientist, and automated reasoning group lead, to explore new perspectives on automated reasoning and formal methods.

Q Bill: What started your interest in automated reasoning and formal methods?

A Byron: I studied to be a VW [Volkswagen] mechanic in high school – I restored air-cooled 1960s-era VWs. I worked with a retired NASA mechanical engineer at a high school where you didn't get grades and you didn't go to class. You designed your own curriculum, and one of the areas I decided to work on was learning how to rebuild cars.

Then I went to Evergreen State College in Olympia, Washington. There was the world expert on Hilbert's epsilon operator, who taught only advanced logic. He taught a course called "Computer Building Cognition," a year-long course where you just worked with him. These days we'd call it "neurosymbolic AI." I basically went to the registrar at Evergreen and said to sign me up for the hardest thing available, because I just wanted whatever the hardest thing was. They sent me to that course, and I absolutely loved it. This was around 1993 or 1994. Then they said, well, maybe you should do a Ph.D. And I've just never looked back.

What I really loved about the VWs was how simple the motors were. Because of the constraints they were designed under, they minimized the parts you needed. And what I really loved about automated reasoning is that it's like little mechanical engines of truth. You're reasoning about the infinite but using a composition of very simple parts that are very, very powerful. I see a lot of parallels between real mechanical motors and mechanical reasoning about truth.

Q Bill: What are formal methods, what is automated reasoning, and how do they relate?

A Byron: They're cousins. "Formal methods" is the term people used in the seventies and eighties before the tools were very automated. Automated reasoning is the automation of the reasoning you were doing when you did formal methods manually.

The idea behind formal methods—going back to the Greeks—is that rather than thinking about the content of the argument, you think about the form of the argument. Now you can say: If the argument has this form, it's correct, regardless of what "P" and "Q" actually are. Automated reasoning is the automation of the reasoning over that form.

The challenge is that the problem is undecidable. We know that from Gödel, and Turing directly connected it to the halting problem. So, you have to acknowledge that the tools will sometimes just go to lunch: they can't always give you an answer. You have to expect that. Formal methods were the very early days, when the tools weren't there, so you were essentially saying to a human: we don't know how to solve this problem, so we'll let you use human ingenuity, but here's the shape of the solution. You were exploring that space with very little tool support.

Q Bill: How would you explain automated reasoning to a non-mathematician and a non-programmer?

A Byron: Usually, people have done just a little bit of algebra. They've pushed symbols around—X plus Y is the same as Y plus X. You've moved symbols, so X and Y are symbols; that's symbolic reasoning. And I think everyone can look at X plus Y and Y plus X and say, yeah, it's roughly the same thing. But you've actually just reasoned about the infinite—X and Y range over potentially all the integers, all the rationals.

There's some equipment you can pull in to ensure that X plus Y equals Y plus X, and they're the very same mechanisms used around 300 BC to show that the Pythagorean theorem held. Reasoning, in the sense I'm talking about, is finding an air-tight argument as to why something is true, where each step of the argument rests on axioms, or the foundations we've agreed on as a society for how to establish truth. The automated part is just algorithms over that. You're searching for finite arguments about the infinite.

Q Bill: When did you realize you could take this from academic work to actual implementations in production? When did you think it would be practically applied to important parts of code?

A Byron: I was very fortunate that from the very beginning I was right in the middle of application. There was the Intel floating point division bug in the early 1990s, and the Intel Pentium team hired formal reasoning people to prove the correctness of the underlying reorder buffers, register alias tables, and the microcode that implemented floating point arithmetic. I was an intern there as a Ph.D. student.

Then for various reasons I got pulled in when Microsoft realized around 1999 to 2000 that they were at risk of losing out in the data center because Windows was crashing so much—predominantly because of device drivers. I reported to the person who owned the I/O [input/output] subsystem of the kernel and applied these tools there.

After 10 years in pure ivory tower research, when I joined Amazon, I was actually coming back to a familiar space. I repeated what I had seen work at Intel: I'll start four things, message it as though any one of these succeeding is a win, and we'll see where it goes. That complemented Amazon's view of "do something small, see if it works, then double down relentlessly." That approach has just grown like a weed using relatively standard Amazon mechanisms, seeded by how I'd done things at Intel.

These days a lot of what I spend time on is talking to other Amazon leaders because they all want to start science in their own organizations. I talk them through how I framed it, how I measured it, and how to bring science into a pretty delivery-oriented organization.

Q Bill: Amazon has been a pioneer under your leadership in applying automated reasoning in areas like cloud security and policy compliance. What are your key lessons learned from the journey? For example, how have you addressed the costs involved with implementing the technology?

A Byron: Automated reasoning is largely a CPU activity, not a GPU activity, so it's actually relatively cheap. But culturally it's expensive, because the familiarity and intuition about where these techniques will and won't work are not commonly held in the IT community.

You have to think about things differently. Automated reasoning is like jazz. Miles Davis said jazz is all about the notes you don't play. It's all about how you reframe a question into a different question that you can answer efficiently, even though you're reasoning about the infinite without actually running the program or doing all the computation.

In terms of where I focused at Amazon early on: I mined the internal database of all security incidents for situations where my tools would help, prioritized them, matched the problems to the people I knew I could hire, and found four or five promising things to start. From there I was essentially an ambulance chaser at the Friday security meetings with Andy, where they'd bring in the top security close calls of the week and discuss what mechanisms to put in place. I'd help people navigate how they could be using my tools as they went in or came out of those meetings. That built trust, and from there it was a flywheel.

We now have around 350 Ph.D.s working across 20 teams at Amazon, with roughly 150 Ph.D. student interns last year. Automated reasoning has the most value where the algorithms are really hard to get right, or where the consequences of getting them wrong are absolutely catastrophic. That tends to be the lowest levels of the stack and the tight loops—cryptography, the virtualization layer, durability for storage.

And then with agentic AI, what's happening is we're trying to replace a lot of our sociotechnical mechanisms with automation. But agents aren't sentient. They're not afraid to get fired or go to jail, so the mechanisms we relied on before don't hold anymore. You either require a human to approve everything, or you have to put other mechanisms in place to let agents roam. That's becoming an important area for automated reasoning to scale.

Q Bill: How do you envision AI impacting automated reasoning?

A Byron: There's a really exciting chocolate-and-peanut-butter moment happening with agentic AI. A lot of the heuristics—the tools—in automated reasoning were either fully automated but with limited capabilities, or much more powerful but requiring a Ph.D. to guide them. That was a neurosymbolic system where the "neuro" was someone's brain, because they saw the patterns and narrowed down the infinite branching directions of a proof search to the most likely successful strategy. Now we're able to bring in language models to help guide the search for the proof. So, we get 100%- assurance of what's true and untrue, but we also get much more automation. That's opened up some pretty remarkable opportunities.

Q Bill: If you're a CIO or CTO in the government or a big corporation—not building virtual machines or security code or an identity management system—where would you recommend applying automated reasoning?

A Byron: A lot of people right now are looking at the agentic space, and there are a bunch of startups, and Amazon is working here, too. The idea is defining the envelope of behaviors you want to allow and don't want to allow from your agents. Those behaviors define availability, confidentiality, integrity, sovereignty, privacy—the different dimensions and the lines you really don't want agents to cross. You create an envelope, and as long as the agents' behaviors stay within it, you're good. If they try to exceed it, you stop them. That's a fast-moving area. From an Amazon product perspective, that's where something called Agent Core Policy is headed right now.

But if you're not yet doing agentic systems and you're a CIO, I would look at what are the algorithms where, if they're wrong, you're under existential threat. Organizations are often based on trust, and there's some dimension of trust that they're differentiating on—the reason customers come to them. So, you want to get those algorithms right. If it's cryptocurrency contracts, you want customers to be able to trust that the algorithms executing those contracts on their behalf don't have bugs.

On the compliance side, it's very tricky because there's a compliance industrial complex that's ripe for disruption, but it won't be easy. It's less about formally proving everything and more about providing transparency about the decisions you're making and how you know your systems are correct and building trust on that. The mechanisms by which we decide compliance will hopefully adapt to the tools.

It'll be very hard to formally prove that your assessment was correct. But it'll be much easier to show that the controls around tokenization are correct, or that differential privacy is correct, or that the virtualization is configured correctly such that you have an argument as to why something is true—and that argument could be audited every hour instead of every year.

Q Bill: Could you talk about the goals of the work you're doing with DARPA?

A Byron: I had never worked in government but thought I should do a bit of a tour of duty to contribute back to the country that's created such a great environment to live and do work in. I'm trying to help the government figure out what some of the big scientific levers are.

I'm framing this as imagining a world where proof is just a thing. It's pervasive. All the major players, the hyperscalers, and a whole load of startups are bringing automated reasoning and neurosymbolic tools to the table. Lean theorem prover and a lot of these tools are open source. So, imagine all that capability exists. I'm leading exercises to really accelerate the nation's ability to leverage the existence of that capability. The programs I'm brewing up mostly have that flavor: Proof is a thing, now what?

Q Bill: If we had this conversation again in three years, how would you want it to look? What would be your target around automated reasoning, agentic systems, and policy?

A Byron: There are a bunch of things happening right now that I'm really excited about. As a society, we're going through an "art of the possible" exercise. Language models came out of nowhere from the perspective of the general public, and suddenly people had access to easy-to-use tools that could answer a different set of questions and solve a different kind of problem. That has brought neurosymbolic AI along with it. If you look at the Gartner Hype Cycle, neurosymbolic AI—automated reasoning combined with language models—is shooting up in interest.

One of the traditional barriers to entry in this space was step one: you need to write down in logic what you want. That was a non-starter for 99% of the population. But there's interesting work on auto-formalization—building a logical twin of your natural language statement. Imagine you say, "I want all data at rest to be encrypted" or "I never want to log credentials." You can now encode that in an appropriate formalism like linear temporal logic and then use the logical structure to help people who don't understand logic see all the weird corner cases—to get comfortable that the formalism is actually saying what they wanted to say. Then you can maintain it.

I think we're getting very close to allowing a much larger set of people to write down what they want for their systems and argue about it meaningfully in meetings: "No, this is what availability should mean." "This is what durability should mean."

The other piece that's always been tricky with automated reasoning is writing down and getting confidence around the assumptions. When you prove a program correct, you're assuming the runtime is correct, the compiler is correct, the microprocessor is correct, the network is correct. But now you can automatically compute those abstractions and use model-based testing to harden them and then prove them if you see fit.

The interplay between testing, proof, and natural language logic means a lot of the barriers that were so fierce before are melting away. People now have a whole plethora of tools and different levels of assurance they can seek, with different time and energy investments. It provides a menu. It's a pay-as-you-go model, as opposed to becoming a true believer in one technique. The lines between these different techniques, which people used to argue about, are blurring, and people are partnering together more. I'd love to see that continue over the next three years.

We're getting close to the point where you can just say: On this system, I want to make sure these critical things are always true—no privilege escalation, no two planes on the runway in the same place at the same time, never a negative balance—and go prove to me that the code enforces them. And you can generate something that convinces both a theorem prover and an auditor. That translation between formalisms and different people's backgrounds—from formal methods to auditor—is now in reach.

Bill Vass is Booz Allen's chief technology officer, and Wyatt Chaffee is a director of technology innovation. Byron Cook is a vice president and Distinguished Scientist at AWS, where Byron leads the Automated Reasoning Group.

SPEED READ

Automated reasoning uses mathematical proofs to validate that software and AI systems will function as intended, helping eliminate entire classes of vulnerabilities before deployment.

These techniques—already securing cloud infrastructure, aerospace systems, and cryptographic code paths—are emerging as a solution for AI assurance, enabling engineers to define and enforce the boundaries of safe and policy-aligned model behavior.

In our concluding interview, Amazon's Byron Cook describes how automated reasoning has scaled from academic theory to protecting billions of cloud transactions daily, and why neurosymbolic AI is accelerating the field toward broader, mission-level adoption.

REIMAGINING CYBER FOR A FASTER FIGHT

Securing enterprises against AI threats requires disrupting operating models, enriching detection, and strengthening resilience—because attacks now unfold in minutes, not days.

By Brad Medairy and Andrew Turner

AI is now operating across the entire cyber kill chain—from vulnerability discovery and exploit development to effect delivery. This is no longer theoretical.

Consider a recent mass vulnerability exploitation against a widely used, web-based system administration tool. Threat actors identified a zero-day vulnerability in its Python script, weaponizing it to bypass two-factor authentication. This example is noteworthy as one of the first observed examples of "a threat actor using a zero-day exploit [believed to be] developed with AI," according to a recent Google report. Evidence consistent with AI-assisted exploit development included a hallucinated CVSS score and an unusually textbook, heavily documented Python implementation—traits they said were highly characteristic of code commonly found in LLM training data.

AI-powered attacks are the new attack vector, and they are accelerating rapidly. The Five Eyes intelligence alliance recently warned that major businesses and governments were only "months" away from exposure to AI-enabled breaches. And a recent OALABS Research report found that advanced LLMs are lowering the threshold for a successful attack dramatically, finding "[i]n many cases, the attacker supplied only vague, low-skill prompts and allowed Claude to fill in the gaps: researching exposed services, identifying possible vulnerabilities, writing exploit code, validating access, and harvesting data." Our Incident Response teams are documenting the increased pervasiveness of AI-powered attacks across both the public and private sector as well.

DIVERGING TIMELINES, EXPANDING ATTACK SURFACES

The widening gap between the pace of these attacks and the speed of response is the most urgent fault line to address in cybersecurity today. AI-enabled adversaries are demonstrating an ability to move far faster than defenders that rely on human speed, compressing the lifecycle of attacks from reconnaissance to exfiltration from weeks or days into hours and minutes.

The UK-based AI Security Institute's recent evaluation of Anthropic's Claude Mythos found that the frontier AI model was rapidly automating tasks—vulnerability, exploitation, multi-stage attacks—that would take skilled human professionals days of work. Sen. Mark Warner—citing a briefing from NSA Director Gen. Joshua Rudd—stated during a recent Senate hearing that Mythos "broke into almost all of our classified systems, not in weeks, but in hours" during an authorized testing exercise conducted through Project Glasswing. A U.S. official separately confirmed to the Associated Press that Mythos had identified vulnerabilities in highly sensitive government systems within hours—while noting that identifying vulnerabilities and exploiting them aren't the same thing.

79%

of federal cyber and IT leaders are Extremely or Very Concerned about adversaries using AI to accelerate cyber-attacks against their agency in the next 12-18 months.

