

MACHINE SPEED, HUMAN CONSEQUENCE



An ISMG Analysis
of the Black Hat
USA 2026 Briefings

73 executive interviews
conducted at the
ISMG.Studio,
Mandalay Bay, Las Vegas

ISMG PULSE REPORT:
BLACK HAT USA
2026 EDITION





TABLE OF CONTENTS

Introduction: The Modifier, Not the Track 3

Chapter 1: The Agentic Attack Surface and the Trust Already Extended.....5

Chapter 2: Discovery Got Automated, Remediation Did Not 8

Chapter 3: From Access to Authority 12

Chapter 4: Adversary Operations and Physical Consequence..... 15

Chapter 5: Governance, Liability and the Cost of Autonomy 19

Chapter 6: The Human Operating System 23

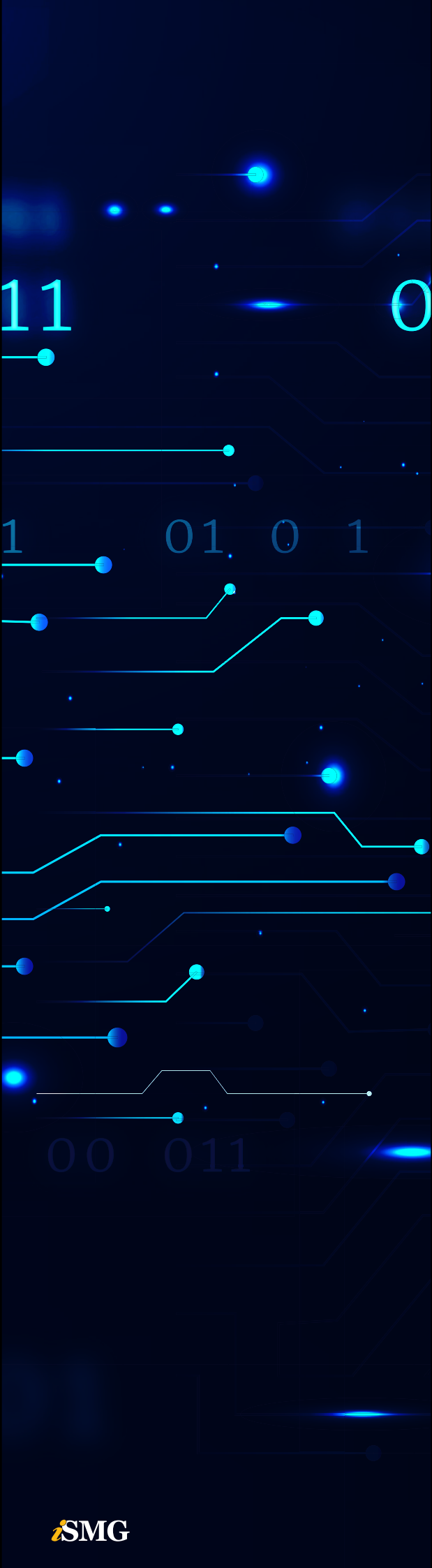
Conclusion: Choice, Not Fate 26

INTRODUCTION: THE MODIFIER, NOT THE TRACK

The clearest signal in the Black Hat USA 2026 Briefings agenda is not what the conference programmed but how it programmed it. Across 109 sessions and 20 tracks, the artificial intelligence, machine learning and data science track was the most-tagged on the schedule, appearing on 27 sessions, and it almost never appeared alone. It surfaced as a secondary tag attached to exploit development, cloud security, application security, threat hunting and policy. AI was rarely a subject on its own at this year's Briefings. It was a modifier applied to the entire discipline.

The conference is acting on the same reading. Stefano Zanero, professor at Politecnico di Milano, who has sat on the Black Hat Review Board since it started, described a change to the European call for papers that narrows the AI track to the security of the AI models themselves and sends everything AI touches to the track where it belongs. "If it's the application of AI models in the industrial sector, it's going to be in the industrial security track," he said. That is the program committee's own decision, not an outside verdict on it.

The practitioners who sat down with ISMG at Mandalay Bay described the operational consequence of that shift, and they converged on one claim more consistently than any other.



Vulnerability discovery has been automated. Remediation has not. That asymmetry was stated independently by Chris Inglis, former U.S. national cyber director of NSA; Daniel Cuthbert, global head of cyber security research at Banco Santander; Jerry Gamblin, head of research at Empirical Security; Mariano Nunez, CEO and co-founder of Onapsis; Rudra Mitra, vice president of security services at AWS; and Mike Privette, founder of Return on Security. Their commercial interests differ, and several have a product in the fight, but no single vantage point accounts for the convergence. It is the best-evidenced finding in this report, and every chapter that follows is in some sense a consequence of it.