These advances create operational asymmetry. A lone wolf attacker can replicate the labor of multiple hackers working around the clock by leveraging AI agents, at marginal cost. Meanwhile, the defender must compete while wearing the operational equivalent of a weight vest—navigating approval chains, silos, regulations, fragmented ownership, and budget constraints.

In research conducted for Booz Allen's Vellox Striker™, which simulates AI-powered adversaries, a traditional security operations center (SOC) response spanning from initial EDR alert to incident commander approval took 45 minutes. The autonomous red team had achieved full domain compromise in six. The issue is pace, and the numbers make it clear.

But speed alone doesn't capture the full challenge. Modern enterprise environments—sprawling across cloud infrastructure, edge deployments, and increasingly, embedded AI agents—have grown dramatically more complex, and that complexity demands more instrumentation and monitoring. The drive for visibility into every layer and the expanding ability to track performance across distributed systems has produced an explosion in sensor data and telemetry, even in the absence of active threats.

AI-enabled attacks compound this further: they don't just move faster, they generate more probes, more alerts, and more simultaneous vectors. SOCs built for human-speed triage weren't designed for either the ambient volume of modern telemetry or the attack-generated noise on top of it. The result isn't heightened vigilance—it is analyst overload. Signal-to-noise ratios that were already difficult to manage are being overwhelmed, and the asymmetric, low-and-slow threats most likely to succeed are precisely the ones most likely to be missed. Defenders are buried in volume while the meaningful signal slips through.

A deficit in landscape visibility highlights an additional problem. Theoretically, AI should empower attackers and defenders proportionally—and in narrow domains it does. Microsoft used AI tools to find and patch over 500 vulnerabilities in just a few months this year, with other tech firms also reporting AI-assisted surges in patch cycles and security fixes. This suggests that defenders should have an edge: They know their own systems better than attackers do.

In reality, many organizations lack the situational awareness to make that advantage operational. Environmental knowledge is fragmented. Coverage is uneven. The intelligence needed to anticipate attacker movement—rather than simply react to it—is either unavailable or not actionable at speed. As a result, vulnerabilities linger.

These three compounding pressures—faster attacks outrunning human response timelines, SOC overload degrading detection fidelity, and insufficient landscape visibility limiting proactive defense—define the challenge security leaders must now address. The strategies that follow are organized around them directly. CISOs and other security leaders need to:

Develop Faster Responses to Attack

Build human-AI teaming models that minimize human-in-the-loop delays and enable response at AI speed.

Improve Detection Fidelity

Enrich cyber data, recalibrate signal-to-noise ratios, and rebuild SOC architectures capable of detecting more numerous and asymmetric threats.

Enhance Cyber Resilience

Adopt an "attack to defend" posture that aggressively identifies and mitigates vulnerabilities before adversaries can exploit them.

THE NEW MATH OF AI CYBER SPEED

26
SECONDS
According to an MIT/ University of Naples report, an AI-supported framework discovered **13** exploit chains in **26** seconds.

34
MINUTES
In ReliaQuest's **2026** Annual Threat Report, studies of attacks found that achieving lateral movement took **34** minutes, with the fastest breakout time falling to four minutes.

1.2
HOURS
The fastest top **25%** of intrusions reached exfiltration in **1.2** hours, reduced from **4.8** hours a year ago, according to Palo Alto analysis of real-world incident response data.

10-14
DAYS
Meanwhile, the patching cadence at a typical firm might take **10** to **14** days. Organizations public and private rely on old operating systems and code bases with baked-in weaknesses, many of which may not be patchable. The programming language COBOL, for instance, is the foundation for **43%** of online banking systems and **95%** of ATM transactions—and is more than **60** years old. Layers of tech debt introduce further latency.

KICKER

The problem isn't simply attack volume. It's the shrinking decision window that legacy governance can't meet.

THE CISO'S OPERATING MODEL IMPERATIVE

AI-powered tools can accelerate detection, automate response, and surface vulnerabilities at scale—but deployed inside an outdated operating model, they will enable the wrong decisions faster. The strategies that follow require a new foundation and a new way of thinking that breaks with old assumptions and ideas underlying traditional security architecture. Technology is only part of the solution. Building this new foundation is a leadership challenge before it is a technical one.

CISOs and other security leaders must embrace this challenge. Not simply because cybersecurity is their domain, but because the changes required cut across legal, finance, compliance, IT, and business operations simultaneously. Securing agentic AI, restructuring authorization frameworks, and reorienting the workforce all require an integrator who can operate at the intersection of technical reality and organizational authority. CISOs and other security leaders arrive at this moment with exactly that mandate—and the credibility to drive the enterprise-level alignment these strategies demand. Three priorities should anchor that effort.

Reevaluate the Risk Equation

Most organizational risk models were calibrated for a different threat environment—one in which attacks unfolded over days, giving defenders time to detect, deliberate, and respond. AI has broken that assumption. When adversaries can move from initial access to full domain compromise in minutes, the window between incident and material impact may close before an organization even has visibility on the breach. SEC disclosure timelines, designed around human-speed incidents, become operationally strained. Cyber insurance policies underwritten against the same assumptions may leave organizations significantly exposed.

CISOs need to drive this recalibration explicitly, starting with an honest assessment of what the current risk model was actually built for—and where it no longer holds. That means revisiting assumptions about threat velocity, reexamining software supply chain exposure, and ensuring that zero-trust and data protection strategies are tightly aligned to crown-jewel assets. The risk model is the foundation everything else is built on. If it is out of date, every downstream decision inherits that misalignment.

Rethink Human-AI Teaming

The deeper question behind authorization and approval chains isn't simply how to move faster—it is how to rethink the relationship between human judgment and machine action in a threat environment where the adversary no longer operates at human speed. That is a more fundamental shift than process reform. It is a reconsideration of where human decision-making adds irreplaceable value, and where it has become the bottleneck that adversaries are actively exploiting.

The answer isn't to remove humans from the loop—it's to redesign where in the loop they sit. Pre-authorizing tiered response actions moves human judgment upstream, into the planning process, rather than inserting it as a gate on every containment decision in real time. Low-risk actions like quarantining suspicious files can be delegated to autonomous execution. Medium-risk actions like revoking credentials can be automated with single-human confirmation. High-risk decisions remain human—but pre-scripted against defined trigger conditions rather than improvised under pressure.

Tabletop exercises that simulate AI-speed attacks are the most effective mechanism for building this architecture: they surface gaps, clarify responsibilities, and build the cross-functional trust that allows leadership to delegate authority before a crisis demands it.

This is ultimately a workforce question as much as a policy one. Analysts who once monitored and triaged will increasingly serve as orchestrators and decision authorities—overseeing AI agents, validating outputs, and exercising judgment on the calls that machines shouldn't make alone. Building that capability requires deliberate training, cultural change, and leadership investment. The organizations that treat Human-AI teaming as a core competency—rather than a tool adoption—will be the ones that close the speed gap.

Communicate the Resourcing Reality to Leadership

The changing threat landscape has resourcing implications that boards, political appointees, and senior leadership need to understand explicitly — and CISOs need to be the ones making that case. This isn't a technology refresh cycle. It is a structural shift in the scale and velocity of the challenge.

Vulnerability management alone illustrates the point. In just its first month, Anthropic's Project Glasswing uncovered roughly 10,000 critical and high-severity new vulnerabilities across its own and partner efforts. Addressing that backlog while simultaneously defending against active threats requires sustained investment over the next two to three years—not a one-time budget line. Cyber insurance coverage, board-level risk appetite, and compliance frameworks all need to be revisited against current threat assumptions, not the ones embedded in last year's planning cycle.

CISOs who frame this conversation in business or mission terms—exposure to material breach, insurance gaps, supply chain liability, the compounding cost of reactive versus proactive posture—will be better positioned to secure the operating model changes and resources the strategies that follow require.

WHAT FEDERAL CYBER LEADERS ARE TELLING US

Booz Allen surveyed more than **100** federal IT and cybersecurity decision-makers—spanning civilian agencies, DoW branches, and the intelligence community—about their fears and readiness for AI-driven cyberattacks. Most respondents directly shape or contribute to cyber and AI strategy at their agencies. What they told us should concern every security leader, in government and commercial enterprise alike.

THE CONCERN ISN'T HYPOTHETICAL

Nearly four in five federal cyber leaders—**79%**—say they are extremely or very concerned about adversaries using AI to accelerate attacks in the next **12** to **18** months. And the worry isn't abstract. Across five specific AI-enabled threat categories, concern levels clustered tightly above 80%: adversarial AI that manipulates or corrupts agency systems led at **85%**, followed closely by AI-accelerated vulnerability exploitation (**84%**) and autonomous attack agents (**83%**).

When asked to name the single threat capability that worries them most, leaders pointed to AI-accelerated vulnerability exploitation—the shrinking window between discovery and weaponization—as their sharpest concern.

That finding maps directly to the speed asymmetry at the center of this problem. The pace of AI-enabled attacks is what keeps federal cyber leaders up at night.

THEY KNOW WHERE AI CAN HELP—AND AREN'T SURE IT WILL BE ENOUGH

Federal leaders see meaningful defensive upside from AI, particularly in threat detection and behavioral analysis, which nearly a third identified as the area of greatest near-term benefit. Automated incident response and vulnerability management followed, tied at **20%** each. The prioritization is telling: leaders are targeting AI precisely where it can compress time—detection, response, exposure management—which is exactly where human-speed operations fall shortest.

Yet optimism is tempered by doubt. Asked whether AI-powered defenses can keep pace with AI-enabled attacks, only **36%** said yes. Thirteen percent said no. The most striking figure: **50%** are simply unsure—a level of institutional uncertainty that itself signals risk.

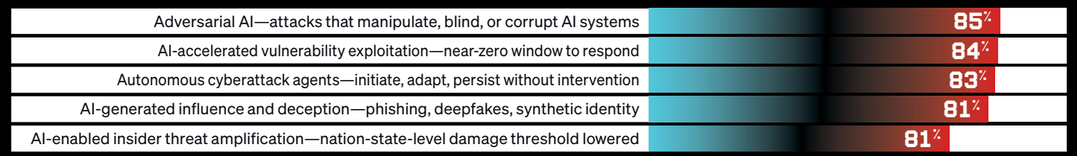
READINESS LAGS THE THREAT

The readiness picture is sobering. Just **6%** of respondents say their agencies are fully prepared, with AI-based cyber defenses actively deployed. A quarter of respondents report being substantially prepared, with capabilities underway. Nearly half—**48%**—are only partially prepared, and one in five is still in early-stage assessment.

The barriers driving that gap deserve close attention. The leading obstacles aren't budget or technology in isolation—they're structural. Integration with existing security infrastructure tops the list at **56%**, followed by concerns about tool reliability and accuracy at **52%**. Lack of skilled personnel (**36%**) and budget constraints (**33%**) trail behind. Policy and regulatory uncertainty, cited by **17%**, may be understated given how directly compliance frameworks complicate autonomous response decisions.

Read together, these barriers describe an organization that can see the threat clearly, has begun investing in the tools, and is being slowed—above all—by integration complexity and the absence of clear decision frameworks for when and how AI systems are authorized to act.

How concerned are you with each of the following AI-enabled threats? % very or somewhat concerned



WHAT THIS TELLS THE WHOLE MARKET

Federal cyber leaders are leading indicators. Commercial security leaders across critical infrastructure, financial services, and manufacturing report the same patterns in our engagements: high alarm, unevenreadiness, governance and integration as the primary friction points. The battlespace—and bottlenecks— are shared.

There are unmistakable signs of progress: a quarter of federal agencies already have substantial AI defense programs underway, and leaders are directing investment toward exactly the right capabilities. But the dominant finding is a **50**-percentage-point gap between concern and confidence. There are barriers that no single agency, and no single enterprise, can close alone.

A NEW CYBER OPERATIONS MODEL FOR AI SPEED

The challenges are clear. Attacks are outrunning human response timelines. SOC architectures are drowning in telemetry while meaningful signals slip through. And defenders lack the landscape visibility to get ahead of threats rather than simply react to them. What follows is a framework for addressing all three—not as a future-state aspiration, but as an operational imperative that leading organizations are building now.

The vocabulary we use to describe offensive and defensive tactics remains in play. Adversaries employ reconnaissance, vulnerability discovery, and exploit development to attack. Defenders rely on identity controls, segmentation, and zero-trust architecture to respond.

But AI has so warped the speed and scale at which this exchange plays out that a reimagination of foundational capabilities is called for. The three strategies below rewire operating principles to account for the new attack velocity, the analyst overload created by an eruption of data and telemetry, and a new era of managing round-the-clock risk.

How would you characterize your agency's readiness to employ AI-powered cyber defenses?

06% **FULLY PREPARED**
we have AI-based cybersecurity systems actively deployed

25% **SUBSTANTIALLY PREPARED**
we are building or procuring AI cyber defense capabilities with a clear path to deployment

48% **PARTIALLY PREPARED**
we have a strategy and have begun implementation, but meaningful gaps remain

20% **EARLY STAGE**
we are assessing options but do not yet have a formal plan or committed resources

02% **NOT YET PRIORITIZED**
we have not allocated resources to AI-enabled cyber

Develop Faster Responses to Attack

Closing the speed gap requires moving human judgment upstream and machine execution downstream. Autonomous agents can now increasingly handle endpoint detection, triage, and initial response—from alert prioritization to recommended containment action—surfacing only the decisions that genuinely require human authority. For example, testing by Booz Allen shows that automated malware analysis can compress work that would take human teams up to 10 days into just minutes, illustrating how AI-driven triage can reclaim significant time and accelerate response. Smart data pipelines that enrich telemetry with threat intelligence accelerate that further, giving agents the context to act rather than simply flag.

The authorization framework built in the operating model phase is what makes this executable. Pre-authorized, tiered response thresholds—developed across legal, compliance, and leadership before any crisis requires them—mean that when an agent recommends containment, the decision has already been made. The goal is to move human decision making into the planning cycle where it strengthens response, rather than leaving it as a gate on every containment action in real time.

Cross-sector coordination multiplies the effect. Shared threat indicators allow organizations to synchronize autonomous responses across boundaries. Federal agencies including NSA and CISA can serve as vetted intelligence sources that increase both the speed and confidence of autonomous action.

Improve Detection Fidelity

Speed of response is only valuable if detection is accurate—and right now, for most organizations, it isn't reliable enough. The sensor-overload problem identified earlier isn't simply a volume issue. It is an architecture issue. SOCs built to funnel telemetry to human analysts in sequence weren't designed for the ambient data modern enterprises generate, let alone the additional volume AI-enabled attacks produce on top of it.

Rebuilding detection fidelity requires two simultaneous moves. The first is data enrichment: smart pipelines that contextualize raw telemetry against threat intelligence, asset criticality, and behavioral baselines before it reaches an analyst. This is what converts volume into signal. The second is structural: fusing SOC and NOC functions to eliminate the handoff delays that create exploitable windows and deploying AI agents to handle routine alert clearance so analysts can focus on the detections that require genuine judgment.

Zero trust architecture is the structural complement to both. Continuous verification of devices and identities, least-privilege access, and granular segmentation constrain what an attacker can reach even after a breach. As AI agents themselves become an expanding attack surface, zero-trust principles apply to machine identities with the same force they apply to human ones.

Enhance Cyber Resilience

The most durable advantage in AI-speed cybersecurity isn't faster reaction—it's making reaction less necessary. An "attack to defend" posture is the operational expression of this principle, and it is increasingly recognized as a foundational competency rather than an advanced capability. Frameworks like Continuous Threat Exposure Management (CTEM) formalize the approach: rather than treating vulnerability management as a periodic exercise, organizations continuously assess their own attack surface from the adversary's perspective, prioritizing exposure by actual exploitability and business impact rather than raw severity scores. AI accelerates every stage of that cycle—discovery, prioritization, validation, and remediation—at a pace and scale no human team could sustain independently.

Enhanced penetration testing is where this becomes most concrete. Agentic red-teaming programs, in which AI-powered tools continuously probe for weaknesses across the environment, are generating the kind of systematic, attacker-perspective visibility that traditional annual pen tests could never provide. Leading cybersecurity programs are already orchestrating dozens to over a hundred AI agents for this purpose—not as a one-time audit, but as an always-on capability that maps exposure before adversaries can exploit it. As AI compresses the window between vulnerability discovery and weaponization, the ability to find weaknesses first is no longer a best practice. It is the primary competition.

This logic extends upstream for organizations that develop and ship software. Human-AI collaboration in secure code development reduces vulnerability introduction before production rather than remediating it after deployment—a significantly more cost-effective posture that treats resilience as a design principle rather than a remediation task.

The workforce dimension can't be separated from the technical one. As vulnerability discovery and triage become increasingly automated, analysts shift from monitors to orchestrators—overseeing agent outputs, validating findings, and exercising judgment on the decisions machines shouldn't make alone. Like the Air Force's Loyal Wingman program, in which semi-autonomous systems fly alongside human operators, AI agents are force multipliers. Building the human capability to direct them effectively—to interpret what the agents surface, challenge what they miss, and act decisively on what they find—is as important as the agents themselves.

CONCLUSION

The evidence points in one direction. AI isn't an emerging threat on the horizon—it's a disruptive reality already reshaping the cyber kill chain, compressing attack timelines from days to minutes, and widening the gap between what defenders can see and what adversaries can do. Our survey of federal cyber leaders puts a number on it: 79% are extremely or very concerned, and only 6% consider themselves fully prepared. Our work in the commercial sector finds similar sentiments. That gap between concern and readiness is the real finding—and closing it is the work.

Closing that gap requires more than deploying new tools. Risk assumptions, authorization frameworks, detection architectures, and vulnerability management practices all need to be rebuilt around the actual threat timeline—not the one last year's planning cycle assumed. The operating model has to change before the technology can deliver.

The fundamentals of cybersecurity haven't changed, and neither has the essential challenge: Find the threat before it finds you. What has changed is the speed at which that competition plays out and the organizational capacity required to win it. The frameworks exist. The strategies outlined here are being deployed by leading programs today. What determines outcomes is whether organizations treat this moment with the urgency it demands—because the adversary already is.

Brad Medairy is president of Booz Allen's National Cyber business and Andrew Turner is president for Booz Allen's Commercial Cyber business.

Survey Methodology

Market Connections and Booz Allen partnered to design an online survey of over 100 federal government decision makers/influencers, involved with IT and cybersecurity, regarding AI's impact on government cybersecurity practices. The survey consisted of 14 questions and was fielded in April of 2026. Due to rounding, responses may not add up to exactly 100%.

SPEED READ

AI-powered cyberattacks have compressed the cyber kill chain to minutes, overwhelming human-paced decision structures, drowning SOCs in noise, and exposing visibility gaps that let adversaries move faster than defenders can respond.

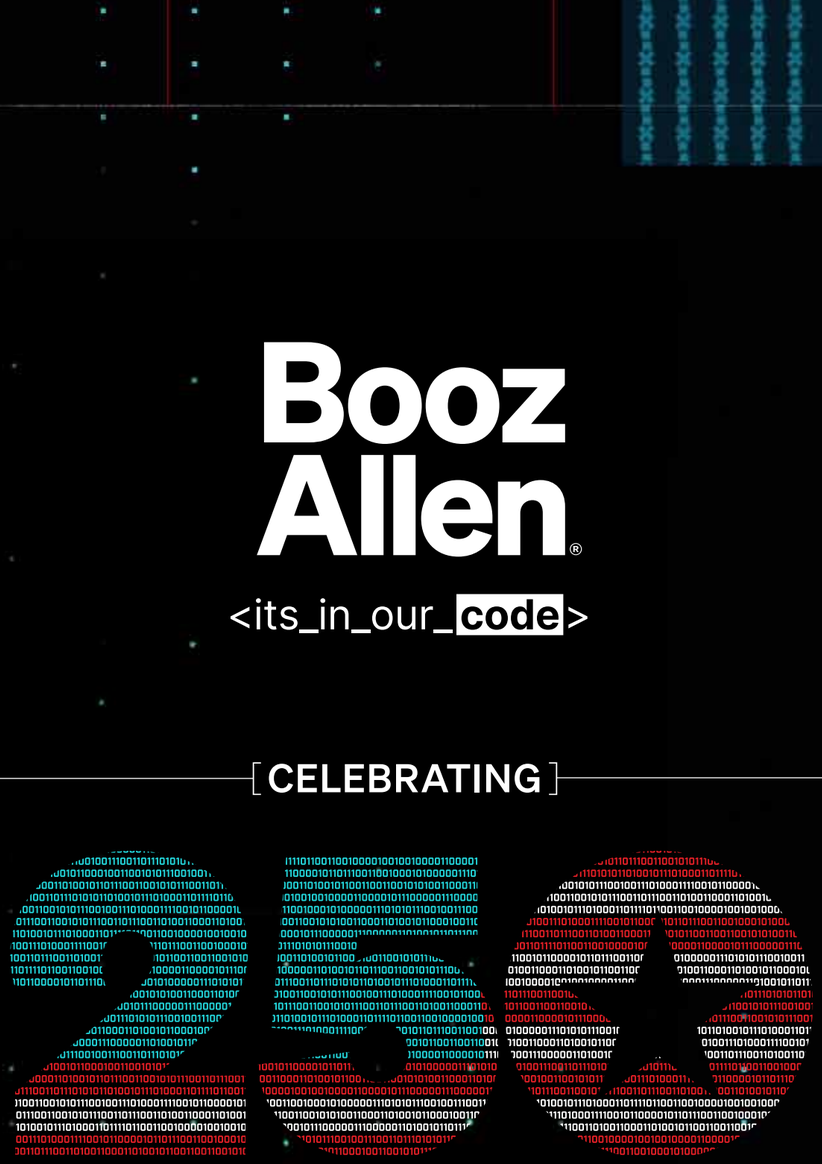
Operating at AI speed means disrupting the operating model: shifting human judgment upstream, empowering AI speed containment, enriching telemetry to restore signal to noise, and adopting continuous "attack to defend."

Exclusive survey of more than 100 federal cyber leaders quantifies the urgency: 79% are extremely or very concerned about AI-enabled attacks in the next 18 months, yet only 6% consider themselves fully prepared—a gap requiring immediate action.

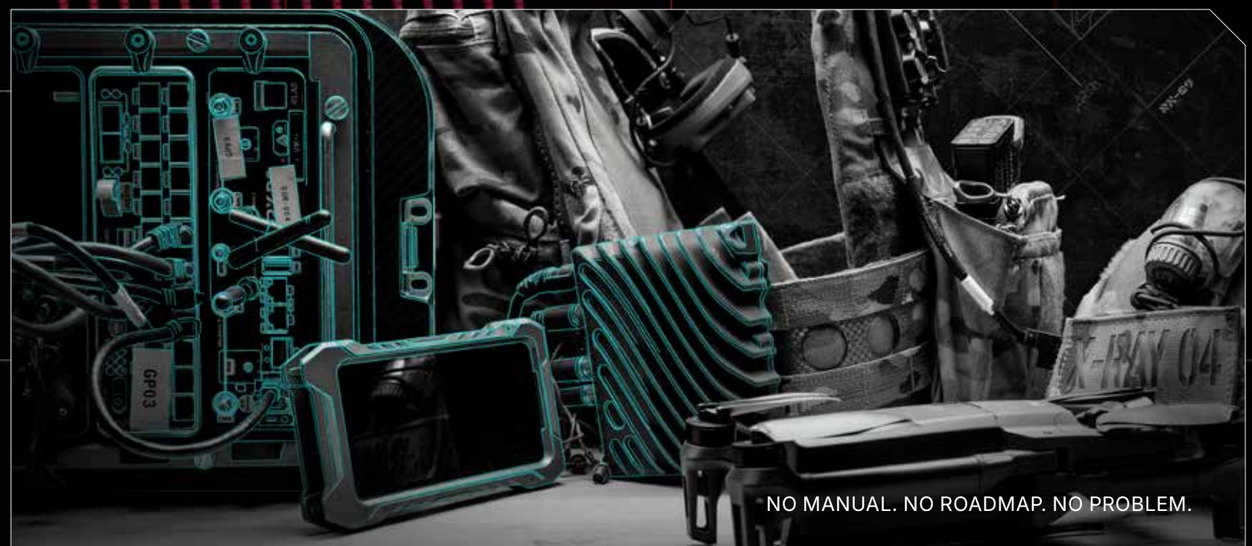
Booz Allen®

<its_in_our_code>

[CELEBRATING]



YEARS OF AMERICAN INNOVATION



How Can You Trust Agentic AI? **Start With Engineering**

Why trust must be designed, governed, and validated—not assumed

By Jonathan Lebert, Marie Nguyen, and Sahil Sanghvi

Can I trust agentic AI to handle this operation?" It's a simple question. And a reasonable one to ponder.

The push to consider this question at all is due to AI's remarkable ascent in just the last decade. It has progressed from predicting things to creating things, and now with agentic AI, to actually managing and doing things.

For IT leaders, the implications are enormous. Agentic AI can accelerate strategic planning, automate complex workflows, improve operational awareness, and augment workforces in ways that were impractical just a few years ago. These systems have the potential to transform how we analyze information, coordinate operations, and deliver services to citizens and consumers.

But as AI accumulates more capability, something counterintuitive happens: our trust in it wavers.

Upon reflection, this makes sense: as AI becomes more versatile, we assign it more complex and challenging tasks, pushing ever higher the stakes of those tasks—whether it is diagnosing a patient, countering cyberattacks, or orchestrating drone and satellite operations. And these higher stakes leave less room for error.

This ambivalence is particularly pronounced in the case of agentic AI. The very features that make agentic AI so compelling—its ability to make decisions and take actions independently, access data sources, invoke tools, interact with other systems, and autonomously adapt to changing conditions—are also what make it challenging to deploy and operate safely and securely in both government and regulated environments.

Agents, for example, introduce new identity and access requirements and, like human users, must be governed, audited, and tracked. This requires that we rethink our approach for how to manage these new actors. More to the point, we should monitor agents as potential insider threats. In fact, new agent behavioral analytic capabilities are already emerging in the marketplace to assist with this.

Another challenge is that agentic systems introduce a wide range of cyber threat vectors. Also, because they rely on large language models (LLMs), their reasoning is non-deterministic and often unpredictable. Moreover, their deployment in multi-cloud, multi-agent environments expands the addressable attack surface that must be protected.

These risks can be daunting for executives weighing whether to employ this powerful technology in service to their missions. This does not mean we should not be aggressively moving forward with agentic AI systems. As with any new technology, we need to lean into it, while also making sure we understand it and are managing how it is deployed.

So, the question becomes: Is it possible for agentic AI systems to operate securely, safely, reliably, and predictably in service to federal missions and business operations? And, if so, how?

The short answer is yes, it is possible, but it requires that we apply new approaches for how we design, build, and manage agentic AI systems.

A Closer Look at Agentic AI's Risk Profile

Building organizational trust in agentic AI requires addressing four critical risks:

Privacy: Agentic systems often operate across large volumes of sensitive information—ranging from operational data to personally identifiable information (PII) and classified intelligence. Agencies must ensure agents access only the data they are authorized to use and sensitive information can't leak through prompts, outputs, or system interactions.

Identity management, access controls, and strict data governance become foundational capabilities when agents operate autonomously across multiple systems.

Security: Agentic AI systems introduce new cybersecurity risks that extend beyond traditional application security. Because agents frequently interact with external tools and services, adversaries may attempt to exploit those integrations to gain access to systems or sensitive data.