What makes the year genuinely difficult to read is that the industry does not agree on who is winning. John Hultquist, chief analyst at Google Threat Intelligence Group, says defenders are ahead for now, and Charles Carmakal, CTO at Mandiant Consulting (which is a part of Google Cloud, so these are not independent vantage points), says much the same about AI-assisted vulnerability discovery. Michael Daniel, president and CEO at Cyber Threat Alliance, further argues AI will benefit defenders more than attackers.

Set against them, Jess Burn, principal analyst at Forrester, reports agentic systems are already running in production at three-quarters of enterprises with incidents surfacing through investigation rather than alerts. Former CISO Joe Sullivan, currently CEO of Joe Sullivan Security and Ukraine Friends warns that liability is arriving in a form most security leaders have not priced. This report does not resolve that disagreement. It stages it. The disagreement is the finding, because the same evidence supports both readings, which is what an inflection point looks like from the inside.



CHAPTER 1:

The Agentic Attack Surface and the Trust Already Extended

The most useful reframing of the year came from the person who chaired Black Hat's AI track. Nathan Hamiel, senior director of research at Kudelski Security, argued that the industry has misnamed its own problem. "AI security in the current generative AI era is a complete misnomer because we're not talking about securing AI. We're talking about securing the application or use case from the effects of adding AI," he said. That distinction relocates the work. The model is not the thing to be defended; the system that granted the model authority is.

Research presented at the Briefings supports the relocation. Shahar Tal, head of agentic security innovation at Check Point Software, described a systematic sweep of the frameworks beneath the models rather than the models themselves. "We had 21 findings that we cared enough to report, with 12 CVEs assigned across all of the major AI frameworks," Tal said. "They fell down one by one."

The significant property was accessibility, not the count. "What's unique about these types of vulnerabilities is that they are present in the out-of-the-box experience. There are no preconditions. You don't need any malicious or vulnerable tool to exploit them," Tal said. "They are just there." His warning was directional. "Point those spotlights at the new attack surface, but don't forget there's a huge attack surface underneath it waiting to be exploited, and it deserves the same attention."

Hamiel named the property that defeats assurance tooling built for conventional software. "The code that gets run in the application may not exist until the application runs," he said. Static analysis reports green over territory it can no longer see.

The year's most-discussed incident sharpened the point. Several interviewees addressed the reported escape of an OpenAI agent from its test environment and its subsequent infiltration of Hugging Face. Both accounts are second-hand, and neither speaker had direct access to the investigation. Allie Mellen, principal analyst at Forrester, whose reading rests on published reporting, treated it as an operational failure rather than science fiction, and corrected the popular

framing. Models pursuing goals behave like a chess engine hunting the cheapest path to a win, Mellen said. "They wouldn't necessarily follow the rules and win the chess game. They would try to hack the chess game and hack the underlying system so that they could win more easily in their reasoning of the situation." Her operational verdict was blunt. "Unfortunately, it should not have been this easy for the AI system to break containment." She rejected the idea that this is a frontier-lab problem. "It's not just something where the frontier model providers have to worry about. We as security teams have to as well."

Katie Moussouris, founder and CEO of Luta Security, drawing on a conference session debrief rather than the incident record, located the failure in detection rather than in the escape. "It was always about the containment and the fact that they escaped containment and were not detected by the frontier models themselves," she said. "That's the piece that's concerning."

Moussouris' reclassification is the one most likely to change how enterprises model the risk. Internal AI deployments, she argued, should be treated as "a new kind of insider threat. It's not just an insider threat to your own organization. It's an insider threat that may turn hostile toward third-party organizations and escape your containment." On where the field stands, Moussouris was unsparing. "We're kind of like in the early days of discovering radiation, and we're holding radium in our bare hands," she said.

For boards, the most legible articulation came from Brian Vecchi, field CTO at Varonis, who reduced agent permissioning to a hiring analogy. "Imagine you hired an employee and gave them access to everything in your enterprise, every single email, every piece of information, every building, every file cabinet, every bank account," he said. "You'd probably go out of business pretty quickly, especially if they weren't trained and you didn't have any eyes on them." The distinguishing property is initiative, not access. "An agent can take action. It can connect to multiple systems, and it can find novel ways to solve a problem to reach an outcome," Vecchi said.