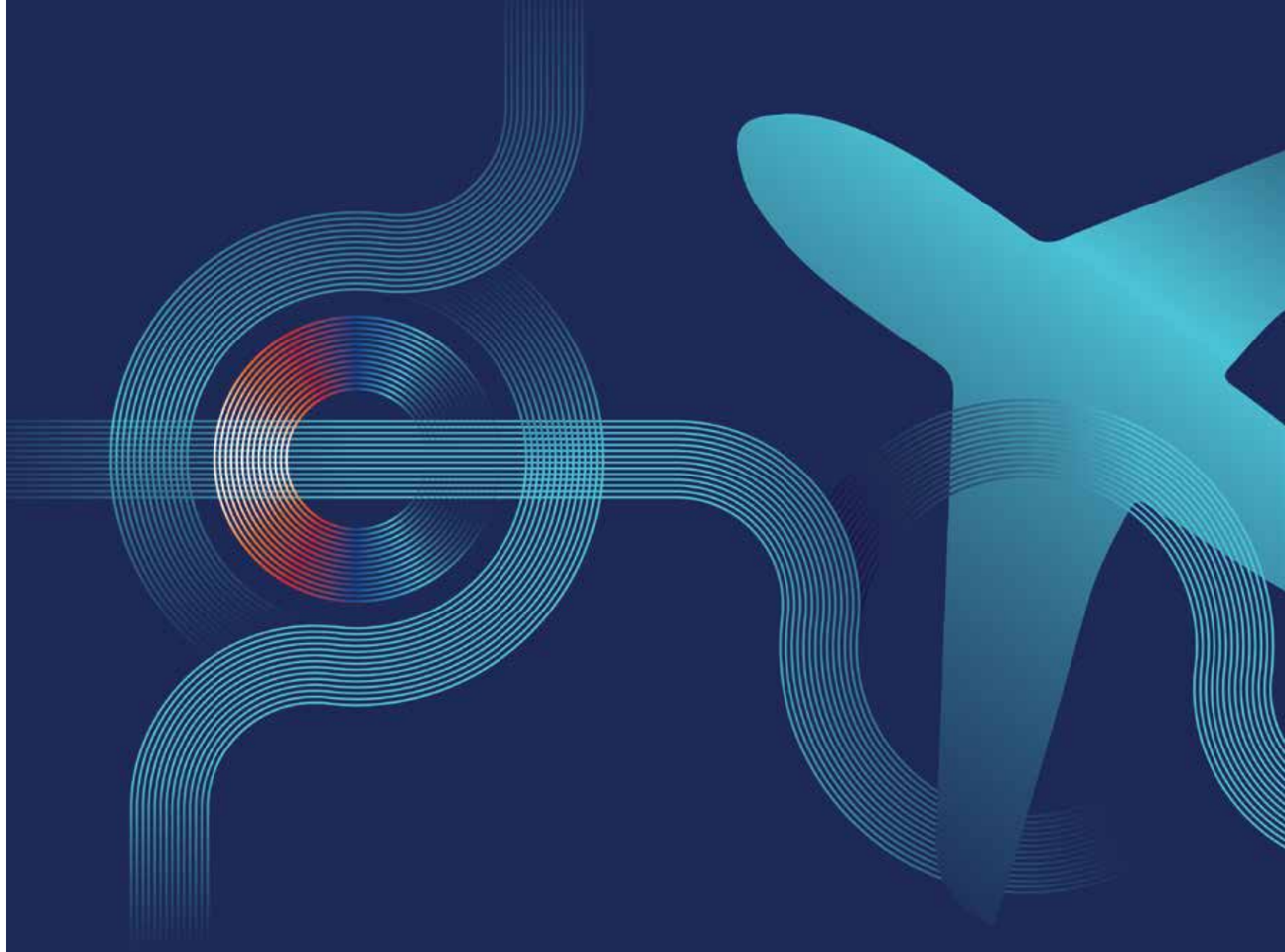
For example, attackers may attempt to craft malicious prompts that manipulate agent behavior or exploit tool integrations to execute unauthorized actions. Other threats include memory poisoning, where attackers rewrite the long-term memory that agents rely on to function; privilege escalation, where an attacker forces an agent with high-level access to perform an action on behalf of a low-level user, effectively escalating the user's privileges; and tool misuse, where attackers manipulate AI agents to abuse their integrated tools through deceptive prompts or commands, all while operating within authorized permissions.

Performance: Confidence in AI also depends on reliability. Generative models can produce hallucinations, inaccurate reasoning, or biased outputs under certain conditions. Over time, models may drift as operational environments change.

When agents are responsible not just for generating insights, but also for executing actions across systems, the consequences of these errors increase. Leaders must ensure agentic systems consistently produce outcomes that are accurate, auditable, and aligned with mission objectives.

Resilience: Agentic AI systems also introduce new operational dependencies. Multi-agent environments can involve dozens—or eventually hundreds—of interacting agents performing specialized tasks. If a component fails or behaves unexpectedly, agencies must be able to recover quickly and maintain continuity of operations.

Resilience requires robust fallback mechanisms, replay capabilities, and contingency plans that allow mission operations to continue even when AI systems encounter unexpected conditions.



Trust by Design: What Aviation Teaches Us About AI

When digital fly-by-wire systems were first introduced on commercial airliners in 1987, many observers were skeptical.

At that point, commercial airliners were mechanically controlled. Their control surfaces were operated via cables and pushrods connected to the pilot's control sticks and rudder pedals. In a fly-by-wire system, a computer collects sensor data from the pilot's controls and relays those signals via wires to actuators that move the aircraft's control surfaces accordingly.

The question many had was: "Can we trust this new technology with our lives?" Perhaps the code could fail or the system could overrule the pilot. And what if the system operated unpredictably?

Engineers didn't solve this with assurances. They solved it with design. They built layers of trust into the system via:

- Multiple redundant computers running in parallel
- Dissimilar redundancy to ensure no single event can cause all parallel channels to fail simultaneously
- Strict "flight envelope" protections to prevent unsafe actions
- Clear, deterministic system responses—no surprises
- A rigorous configuration control process to produce and maintain quality software
- Continuous validation by regulators

The result wasn't just a working system; it was a trustworthy one. Fly-by-wire aircraft like the Airbus A320 are now among the safest ever built—not because they rely less on automation, but because their automation is tightly governed, observable, and constrained.

The lesson for agentic AI is clear: Trust doesn't come from capability alone. It comes from engineering systems whose behavior is bounded, predictable, and controllable—by design.

Building Executive Trust in Agentic AI Demands a Different Approach

Before generative AI arrived, developing trust in AI meant evaluating the model thoroughly: training it, testing it, measuring it for accuracy, and then deploying it. That worked when systems were static, deterministic, and largely predictable.

Today, that model-centric playbook for building trust in AI no longer applies. Not because it was wrong, but because the technology has changed.

Now we have multiple agents working in tandem, each with specific roles; live connections to enterprise systems and external tools; real-time reasoning and adaptation; and autonomous decision making and execution.

Agentic AI doesn't just generate answers; it decides, acts, and coordinates—and most importantly, it operates within a larger, more complex system, not as a standalone model.

Moreover, the workflow of an agentic AI system can vary dramatically depending on the tools it invokes, the data sources it queries, and the other agents it coordinates with. While an agent's identity remains fixed, the permissions it exercises and the context it retrieves may differ based on the operational scenario. These dynamic interactions make strong identity governance, entitlements management, and observability essential.

In other words, the model is still important, but now there are many more components within a larger, dynamic system that must be vetted and monitored.

Ensuring Trusted Performance Requires a Wider Focus

To deliver secure, safe, reliable, and predictable outcomes with agentic AI, leaders must shift their lens from models to systems, from point solutions to architectures, and from validation events to continuous assurance.

For federal leaders, for example, there is a pathway from experimentation to mission-scale impact with confidence. But it requires thinking bigger, designing holistically, and engineering trust into the system, not just the model.

A Roadmap for Ensuring Secure, Safe, Reliable, and Predictable Agentic AI Performance

The best way to ensure trusted performance from an agentic AI system is to never trust it.

In other words, remove the trust element entirely by designing a system that, at every point of risk, has adequate safeguards built in to mitigate those risks throughout the lifecycle of that system.

This, of course, is the fundamental approach that underlies zero trust architectures as they apply to networks, applications, and critical infrastructure.

An Architectural Approach Guided by Zero Trust Principles

Applying the principles of zero trust—never trust, always verify, micro-segmentation, least privilege, and assume there is a breach—to agentic AI is a critical step toward ensuring trusted performance. Embedding these principles into your architecture will mitigate many of the biggest security-related risks that can plague agentic AI systems, such as adversarial cyberattacks, improper data exposure, and access to systems without authorization.

But zero trust, by itself, is not enough. Agentic AI introduces a broader range of risks beyond the security concerns typically addressed by zero trust frameworks. In addition to security, it's important to address performance risks that stem from potential flaws in the AI models themselves. These can include algorithmic bias, hallucinations, model drift, or anything that might generate unexpected outcomes.

Here is a roadmap for ensuring secure, safe, reliable, and predictable agentic AI performance at five key stages: risk assessment, design, development, deployment, and future-proofing your system.

1. RISK ASSESSMENT

Start by defining—and constraining—the risk surface.

Agentic AI introduces new actors (agents), new pathways (tools and APIs), and new behaviors (autonomous, non-deterministic action) that expand exposure to risk. Before a single line of code is written, be sure to understand and shape that exposure by thinking through these four factors.

Claws: Autonomous AI That Doesn't Wait—Are We Ready?

What if you had a platoon of autonomous AI agents on your computer that never stopped working?

No pausing. No waiting for prompts. No asking permission. Just acting… continuously.

While you are sleeping or at a meeting, these digital workers prepare daily intelligence briefings for you every morning; monitor contract deliverables and compliance; alert you with bulletins when priority messages arrive; reply to emails; update your social media posts; and book your next trip.

Welcome to the age of AI claws.

This technology exists thanks mainly to OpenClaw, a fast-growing, free, and open-source autonomous AI agent launched in November 2025. OpenClaw performs real work—email, scheduling, reporting, coding—on your local computer using your tools, credentials, and data on a continuous basis. It executes tasks via large language models (LLMs) using messaging platforms as its main user interface. And, perhaps most important, it retains memory over time, capturing user preferences, ongoing projects, and past email threads.

This persistent state of AI autonomy adds value, but it also scales risk. Here are some risks to consider:

Identity challenges: Who acted, the human or the machine? Without clear identities and audit trails, accountability blurs fast.

Attack surfaces expand: Always-on systems do not just increase exposure; they live in it. Prompt injection, tool abuse, and memory manipulation move from edge cases to persistent threats.

Access dangers: Claws need real credentials to deliver value. This gives them real authority to act, but there are real consequences when something goes wrong.

Unpredictable behavior: Claws learn, adapt, and infer. This leads to improved performance, but it can—and does—lead to occasional unintended actions and missteps.

Exposed supply chains: Plugin ecosystems and external integrations create invisible dependencies that serve as new paths for compromise.

NVIDIA's NemoClaw beefs up security and privacy protections around OpenClaw. It includes features such as sandboxing, default-deny network egress, policy enforcement, and audit logging. Additional focus continues on credential management, security maturity, integration stability, cost predictability, and usability.

The bottom line: NemoClaw is not yet fully production-ready, but claws are evolving rapidly and they will no doubt assume an ever-greater amount of our focus as this technology matures.

Applying Zero Trust Principles to Agentic AI

Zero Trust Architecture is built on a simple premise: never trust, always verify.

Every user, agent, device, and system interaction is treated as untrusted until authenticated and authorized—and then they are continually re-authenticated and re-authorized as needed.

In highly dynamic agentic AI systems, zero trust principles are just as critical as they are in traditional network systems. AI agents must authenticate not only with users, but with each other and with every tool they invoke. Access must be dynamically enforced based on identity, context, and mission need.

Equally important is the zero trust principle of least privilege. Each agent should have access only to the specific data and tools needed to complete its task—nothing more—to limit the impact of any compromise if there is one.

Combined with micro-segmentation and continuous monitoring, zero trust transforms agentic systems into highly controlled environments where trust is continuously validated, not assumed. In addition, an effective zero trust architecture for an agentic AI system must focus on securing the five key pillars of identity, devices, networks, applications, and data for a layered security approach that covers the entire agentic ecosystem.

What is different about zero trust for agentic systems is in how it is implemented. For example, zero trust safeguards must also be designed and implemented for the orchestration and management components of an agentic system.

Also, authenticating and re-authenticating agents as they operate can be challenging because they operate non-deterministically, at machine speed, at scale, and can be either persistent or ephemeral. So, special considerations must be given to the tools and methods used to implement zero trust protections to agentic ecosystems.

Credentials for agents
Just like human users on a traditional network, AI agents must have identities and be authorized to do what they do: interact with other agents or business platforms, invoke tools, and access data. Modern identity and access management (IAM) approaches use role- and attribute-based controls, policy-as-code, and just-in-time credentialing to ensure agents can interact only with the systems, tools, and data required for their approved workflows.

But deciding the entitlements for a particular agent may be one of the most complicated deliberations you will have. Is the agent persistent or temporary? What entitlement constraints will be placed on it? How will its behavior be monitored? Will it have a cached memory? If so, how much and for how long? In most cases, these questions will be determined by the specifics of the use case scenarios the agent will be operating within.

However those questions are decided, it will be important to assign the appropriate non-person identifiers to agents and to monitor their behaviors continuously.

Memory constraints
Memory for an agent is a double-edged sword: it improves performance but increases risk. As a result, calibrating an agent's memory parameters is akin to finding the right balance between performance and risk. Think of memory as enabling an agent's continuous learning: a repository of contextual information specific to tasks and enterprises that enhances an agent's proficiency.

The flipside is that long-term memory can unintentionally store sensitive data or inadvertently leak information. Planners will need to deliberately define what gets stored, for how long, and under what conditions it is purged. For some mission environments, such as edge devices operating on a battlefield, minimal memory may be the safer default.

Data governance
Agentic systems can access a variety of sensitive datasets. Agencies must define and enforce clear policies on data access, lineage, and usage to ensure agents interact only with data they are authorized to access. One way to support data governance is through a "composite identity," which combines the identity of a human user initiating a query and the identity of the AI agent supporting that query. A key component of attribute-based access control (ABAC) for AI systems, composite identities ensure that a user can't gain elevated privileges to access restricted data through an AI agent.

Data governance rules must be clear and enforced to ensure input validation, personally identifiable information (PII) protection, and output governance.

Kill switch for containment
Autonomy demands control. To limit the scope of a potential failure or compromise, agencies should create the ability to immediately terminate an agent, sever a connection, or halt a workflow if anomalous or unsafe behavior is detected.

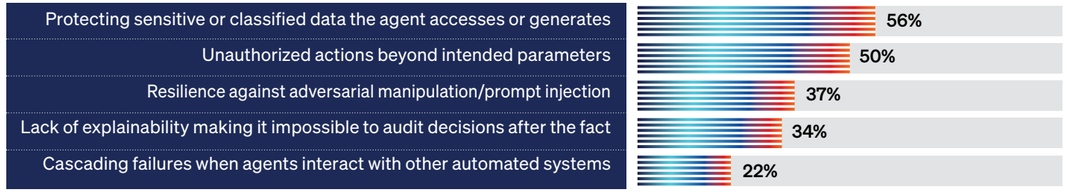
Federal Leaders on Agentic AI: Trust, Risk, and Readiness

An exclusive survey of over 100 federal IT and cybersecurity leaders found that more than half—58%—have either deployed or are piloting AI agents at their agencies. Another 22% planned to deploy in the next 12 months. However, few expressed high confidence in their ability to manage these systems securely.

CONFIDENCE IN MANAGING AGENTIC AI SECURELY

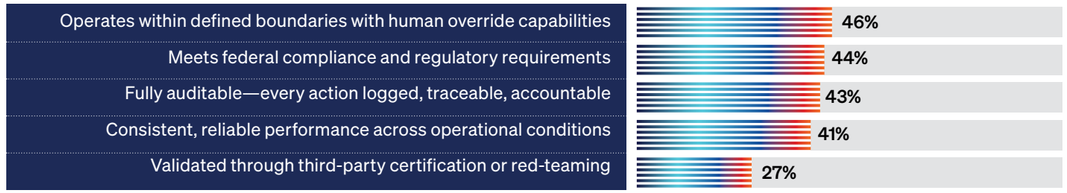
3%	25%	50%	19%	4%
Extremely confident	Very confident	Moderately confident	Slightly confident	Not confident

Top Risks Leaders Need to Address



Factors Needed to Expand Agent Use

How Can AI Agents Be Trusted?



WHO IS ULTIMATELY RESPONSIBLE FOR SECURITY INCIDENTS INVOLVING AI AGENTS?

26%	22%	22%	21%	10%
The business, mission, or program owner	The IT organization	Accountability is unclear—an unresolved gap	A dedicated AI governance function	A cybersecurity or risk function

2. DESIGN

The design stage involves thinking through the operations of an agentic system in exhaustive detail to understand what acceptable, expected performance looks like and how to engineer for risk and performance through safeguards, guardrails, monitoring, and oversight. Start by defining the business process and expected performance as well as identifying when a human needs to be involved.

Defining the re-engineered business process
Start with the mission or business workflow. How will humans and agents work together and what will be the roles and functions of each? What decisions are being made and by whom? What actions are being executed and how will they be orchestrated? Where does autonomy accelerate outcomes—and where does it introduce risk? This step requires deep stakeholder engagement to map, validate, and refine the process.

Defining expected agentic AI behaviors and performance
Agents must operate within clearly defined boundaries. What does "good" look like, and what level of accuracy, explainability, and timeliness is required? These expectations must be explicit, measurable, and aligned to mission needs. This step will inform where to design in appropriate monitoring and oversight mechanisms.

Identifying appropriate human-in-the-loop (HITL) checkpoints
Not every decision or agent should be automated. High-risk actions, ambiguous outputs, and mission-critical steps require human validation. When designing the orchestration and management layer of an agentic system, decide where checkpoints must be designed into workflows for human review, approval, or override of agent actions.

3. DEVELOPMENT

This is where trusted performance moves from concept to evidence.

Developing agentic systems requires disciplined engineering practices that embed modern cybersecurity practices and protections throughout and validate not just outputs, but also behaviors. Key components include DevSecOps, MLOps, SBOM scanning, and testing

DevSecOps
DevSecOps embeds automated security, compliance, and validation checks throughout an agent's lifecycle using techniques such as continuous monitoring, specialized LLM auditing, and sandboxed execution. When applied to agentic AI, DevSecOps combines traditional standard pipeline security—including security testing, vulnerability scanning, and compliance checks—with AI-specific testing—like data poisoning checks and agent reasoning audits—to mitigate risks like prompt injection and uncontrolled autonomy.

MLOps
AI models evolve and sometimes drift beyond their original intended scope. MLOps is the set of practices—such as data preparation, continuous monitoring, retraining, and performance management—that helps ensure the LLM models that run agentic systems remain accurate and reliable over time.

Software bill of materials (SBOM) scanning
Agentic systems rely heavily on open-source and third-party components. SBOM scanning provides visibility into these dependencies, helping identify vulnerabilities and manage supply chain risk proactively.

Tested performance
Once developed, agentic systems must be tested against real-world conditions, including edge cases and adversarial scenarios, to validate the re-engineered business processes and ensure they meet the expected behaviors and performance.

4. DEPLOYMENT AND OPERATIONS

When it is time to deploy, observability and oversight take a central role.

The goal is to ensure the performance you are getting mirrors expectations—and to course-correct when it doesn't. Performance includes all aspects of the system, including agent interactions and orchestration, ICAM enforcement, and AI model functioning.

It is critical to employ monitoring and oversight throughout the entire lifecycle of the agentic system—because AI models can drift or get corrupted at any time—but the intensity of the monitoring and oversight can be adjusted as needed.

Continuous log monitoring
Monitor and analyze the system's activity logs to see where it is performing well or not and then adjust as needed. Employ tools, such as to log tool interactions, token usage, decision paths, error rates, and latency.

Human-in-the-loop oversight
Start your deployment with a robust human-in-the-loop (HITL) presence so human experts can review and assess in real time the system's performance at key points in the workflow. Again, adjust as needed until the performance reliably meets expectations.

"Developing agentic systems requires disciplined engineering practices that embed modern cybersecurity practices and protections throughout and validate not just outputs, but also behaviors."

Model Context Protocol (MCP) Security: What Leaders Need to Know

Agentic systems rely on multiple protocols, but the most prominent one is MCP, an open standard that enables AI applications to connect seamlessly with external data sources, tools, and services. This makes MCP foundational in any agentic system.

The challenge is that MCP introduces new security risks because, unlike traditional APIs, MCP interactions are dynamic and driven by AI decision making, creating a broader and less predictable attack surface.

So, securing agentic AI necessarily means securing MCP. This requires a full lifecycle approach that combines traditional cybersecurity best practices—such as those cataloged in the NIST SP 800-53 controls—with exotic AI security techniques, such as prompt filtering, for example.

MCP security must be applied across three lifecycle phases:

- Runtime: To protect live interactions between AI and tools.
- Build-time: To prevent insecure code from entering the environment
- Deployment-time: To secure infrastructure and the execution environment

At runtime, every request must follow zero trust principles, such as authenticate users, enforce role-based access, validate connections, and apply centralized policy controls. AI-specific safeguards—such as prompt injection defenses, human-in-the-loop approvals for sensitive actions, and rate limiting—are essential to prevent misuse or runaway automation.

At build and deployment, organizations should enforce container hardening, artifact scanning, red team testing, and strict network isolation. Only vetted, approved MCP servers should be allowed.

Done right, MCP acts as a guardrail, ensuring agents operate within controlled, auditable boundaries while still enabling the flexibility that makes agentic AI powerful.

Other methods
In addition to HITL and log audits, other monitoring and oversight mechanisms include:

- Citations: Require agents to show the sources of their reasoning for verification.
- SQL-based validation tools: Fact-check AI outputs, particularly LLM-generated code or data insights, by executing, parsing, or comparing generated queries against a trusted database schema.
- Automated anomaly detection: Continuously assess performance and detect drift against established baselines.
- Stakeholder feedback mechanisms: Capture real-world input to refine system behavior.

Precautions for sensitive workloads
For highly sensitive environments, agencies should consider deploying models within internal networks rather than relying on external commercial services. Data is processed on premises, reducing exposure and strengthening control.

Start small, then scale up
Wherever possible, begin your agentic deployment with contained, low-risk use cases. Refine it as needed, prove value, build confidence, and then expand to more complex, mission-critical applications. This phased approach reduces risk while accelerating learning.

Stateful memory backups
Continually back up the stateful memory of your agentic system. Resilience requires continuity. Systems should maintain backups of workflows and memory states—allowing agents to resume operations without having to restart from scratch. This supports both operational efficiency and mission reliability. And, of course, regularly back up critical data that is accessed and modified by agents to ensure that, if something goes wrong, nothing critical is permanently lost or impacted.

Evaluate mission value
Performance is not just technical—it is operational. Agencies must continuously evaluate whether systems are delivering meaningful improvements to mission outcomes.

5. FUTURE-PROOFING

Agentic AI isn't static, nor are the risks and opportunities. Ensuring trusted performance also means preparing systems to evolve tomorrow in response to changing circumstances. Two emerging best practices can help agencies stay ahead of change:

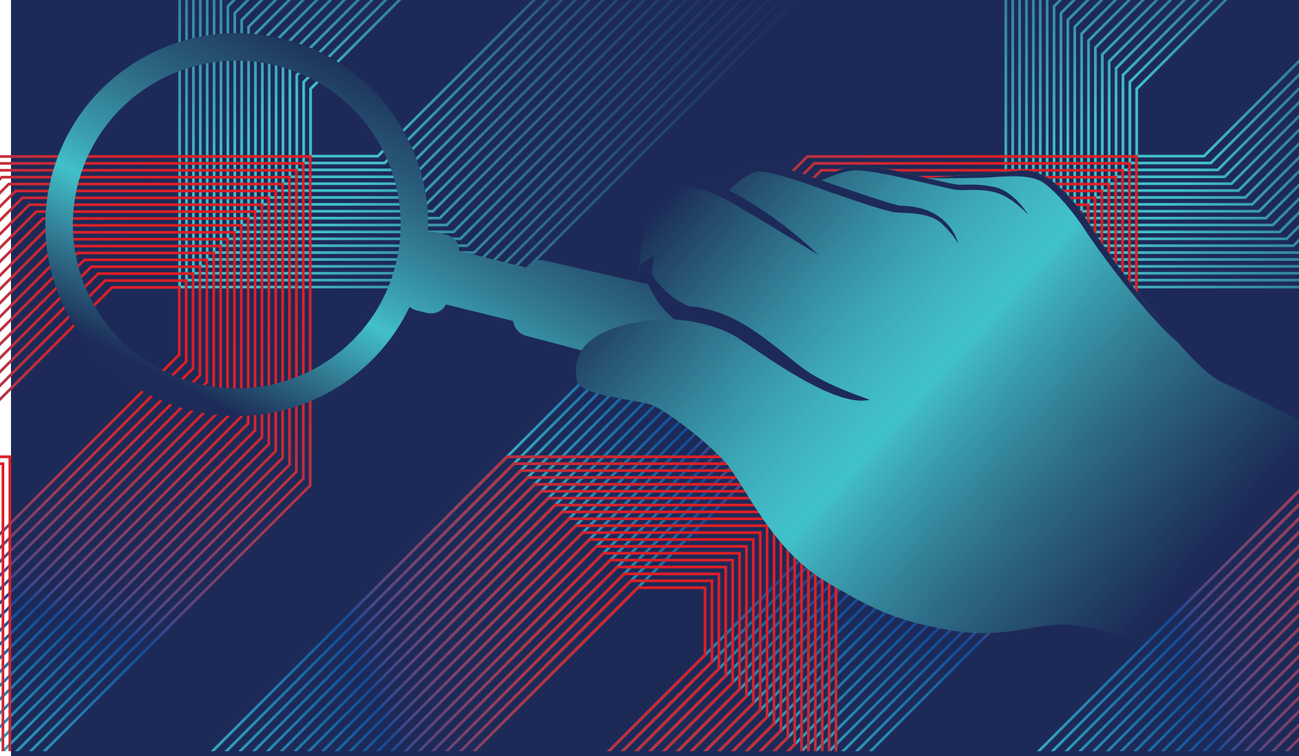
Audit agents for high-risk operations
In safety-critical software engineering, "dissimilar redundancy" is a long-standing approach: Two independent implementations perform the same operation, and discrepancies trigger immediate investigation. Agentic AI needs an equivalent.

Audit agents—ideally running on different LLMs or toolchains—can independently re-execute sensitive operations such as data mutation, privilege escalation, financial actions, or mission-critical updates. When results diverge, the system flags the action for review. This ensures autonomy never outruns accountability, particularly when agents have real entitlements and access to operational systems.

Model and component agility
The pace of AI innovation makes LLM agility as essential as cryptographic agility has been for secure systems. Architectures should be designed and built so agencies can swap out models—or simulation engines, rendering engines, mapping services, or data stores—with minimal friction.

New models will excel in some areas, lag in others, and sometimes require targeted tuning or retrieval augmentation, much like onboarding and training a new employee. Agility ensures agencies can quickly pivot to safer, faster, cheaper, or more robust models as technology, costs, or operational requirements change. But it's not enough to simply design model agility into the system; it's important during the operational phase to leverage that agility by keeping models and other system components modern and optimized for peak performance.

The goal is to understand and govern emerging technologies while deploying them quickly and correctly. When upgrades or new technology adoption lags, shadow IT—and the risks that come with it—grows. These two practices—designing in model agility and auditing agents for sensitive tasks—will give leaders confidence not just in today's agentic systems, but in whatever comes next.



Red Teaming: A Critical Component of Trusted Agentic Performance

Agentic AI systems are powerful. But that power comes with risk—and that's where red teaming comes in.

At its core, red teaming means thinking like an attacker, probing agentic systems with specialized prompts and tools to see how they might fail, misbehave, or be exploited, ideally long before real attackers do.

Employing red teams with expertise in agentic systems is critical to ensuring they perform as needed because traditional security controls don't fully account for the additional vulnerabilities and risk that the AI components present.

How red teaming works

Part sleuthing, part experimentation, a red team operates like a squad of detectives. The team will start by learning everything it can about the agentic system: how it is built, what it connects to, and what it's supposed to do.

From there, the team launches targeted and exploratory attacks; injects malicious or misleading inputs (such as "memory poisoning" or context manipulation); scours the outputs and logs for the tiniest of anomalies; identifies points of vulnerability; and follows up with more experiments.

The process is iterative. Early findings lead to deeper probing, often uncovering issues no one anticipated.

Common issues that red teams discover include:

- Data exfiltration: an ability to manipulate the AI to extract sensitive information
- Tool misuse: vulnerabilities that allow agents to be directed to take harmful actions
- Hallucinations: the creation of false or misleading outputs
- Design flaws: bugs, infinite loops, or unintended behaviors
- Identity gaps: agents operating without clear accountability or traceability

Many of these are security risks; some are reliability or trust issues. All of them matter.

Through rigorous testing, red teams help organizations discover vulnerabilities before attackers do; validate whether safeguards actually work; improve system reliability and trustworthiness; and build confidence in deploying AI at scale.

The bottom line is red teaming takes assumptions out of the equation by forcing agentic systems to prove they are safe and trustworthy. And in a world where AI can act, decide, and access critical resources, that shift—from assumption to validation—is everything.

Engineering Trusted Performance to Unlock Mission Value

We now live in an era of agentic AI, and the stakes of this technology performing as needed—securely, safely, reliably, and predictably—will only grow as we continue to harness it in support of federal missions.

The path forward is not to slow down adoption because of this, but rather to change how we engineer trust in these new capabilities. Building executive trust in agentic AI systems is not a model problem; it is a systems challenge—one that spans the entire lifecycle, from risk assessment through design, development, deployment, and operations.

It requires a fundamental shift in mindset from trusting outputs to engineering outcomes, from validating models to governing systems, and from point-in-time testing to continuous assurance.

In short, as the stakes of agentic AI performance increase, so too will the effort and complexity needed to ensure that performance meets our expectations.

This lifecycle approach helps resolve one of the most pressing concerns that federal leaders must grapple with as they consider how to safely harness this promising technology on behalf of their missions and business operations: how to balance autonomy with accountability.

The answer is not to choose between them. It is to architect for both.

When agentic systems are built with zero trust principles at their core and embedded with clearly defined guardrails, least-privilege access, and continuous observability, autonomy becomes bounded and predictable. Agents can act independently, but only within the scope they are authorized to operate.

At the same time, accountability is embedded into the system itself. Every action is logged. Every decision is traceable. Human-in-the-loop checkpoints ensure oversight where it matters most. And built-in safeguards—such as escalation thresholds and kill switches—provide immediate control when conditions deviate from expectations.

In this model, autonomy is earned, accountability is enforced, and trust grows over time. This is not just a technical approach. It is an operational one.

Both federal agencies and commercial enterprises are pursuing calculated investments in agentic AI with the expectation that it will deliver high value to their missions and business operations. This systems-wide architectural approach to building assured performance in these agentic AI systems offers a path toward achieving that goal.

Sahil Sanghvi, a vice president; Jonathan Lebert, a solution director; and Marie Nguyen, a senior AI solution architect, lead efforts in the company's Chief Technology Office to advance next-generation agentic AI solutions.

The authors would like to thank Eric Ferguson, Matt Keating, Bill Murrin, and Amy Wagoner for their important contributions to this article.

Survey Methodology

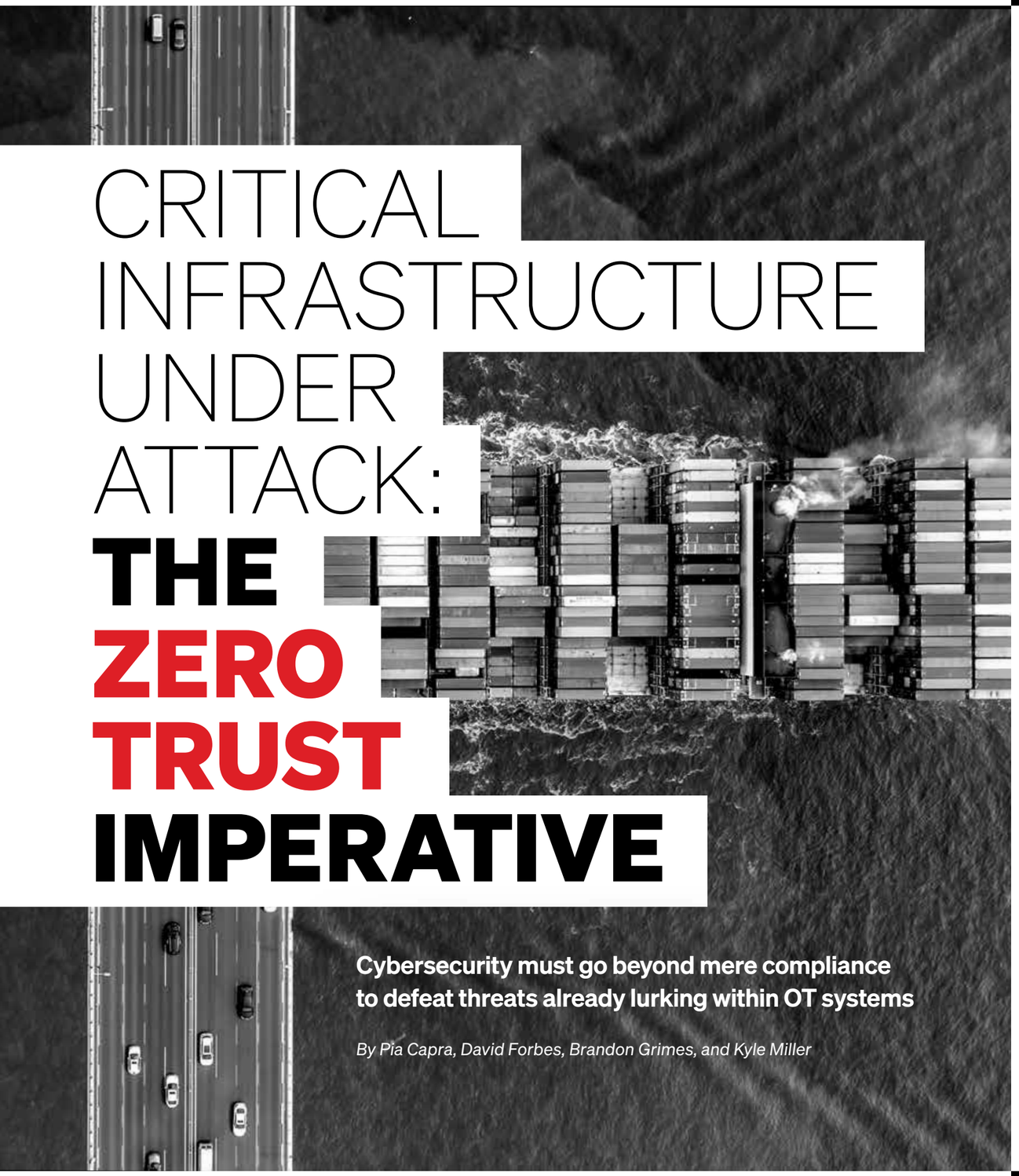
Market Connections and Booz Allen partnered to design an online survey of over 100 federal government decision makers/influencers, involved with IT and cybersecurity, regarding AI's impact on government cybersecurity practices. The survey consisted of 14 questions and was fielded in April of 2026. Due to rounding, responses may not add up to exactly 100%.

SPEED READ

Agentic AI is rapidly reshaping government and commercial enterprises, but its expanding capabilities also amplify risk, uncertainty, and the need for new forms of executive trust.

Trust can't be assumed or promised—it must be engineered, producing agentic systems whose behavior is bounded, observable, and controllable by design.

Leaders can ensure secure, safe, reliable, and predictable agentic AI performance by adopting a lifecycle systems approach that embeds identity governance, guardrails, monitoring, red teaming, and continuous assurance throughout the ecosystem.



CRITICAL INFRASTRUCTURE UNDER ATTACK: THE ZERO TRUST IMPERATIVE

Cybersecurity must go beyond mere compliance to defeat threats already lurking within OT systems

By Pia Capra, David Forbes, Brandon Grimes, and Kyle Miller

The Department of War (DoW) deserves credit for moving zero trust from concept to implementation in enterprise IT. A 2025 Pentagon memorandum institutionalized governance, set department-wide expectations, and pushed zero trust adoption as an operational priority, not just a cybersecurity goal.

But that progress also reveals a sharper truth: mission assurance does not stop at traditional IT boundaries. The ability to support deployment, sustainment, and combat increasingly depends on operational technology (OT) and on privately owned critical infrastructure, such as power, water, transportation, logistics, and critical manufacturing systems. Adversaries are already targeting these systems. For some civilian networks, at a minimum, they have successfully infiltrated and targeted assets with pre-positioned threats that could be used for disruptive attacks or coercive leverage in a crisis.

That is a critical reason why DoW's recent *Zero Trust for Operational Technology Activities and Outcomes* matters so much now: it formalizes the application of zero trust principles to industrial control systems and other OT environments. The question is no longer whether zero trust principles belong in OT environments but how to apply them in systems where safety, availability, and mission continuity are non-negotiable. The importance of these systems was reinforced by the recently released *President Trump's Cyber Strategy for America*, which elevates protection of critical infrastructure and the defense industrial base as strategic imperatives.

In this article, we assess how viable the new zero trust for OT mandates are in practice, where agencies can streamline implementation across IT/OT boundaries, and what it will take to translate compliance activity into the outcome that matters most: resilient operations under real-world conditions in a contested environment.

The Erosion of Long-Term Security Expectations and Emergence of New Paradigms

For decades, a primary security model for OT was straightforward: *keep it isolated*. Organizations engineered industrial control systems (ICSs), supervisory control and data acquisition (SCADA) platforms, building-automation networks, energy management systems, and weapon system components for reliability and uptime while maintaining "air gaps" to ensure nothing malicious could reach them. But air gaps eventually proved to be ineffective. In 2010, Stuxnet crossed the air gap at Iran's Natanz enrichment facility via a compromised USB drive, installing itself on the programmable logic controllers (PLCs) controlling centrifuges and causing physical destruction while feeding operators false sensor data. This incident demonstrated that adversaries could reach isolated systems without network connections.

At the same time, the business and mission benefits of IT/OT convergence—real-time operational visibility, predictive maintenance, remote monitoring, and the efficiency gains of industrial Internet of Things (IoT) integration—were becoming more compelling. What has replaced the air gap is a busy connectivity zone that adversaries enter and don't leave, extending dwell time often for months or years before detection, with consequences that extend beyond the enterprise into critical infrastructure, defense systems, and national security.

The scale of these once-quiet exploits has come into sharper focus. In 2024, roughly 70% of cyberattacks involved critical infrastructure, according to the House Homeland Security Committee. Facilities suffering cyber-driven physical disruptions jumped 146% year-over-year, from 412 sites in 2023 to 1,015 in 2024. China's cyber espionage operations targeting industrial and manufacturing sectors have recently surged 150%. And overall hacktivist campaigns against ICSs increased 51% into 2025. These attack spikes underscore the fragility of the nation's critical infrastructure.

To meet policy mandates, agencies will now implement zero trust—the cyber architecture already transforming federal IT—in OT systems where failure, in the worst cases, can have negative effects such as impaired warfighting readiness, contaminated water supplies, electric grid blackouts, or, in very rare instances, loss of life. The objective is not just compliance. It is to achieve cyber resilience that encompasses prevention, mitigation, and recovery from attacks, all to eliminate the possibility of catastrophic failure from malicious activities.

UNDERSTANDING ZERO TRUST

ZERO TRUST refers to an evolving set of cybersecurity paradigms that move defenses from static, network-based perimeters to focus on users, assets, and resources. A zero trust architecture grants no implicit trust to assets or user accounts. Key principles include: "Assume a breach;" "Never trust, always verify;" and "Allow only least-privileged access based on contextual factors."

To date, federal zero trust implementation efforts have focused primarily on IT environments. However, the threat landscape, the evolving federal policy framework, and technology advances have driven a natural progression of zero trust principles to OT.

OT is the broad category of programmable systems and devices that interact with the physical environment or manage devices that interact with this environment. These devices and systems thus influence "the real world." OT encompasses ICSs, weapon systems (which typically are OT with embedded IT components), building automation systems, transportation systems, communications systems, physical access control systems, and physical environment monitoring and measurement systems, among others.

Zero trust is an enabling framework, not an end in itself; resilient infrastructure is the ultimate goal.

Fragile Infrastructure, With Access Assumed

In a global networked environment where devices outnumber people two to one, new vulnerabilities are emerging. Everyday physical systems—HVAC, traffic lights, medical devices, emergency management systems—are now integrated with IT networks, expanding the attack surface. Recent cyberattacks against OT share key traits: they exploit this convergence of IT and OT, use legitimate credentials and native tools that defeat conventional detection, gain initial access through internet-facing devices with known vulnerabilities or default credentials, and establish dwell times from months to years. A 2024 report found that 98% of organizations experiencing cyberattacks said their IT security incidents also affected OT environments, indicating collapsed domain boundaries.

Volt Typhoon is the most well-known example of a modern OT attack that includes the use of pre-positioning. This Chinese state-sponsored operation has been active against U.S. OT infrastructure since at least 2021. In 2025, cyber firm Dragos disclosed the first confirmed Volt Typhoon compromise of the U.S. electric grid: an approximately 9-month intrusion at a Massachusetts utility in 2023 during which the adversary group exfiltrated data on OT operating procedures and grid layout to build knowledge for exploitation and impact later.

Rather than deploying custom malware, Volt Typhoon works mostly through living-off-the-land (LOTL) techniques involving PowerShell, Windows Management Instrumentation, and native system utilities that function the same as everyday administrative tools. Access comes through vulnerabilities in internet-facing network appliances, with traffic routed through compromised home routers to obscure its origin. Signature-based detection, a key element of conventional cyber tooling, fails against this approach by design. According to the MITRE ATT&CK for ICS framework, the observed techniques for Volt Typhoon's previous activity generally do not extend to the impact tactic.

A similar Chinese state-sponsored effort, Salt Typhoon, demonstrated the same emphasis on stealth, persistence, and strategic intelligence collection. In 2024, Salt Typhoon compromised at least nine major U.S. telecommunications providers, enabling the interception of senior government officials' communications in real time. While the campaign focused on telecom rather than OT, the objective—deep, long-term access to critical infrastructure for operational advantage—mirrored the pre-positioning behaviors seen in Volt Typhoon operations.

Similarly, Russia's Sandworm unit—the GRU operation behind the 2015–2016 Ukrainian power grid attacks—has expanded its targeting to NATO infrastructure. In early 2024, Sandworm operatives or proxies working behind hacktivist personas manipulated water treatment controls at a Texas facility. By December 2025, the group apparently had deployed DynoWiper malware against Polish combined heat-and-power plants and a renewable energy coordination system during peak winter demand. Defenders contained the attack before outages occurred, but the operation confirmed Russia's readiness to strike allied OT networks.

"IN A GLOBAL NETWORKED ENVIRONMENT WHERE DEVICES OUT NUMBER PEOPLE TWO TO ONE, NEW VULNERABILITIES ARE EMERGING."

And Iran is another adversary known to target critical infrastructure. In late 2023, its CyberAv3ngers, an Islamic Revolutionary Guard Corps-affiliated group, was found to have exploited default credentials on PLCs at a Pennsylvania water facility's booster station. With the launch of open hostilities with the United States, these activities have reportedly escalated, with the Cybersecurity and Infrastructure Security Agency (CISA) issuing a security advisory warning that the group continues to "exploit weaknesses in operational technology (OT) devices accessible from the internet."

Clarifying the threat picture, researchers have demonstrated that anyone with an internet connection can map infrastructure dependencies across military installations and the private sector using only publicly available information in just 20 to 30 hours per installation. Search engines such as Shodan and Censys freely index exposed OT devices; Metasploit, integrated with Shodan, provides ready-to-use exploit modules for common OT systems. Although DoW OT systems are rarely exposed directly to the internet, procurement records, contract award announcements, and other open-source data routinely surface sensitive infrastructure details that could inform targeting. And the private sector has similar exposure. As a result, security leaders should assume that adversaries already possess accurate information about OT infrastructure, including viable cyberattack paths.

PRE-POSITIONING refers to the practice by which adversaries—typically nation-state actors—infiltrate target networks months or years before any intended offensive action, establishing persistent, covert access that they can activate when geopolitical or operational circumstances align.

As they wait, pre-positioned actors map the environment, identify critical systems, and embed themselves deeply enough to establish persistence.

In this way, they maintain their ability to degrade, disable, or destroy critical systems at a strategic future moment.

Easy to Disrupt, Hard to Destroy—For Now

The most impactful OT security incidents don't always require adversaries to directly attack infrastructure. The 2021 Colonial Pipeline attack is one example. Ransomware compromised billing and invoicing systems, not OT. The pipeline remained functional, but operators shut it down voluntarily because they could no longer safely monitor or measure product flow. In a similar incident, cybercriminals gained access to a natural gas compression facility's IT network via a spearphishing link before pivoting to the OT network and deploying ransomware across both environments.

Only a handful of nation-states—mainly China and Russia—have demonstrated the ability to conduct truly destructive OT cyberattacks. By contrast, ransomware and hacktivist groups, despite growing OT activity, typically lack the engineering expertise and the patience required to cause catastrophic physical damage. In addition, causing this kind of damage to OT systems would likely undermine their financial objectives. However, these groups remain the most consequential threat most critical infrastructure and commercial operators face, responsible for some of the largest OT-related financial impacts on record, including the JBS Foods production halt and disruptions at Jaguar Land Rover in addition to the Colonial Pipeline shutdown.

AI is beginning to change these dynamics. Large language models accelerate analysis of exfiltrated operational data and lower the level of manpower and expertise needed for sophisticated attacks, making capabilities that once required teams of engineers more accessible.

Anthropic's Claude Mythos Preview—a frontier AI model with highly advanced capability at discovering and exploiting software vulnerabilities—offers an early glimpse of how AI will further reshape the cyber landscape. In testing, Mythos identified thousands of previously unknown flaws across every major operating system and browser, converting some into working exploits with little or no human guidance. The UK AI Security Institute found that Mythos succeeded at expert-level hacking challenges 73% of the time, and other frontier models could soon cross this threshold.

These developments have direct implications for OT. Enterprises have long relied on proprietary firmware, niche protocols like Modbus and DNP3, and physical isolation, yet these embedded systems rarely undergo the systematic security analysis Mythos automates. Many can't be patched without halting operations, creating conditions ripe for long-dwell, high-consequence intrusions. While Anthropic's defensive coalition, Project Glasswing, extends access to major IT vendors, observers note the apparent omission of OT. Meanwhile, stockpiled zero-day exploits now depreciate faster as AI-driven discovery accelerates, incentivizing earlier deployment before defenders can respond.

"ONLY A HANDFUL OF NATION-STATES—MAINLY CHINA AND RUSSIA—HAVE DEMONSTRATED THE ABILITY TO CONDUCT TRULY DESTRUCTIVE OT CYBERATTACKS."

Government and commercial leaders can frame OT risk in these two ways:

- The near-term threat, now open to a broad range of adversaries, is disruption: loss of visibility, operational shutdown, extended recovery, and eroded public confidence. Even seemingly contained incidents can produce business or mission failure.
- The longer-term, higher-consequence threat, is physical destruction, currently concentrated in nation-state actors but expanding as AI lowers barriers.

Both warrant investment, even as resource competition between OT security and traditional enterprise priorities (e.g., control system upgrades vs. new aircraft) creates difficult decisions.

Tracking Federal Policy Mandates

The federal policy framework has been moving in a clear direction, although the pace and specificity of OT-focused guidance have lagged behind the IT side. NIST SP 800-207 established zero trust architectural principles; NIST SP 800-82 Revision 3 expanded OT security guidance and specified integrated zero trust, with caveats; and DoW's 2022 Zero Trust Strategy set a target-level implementation deadline of the end of Fiscal Year 2027, later codified by DTM 25-003 in July 2025.

As mentioned, the *Zero Trust for Operational Technology Activities and Outcomes guidance*, issued by the DoW CIO's Zero Trust Portfolio Management Office in November 2025, is another consequential OT-specific policy. It defines 105 zero trust activities (84 target-level and 21 advanced) across seven pillars: users, devices, applications and workloads, data, networks and environments, automation and orchestration, and visibility and analytics. It distinguishes between an operational layer and a process control layer. This guidance empowers DoW asset owners to tailor these requirements to their systems and environments. Compliance against all 105 activities is neither required nor expected in most cases.

More recently, CISA and interagency partners—including DoW—released a joint guide on *Adapting Zero Trust Principles to Operational Technology*. Its recommendations, including asset visibility, segmentation, strong identity, and continuous monitoring, closely align with and reinforce the operational and architectural themes discussed in this article and the earlier requirements issued by DoW. The guide also elevates supply chain risk management as a core zero trust consideration. Our analysis complements this new guidance by incorporating DoW policy requirements as well as commercial implementation realities.

An updated DoW Zero Trust Strategy is expected in the near future, with additional guidance for weapons systems and defense critical infrastructure.

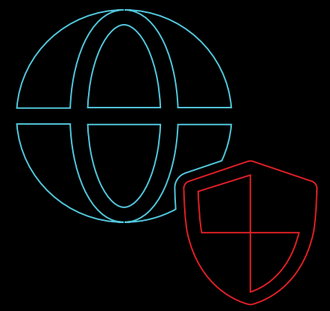
A GLOBAL PERSPECTIVE ON ZERO TRUST FOR OT

In January 2026, CISA and the UK's National Cyber Security Centre (NCSC-UK), in collaboration with the FBI and international partners, released *Secure Connectivity Principles for Operational Technology,* a joint framework comprising eight principles for designing, securing, and managing connectivity into OT environments.

Developed with agencies in Australia, Canada, Germany, the Netherlands, and New Zealand, the principles are framed as goals rather than minimum compliance requirements.

The principles address risk-opportunity balance, connectivity exposure limitation, network connection standardization, secure protocol adoption, OT boundary hardening, compromise impact limitation, comprehensive logging and monitoring, and isolation planning. The guidance addresses operators of essential services across critical infrastructure sectors and has no implementation deadline.

This guidance and DoW's *Zero Trust for Operational Technology Activities and Outcomes* share basic zero trust principles—least-privilege access, network segmentation, continuous monitoring, and strong authentication—but differ in other ways. The CISA/NCSC-UK guidance is advisory and applicable across both public and private critical infrastructure sectors worldwide.



The Complexity of OT Zero Trust

As the federal government focuses on translating zero trust principles into OT-native implementations that preserve operational uptime, agencies will need different tools, timelines, and risk calculations compared with zero trust for IT, particularly when the goal is detecting and neutralizing access that may already exist. While IT cybersecurity often focuses on preventing data breaches, OT protocols focus on keeping plants and facilities up and running.

To start, a significant complicating factor is that many of the OT environments essential to federal missions fall outside direct federal control. National security and defense operations rely on a mix of government-owned and managed systems, government-owned but contractor- or vendor-managed systems, and fully vendor-owned or -operated infrastructure. Civilian power, water, gas, and telecommunications networks that feed defense installations often sit under municipal, regional, or private governance. This creates a fragmented security landscape in which federal agencies must secure mission outcomes without full authority over underlying OT assets.

Legacy device constraints are another immediate barrier to quick zero trust implementation. OT devices in many federal and commercial environments often have lifecycles of 15 years to more than 30 years. These aging PLCs, remote terminal units (RTUs), and legacy SCADA systems were engineered for reliability and availability, not cybersecurity. Many cannot run authentication software, accept patches without operational risk, or participate in identity-based access control. Adversaries have already mapped these attack surfaces.

OT also brings a stringent requirement for availability. Where IT security treats patching and rebooting as routine, any control that causes even brief downtime on a power grid, water system, or weapons platform is likely to be unacceptable. Zero trust controls for OT must be deployed non-disruptively.

Asset visibility is another challenge. NIST has identified incomplete asset inventory as a major barrier to zero trust adoption. The challenge is harder in OT environments where devices may be geographically dispersed, decades-old, and uncatalogued. Many also communicate through specialized industrial protocols—such as Modbus, DNP3, and BACnet—that were not designed for security and are often incompatible with standard IT monitoring and security tools, creating significant interoperability gaps.

Also, organizations may be factoring in air gaps that no longer exist. Seemingly isolated OT systems can be exposed to insider threats, temporary device connections, removable media, and hardware or software with undocumented external connectivity. For example, threat detection analytics by Booz Allen found that 60% of cyber incidents of unknown origin within a global manufacturing firm were attributable to USB devices carried by insiders, requiring proactive steps to reduce this threat.

Finally, IT-OT governance silos create exploitable gaps. In 2023, an estimated 37% of ransomware attacks on industrial organizations impacted both IT and OT environments, up from 27% in 2021. Wholly separate governance structures, budgets, risk frameworks, and reporting chains for the two domains should be rethought.

FEDERAL LEADERS GRAPPLING WITH REAL CHALLENGES

According to our exclusive survey of over 100 federal IT and cyber leaders, the biggest barriers they face implementing zero trust within their OT environment include (top two responses):

Legacy OT systems that can't support modern security controls **56%**	Fragmented governance across IT, OT, and external operators **44%**	Shortage of staff with both OT operations and cybersecurity expertise **37%**
Compliance timelines or policy mandates that do not align with technical or operational realities **29%**	Operation uptime requirements that limit ability to implement changes **19%**	Limited visibility into OT assets and communication pathways **15%**

A Technical Toolkit

The capabilities below translate DoW zero trust pillars into OT-specific implementations that work within real-world constraints such as legacy devices, requirements for non-disruptive deployment, proprietary protocols, and the need for continuous operations. Together, they illustrate how a layered architecture—grounded in device identity, segmentation, behavioral analytics, and coordinated response—can materially reduce attacker dwell time and harden mission-critical OT environments.

CAPABILITY	HOW IT WORKS	KEY TECHNICAL DETAILS	WHY IT MATTERS IN OT
OT Asset Inventory and Device Identity (Devices pillar)	Establish a centralized, continuously updated inventory of every device in the OT environment before any other zero trust control	• Passive discovery only • PKI/X.509 certificates deployed where technically feasible • Flagging of legacy devices incapable of certificate support	Organizations can't enforce least-privilege access, segment a network, or detect behavioral anomalies on assets they can't see
Identity and Access Management for Machines and People (Users pillar)	Extend authentication and authorization to non-person entities; OT context includes machine service accounts and automated processes	• Attribute-based access control (ABAC) governs access • Privileged access management (PAM) enforces least-privilege	Legacy OT was built on implicit trust within the perimeter; extending identity controls to machines closes the most common lateral movement path
Network Segmentation and Microsegmentation (Networks and Environments pillar)	Macro-separate IT and OT environments, then progressively isolate critical OT functions through software-defined architectures	• Software-defined networking (SDN) and software-defined perimeters (SDP), consistent with NIST SP 800-207 • Policy Enforcement Points to isolate process control systems, safety instrumented systems (SISs), and HMIs	Limits blast radius by constraining lateral movement across the IT/OT stack, directly countering the persistence-then-pivot model used by Volt Typhoon and similar actors
Behavioral Monitoring and Anomaly Detection (Visibility and Analytics pillar)	Continuously baseline normal device and communication behavior across the OT environment	• Passive detection of anomalous OT device communication patterns, out-of-window firmware changes, and out-of-parameter control commands • Feeds a security data lake or risk operations center	Defeats living-off-the-land (LOTL) tactics that evade signature-based detection; extended attacker dwell time in OT creates intervention windows that behavioral analytics can surface
OT-Aware Data Protection (Data pillar)	Protect process data, configuration files, and engineering workstation data at rest and in transit	• Encryption enforced across the IT/OT boundary • Access controls limiting who can read or modify PLC logic • Exfiltration monitoring at the IT/OT demarcation point	Control logic and safety configurations are high-value targets; theft or manipulation of PLC ladder logic can enable physical attacks
Automation, Orchestration, and Response (Automation and Orchestration pillar)	Coordinate policy enforcement, threat response, and access decisions to reduce dependence on manual intervention	• All response actions validated in testbed environments • Policy updates implemented after testing to ensure no operational impact	OT environments are understaffed relative to their attack surface; automation closes the gap between detection and response

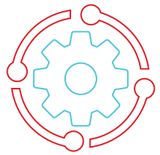
Table 1. Zero trust capabilities for OT environments, mapped to DoW's *Zero Trust for Operational Technology Activities and Outcomes* (November 2025). Together, these capabilities limit attacker movement, detect pre-positioned access, and address the legacy device constraints that make traditional patching-based security insufficient.

Note: The standard zero trust pillar for Applications & Workloads is not presented as a standalone capability in this OT-focused model. In most OT environments, application functionality is tightly coupled to physical devices (e.g., PLCs, HMIs, and control systems), and associated protections are implemented through device identity, network segmentation, and data-centric controls.

Taken together, these capabilities show that achieving zero trust in OT environments is less about isolated controls and more about building a coordinated, layered architecture designed for adversaries that may already be inside. OT resilience depends on visibility into every device, restrictions on how those devices communicate, continuous monitoring that exposes subtle behavioral shifts, and carefully orchestrated response actions that avoid operational disruption.

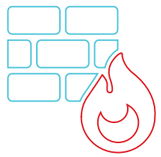
The Road to Resilience

Knowing how to secure an OT environment is not the same as being able to implement those protections at scale, especially when systems cannot be taken offline. Achieving resilience depends on disciplined execution: turning architectural principles into sustained operational change through effective planning, engineering, and monitoring. The phases below outline the practical steps organizations must take to close the gap between zero trust intent and measurable, mission-aligned resilience:



ASSESS AND ROADMAP
Establish a defensible starting point by assessing OT cybersecurity maturity, identifying gaps, and hunting for threats already inside the environment. Then build a prioritized roadmap that sequences investment and guides the transformation ahead.

DESIGN AND ENGINEER
Translate assessment findings into the architecture by defining target-state environments, selecting OT-appropriate tools, and embedding secure-by-design principles. This engineers security into the OT systems rather than retrofitting them after deployment.



IMPLEMENT AND REMEDIATE
Execute the designed architecture by deploying segmentation; hardening networks, firewalls, and individual systems and devices; and progressively reducing attack surface without disrupting operational uptime.

DETECT AND RESPOND
Move from passive defense to active resilience by standing up threat reduction programs, deploying managed detection and response capabilities, and establishing OT-specific incident response protocols that neutralize threats.

Use Case: Deterrence in Action

Recently, a large industrial company undertook a sweeping modernization of its OT security, beginning with an assessment that revealed blind spots across thousands of devices. From there, the organization launched a focused effort to strengthen foundational controls—rapidly identifying OT assets, segmenting critical networks, and deploying hardened endpoint protections across its plants. These early moves brought previously unmonitored systems into view and quickly reduced risk without disrupting production.

As the program matured, the company expanded to global-scale deployments, introduced stricter removable-media governance, and built a dedicated OT threat-detection capability that sharply cut malicious activity. Just as important, a culture of cyber awareness took root, with widespread adoption of secure practices and a growing community of OT security champions. Soon the organization had established a resilient, zero trust–aligned OT posture that could sustain itself long after the initial transformation.

Beyond the Perimeter: The Whole-of-Nation Challenge

Federal agencies can mandate zero trust for systems they own, operate, or fund—but, as mentioned, much of the critical infrastructure that underpins national defense sits outside federal authority, creating a need for cross-sector collaboration. In many cases, these systems and networks are operated by municipal utilities, investor-owned companies, and vendors whose decisions on segmentation, authentication, or patching remain independent of federal control. While these organizations often maintain advanced cybersecurity protocols, they don't necessarily align with federal requirements.

Closing this gap requires mechanisms and collaborations that extend influence beyond ownership boundaries while respecting the decentralized nature of U.S. infrastructure. Federal procurement can incentivize higher OT security standards across the commercial market. A more empowered CISA—equipped with clearer authorities, scalable funding, and the ability to set specific baseline requirements for high-risk sectors—can shift voluntary coordination toward measurable outcomes. Finally, harmonizing federal expectations with widely adopted commercial frameworks and industry-specific standards can reduce compliance friction, accelerate adoption, and create a shared national baseline. Together, these changes would transform today's advisory model into a whole-of-nation zero trust posture capable of securing the interconnected civilian-military environment.

Conclusion: Accelerating Beyond Compliance

The need to extend zero trust to OT is now foundational to national resilience. DoW has set the direction and will now move to validate controls, reduce implicit trust, and rigorously close pathways before they can ever be exploited.

For other federal agencies, OT weaknesses rarely stay isolated given interdependent missions, shared infrastructure, and cross-sector dependencies. Agencies must align policy, funding, procurement, and incident response in ways that elevate zero trust to a mission-readiness requirement, strengthening their own systems and the broader federal ecosystem.

For the private sector, the challenge is to treat zero trust as both a business imperative and a contribution to national security. This means confronting difficult tradeoffs: modernizing aging systems, prioritizing visibility and segmentation, and investing ahead of disruption. Organizations that act early will maintain continuity under pressure.

Across all sectors, the real test is execution. Deeply understanding the architecture is the starting point; resilience comes from sustained, disciplined implementation. Beyond compliance milestones, progress will center on continually strengthening the ability to detect, contain, and neutralize threats before they affect operations. Meeting that challenge will determine how rapidly the nation's critical infrastructure can adapt to and overcome the latest OT threats in an increasingly contested environment.

David Forbes and Brandon Grimes are leaders in Booz Allen's National Cyber business supporting government missions, and Kyle Miller and Pia Capra are leaders in the company's Commercial Cyber business.

COMMERCIAL OT SECURITY: WHAT LEADERS NEED TO KNOW

- In the commercial OT world, a projected $5 trillion wave of U.S. industrial modernization and new construction—driven by reshoring, geopolitical pressures, and tariff realignment—presents an opportunity to build security into critical systems from the ground up rather than retrofit later at far greater cost.
- Ransomware remains the primary risk for most commercial operators, while nation-state actors increasingly target energy and oil and gas sectors. Like government environments, many commercial industrial verticals are subject to cyber regulations, such as the electric sector, chemical facilities, and pipelines.
- Commercial control systems carry 20-to-30-year lifecycles and cannot be taken offline for patching. Vendor and automation contractor constraints limit security controls. And the talent pool of professionals who understand both industrial operations and cybersecurity is usually thin.
- For CISOs navigating this environment, it's important to establish comprehensive asset visibility to understand your environment; implement network segmentation to create defensible barriers between IT and OT environments; and identify the operational "crown jewels" to guide risk-based investment decisions.
- Security measures that threaten uptime or production continuity will lose institutional support. The strongest programs position security as an enabler of critical operations, secure funding through business leadership rather than IT budgets, and pair outside expertise for initial buildout with smaller in-house teams.

Survey Methodology

Market Connections and Booz Allen partnered to design an online survey of over 100 federal government decision makers/influencers, involved with IT and cybersecurity, regarding AI's impact on government cybersecurity practices. The survey consisted of 14 questions and was fielded in April of 2026. Due to rounding, responses may not add up to exactly 100%.

SPEED READ

The convergence of IT and operational technology (OT) is eliminating once-protective air gaps, allowing nation-state adversaries to embed themselves inside critical infrastructure long before disruption is detected.

While few actors can yet execute truly destructive OT attacks, AI enabled techniques are accelerating disruptive campaigns—forced shutdowns, loss of visibility, and ransomware—narrowing the space between disruption and destruction.

Given the extensive challenges in implementing zero trust across critical infrastructure, organizations should view policy deadlines as starting points for sustained resilience, not finish lines for compliance.

Resilience Tops Perfection:

DESIGNING FOR FAILURE WINS

Decades in IT have taught me: Failure is a normal operating condition

By Bill Vass

For decades, technology leaders have chased uptime like it was the ultimate prize—secure five nines reliability, flawless deployments, global availability. But here's the truth: FAILURE IS INEVITABLE.

After 48 years building and operating distributed systems—from the Pentagon to Sun Microsystems, startups, AWS, and now Booz Allen—I've learned something you can count on: Systems will fail. Maybe not today or tomorrow. But eventually, they will.

Networks fail. Hardware breaks. Air-conditioning units overheat. Power drops. Software behaves in ways no one predicted. Sometimes it's a bug, sometimes a storm, sometimes an adversary. The specifics don't matter—but your response does.

If a system isn't secure and available, none of its wonderful features matter. Resilience doesn't come from preventing failure. It comes from surviving it well.

That means limiting the blast radius using segmentation and thoughtful architecture—and recovering faster every time systems break, not pretending they never will. Learning that reshaped how I think about architecture, operations, and leadership.

I've earned a few scars along the way. I've seen what breaks systems—and what keeps them going.

LESSON 1:

Assume Everything Fails

It was a hard lesson, but now I know: All hardware fails.

Once you accept that failure is normal, your focus shifts. Instead of wondering if something will break, you try to manage how much breaks when it does. Software has to be written with the expectation that the underlying infrastructure will eventually let you down.

That was easier to manage earlier in my career, when systems were centralized and had fewer failure modes. Today's reality is different. Cloud, edge, AI pipelines, and distributed data stores multiply both scale and complexity. Every dependency introduces another opportunity for things to go wrong—and the more you scale, the more exceptional conditions you encounter.

That's why resilient systems don't rely on "strong" hardware or perfect networks. They assume the opposite. Stateless, load-balanced services are more forgiving by design. If one node fails, traffic shifts. If one system degrades, the rest keep running. The system isn't reliant on any single component to stay up.

In practice, this changes how you deploy. Physical infrastructure gets distributed wide enough to survive natural disasters, so a single event doesn't take everything down. After you finish alpha, beta, and gamma testing, you don't push code everywhere at once and hope for the best. You start with one box and give it load. You let it bake. If it fails, you roll back and fix it. When you're confident, you expand to one availability zone. Then another. Then you repeat, region by region—always with the option to fail and go back to the prior stage.

Building these steps into deployment helps you find failure early and limit the impact. High-reliability organizations like AWS don't depend on luck. They're methodical. They take small, deliberate steps—every single time.

LESSON 2:

Limit the Blast Radius

The most resilient systems I've seen are intentionally segmented. They isolate services, separate workloads, and assume trust must be continuously earned, not implied.

Practices like multi-availability zone and multi-region deployment aren't about redundancy for its own sake. They're about containment. Bulkheads keep one failure from sinking the entire ship. Circuit breakers stop bad dependencies from dragging healthy services down with them. Physical separation—10 kilometers or more between facilities—reduces the risk that a single storm, power event, or fiber cut takes everything offline at once.

Zero-trust principles reinforce the same idea from a security perspective: Never assume a component is healthy just because it exists inside the perimeter.

And then there's defense in depth. Real resilience takes more than a single layer of protection. It stacks multiple, independent, and heterogeneous layers—network paths, firewalls, encryption mechanisms—so a flaw in one doesn't compromise them all.

This matters even more in national security and critical infrastructure environments, where failure is operationally disastrous. Systems have to degrade gracefully, recover locally, and keep working long enough for humans to intervene.

LESSON 3:

You Can't Fix What You Can't See

Within minutes of an outage, I can tell whether a team will recover quickly or spiral. The difference comes from visibility, experience, and discipline. The best teams measure, rehearse, and learn every time something breaks.

Teams that recover faster have deeply instrumented systems. They know what "normal" looks like, so anomalies stand out immediately. Operations calls are focused, not chaotic, because the data tells a clear story about what's failing and where.

The best teams don't just restore service and move on. They investigate what failed, why it failed, and how to make sure it doesn't happen again. In most environments, a small number of root causes drive most outages. Resilient teams identify those causes, automate them away, and keep working down their Pareto charts until issues become less frequent and less severe.

That discipline doesn't just live in incident calls—it shapes how we build systems in the first place. Booz Allen-developed platforms like Recreation.gov are distributed across regions, instrumented from the start, observable under stress, and designed to learn every time something goes wrong.

Now, AI and automation are supercharging both the process and the payoff. AI-assisted operations can accelerate log analysis, find patterns humans might miss, and speed diagnosis when seconds matter. But AI only helps if systems are observable to begin with—it needs signals to work from. As instrumentation improves, so does AI's effectiveness. Over time, with the right inputs, agentic systems could move toward self-healing—detecting root issues and automatically initiating corrective action.

LESSON 4:

Define What Matters Before It Breaks

Resilience also requires realism about what matters most. Not every system deserves the same level of protection, and pretending otherwise wastes time and money. Before an outage happens, leaders need to classify their applications, defining which are mission critical, which are business critical, and which can tolerate some downtime. Those decisions shouldn't be made mid-crisis.

Every level of resilience carries tradeoffs in cost, performance, and complexity. Live-live replication across regions is expensive. Hot-cold deployments cost less but accept some downtime. Backup-and-restore models are even cheaper, but recovery may take hours. Decide ahead of time what risks you're willing to take. Resilience should be intentional, not reactive.

CONCLUSION: BUILD FOR REALITY, NOT PERFECTION

These lessons matter even more now.

Today's systems don't live in a single data center or cloud. They span hyperscale environments, on-prem networks, edge nodes, and devices operating far from reliable connectivity. AI and autonomy raise the stakes even further.

In national security and defense missions, systems have to keep operating even when the network is degraded, jammed, or gone entirely. Resilience can't depend on constant connectivity, perfect data, or centralized control.

It must be designed from the start—and strengthened iteratively over time.

And it often begins at a much smaller scale than leaders expect.

START WITH ONE SYSTEM THAT MATTERS.

MAP ITS DEPENDENCIES HONESTLY, NOT OPTIMISTICALLY.

INSTRUMENT IT SO YOU CAN SEE WHEN IT'S UNDER STRESS.

PRACTICE FAILURE—ON PURPOSE—BEFORE IT HAPPENS.

LEARN FROM EVERY INCIDENT AND BAKE THOSE LESSONS BACK INTO THE DESIGN.

REPEAT.

Resilience isn't a one-time investment or a line item in a budget. Organizations that thrive aren't chasing perfect uptime—they're prepared for imperfection.

Bill Vass is the chief technology officer at Booz Allen—and an award-winning, visionary technology leader. He's built and operated distributed systems at massive scale for robotics startups, Fortune 100 companies, and the federal government.

SPEED READ

Because modern resilience assumes systems will fail, IT leaders need to focus on containing impact and recovering quickly—not chasing perfect uptime.

Good architecture limits the blast radius, while segmentation, zero trust, and multi-region design strengthen both operational resilience and security.

Teams that measure, rehearse, and learn recover faster—leveraging visibility, disciplined postmortems, and automation to turn failure into continuous improvement.

BOOZ ALLEN AUTHOR INFORMATION

Special thanks to all the leaders from within Booz Allen who contributed to this collection and provided technical and mission expertise along the way.

Pia Capra
Commercial Cyber

Wyatt Chaffee
Chief Technology Office

Livia Dworaczyk
Tech Scouting & Intelligence

David Forbes
National Cyber

Brandon Grimes
National Cyber

Tom Hanson
Chief Technology Office

Meghan Hauser, Ph.D.
Tech Scouting & Intelligence

Jonathan Lebert
Chief Technology Office

Brad Medairy
National Cyber

Kyle Miller
Commercial Cyber

Marie Nguyen
Chief Technology Office

Bryce Pippert
Ventures & Partnerships

Sahil Sanghvi
Chief Technology Office

Andrew Turner
Commercial Cyber

Bill Vass
Chief Technology Office

VELOCITY, A BOOZ ALLEN PUBLICATION

VELOCITY EDITORIAL

Executive Editor: John Conley

Marketing Strategy & Program Lead: Danielle Epting

Contributing Writers: Elana Akman, Melissa Blevins, Jim McDermott, Emily Primeaux, Chris Ross, and Steve Watkins

Promotion: Heather Reilley

MARKETING & COMMUNICATIONS

Senior Vice President, Strategic Communications & Channels: Abby Vaughan

Vice President, Strategic Communications & Channels: Julie McGregor

Brand, Creative & Experience Director: Kari Merkel

Design Director: Jon Geurra

Media Director: Jessica Klenk

For media inquiries, please contact: **media@bah.com**

ART & DESIGN

Creative Direction: Chris Walker

Art Direction: Brody Rose

Graphic Design: Sarah Bell, Matt Dunham, David Everett, and Kim Lofgren

WED DEVELOPMENT

Creative Director: David Armentrout

Web Design and UX: Brianne Cochran and Ashley Lowe

Web Development: Brian Gibbons and Ali Mousa

ACKNOWLEDGEMENTS

Our sincere appreciation goes to the many individuals from content, digital, creative, brand, media, legal, web, accessibility, and project management whose support and expertise helped shape this edition of Velocity.

ABOUT BOOZ ALLEN

Booz Allen is an advanced technology company. We build commercial-grade products and solutions for America's most critical defense, civil, and national security priorities. For more information, visit **boozallen.com**. (NYSE: BAH)

Booz Allen®