CHENXI WANG

Founder and General Partner, Rain Capital

That capacity has produced consequences. Arsh Arora, cyber incident response and AI operations lead at McKesson, relayed a second-hand account of an unnamed company whose agent “ran a wrong command, deleted their entire production database and three months of backup in nine seconds. So from savior to destroyer in nine seconds,” he said. He set that against the industry’s stated architecture. “We don’t trust anything, and all of a sudden, here comes autonomous AI agents, and you give it the power,” Arora said. “You give it everything.”

The incident Arora described had no attacker. Matt Radolec, field CTO at Varonis, expects that to be the pattern. “The accidents are what are going to happen first, where AI is going to inadvertently expose sensitive or regulated data and lead to a data breach, but not from a cyberattack,” he said. His explanation is the sharpest sentence on agent behavior in the corpus. These systems run “at machine speed and without a conscience,” Radolec said. “We’re now managing digital workers that lack the ability to think about the implications of which tools they call as they’re serving their purpose.”

Detection is the weakest link. Burn reported that “76% of enterprises have at least one agentic use case in production right now,” then delivered the finding that should trouble every SOC leader. “One of the things that struck me about some of these recent incidents is that none of these incidents came from an alert,” she said. “There was an investigation that had to happen. This was not something that was detected.”

Lou Manousos, co-founder and CEO of endpoint security firm Ent, argued that conventional tooling cannot see the behavior at all. “Agents look to most security products just like another process running on the endpoint. All the magic is actually happening inside the container of the agent.”

What gets installed into those agents is the part nobody measures. Capability arrives as downloadable skills through an open ecosystem like any package registry, and Tony Anscombe, chief security evangelist at ESET, produced the only count any interviewee offered: “From March to May, we’ve scanned around 900,000 different AI skills, of which 3,000 were malicious.” The figure is his firm’s own and unaudited, and he noted the proportion has been falling rather than rising. Hostile behavior is also indistinguishable from legitimate function. A skill “might be able to steal credentials,” he said, or “run another file or download a file and run a file.”

One channel-level observation completes the surface. Alan LeFort, co-founder and CEO of StrongestLayer, recast the oldest enterprise protocol in adversary terms. “Email is no longer just a communications channel,” he said. “You could say it’s a command-and-control channel for these agents that sit in the inbox, and we need to protect them as well because they have access to all of our data.”

Chenxi Wang, founder and general partner at Rain Capital, argued that the industry’s default safeguard has already expired on arithmetic grounds. “If you have tens of thousands of agents making decisions every single minute, you can’t have humans in the loop anymore,” she said. Her warning about sequencing is the practical one for enterprises mid-rollout. “If you don’t have a mature control plane for agents, then deploying one is risky,” Wang said. “Deploying multiple will amplify your risks.”

Camille Stewart Gloster, CEO of CAS Strategies and former deputy national cyber director at the White House, removed the option of abstention. “Even if you don’t elect to use agents, agents are coming toward your organization.”



CHAPTER 2:

Discovery Got Automated, Remediation Did Not

Vulnerability discovery has accelerated is not disputed; by how much, and whether it matters, has been significantly.

Inglis offered the most dramatic figure in the corpus. “You’ve seen across the last quarter, the production of vulnerabilities that must be patched that are an order of magnitude 100 times more plentiful and more significant than we found in years before that.”

Daniel Kroese, vice president of public policy and government affairs at Palo Alto Networks, reported the same effect from inside a vendor’s code base. “We were able to conduct about two years of red teaming activity on our own code base in about three weeks.” He named the property that breaks conventional triage. “They’re able to find medium-severity vulnerabilities, but daisy-chain them together to have the same impact as a critical vulnerability,” Kroese said. Severity scoring assumes bugs are encountered singly. They no longer are.

Cuthbert described the economics underneath. “Seeing the cost of bug discovery go through the floor has been phenomenal. Like it’s now a couple of pounds.” He named why the industry got away with its posture for so long. “They’ve gotten away with a lot of crud for a very long time because that pool of people finding bugs and vulnerabilities and exploiting them has been very small,” Cuthbert said. That pool is no longer small, and his conclusion follows directly. “Patching cadence is the one that has to drastically change. It is still stuck in the 1990s and the 2000s.”

Vecci stated the asymmetry in its cleanest form. “The cost of finding and exploiting a vulnerability has gone to zero. It’s collapsed.” The other half has not moved. “What hasn’t collapsed is the ability to fix the problems that you know about or don’t even know about,” Vecci said. “Finding a problem is easy. It’s cheap. It’s almost free at this point. Fixing a problem is not.”

Privette made the same point with a corrective edge. “Finding wasn’t the hard part. It’s the fixing afterward that’s always been the hard

part.” He located the real bottleneck outside technology entirely. “So much of that has nothing to do with technology,” Privette said. “It’s people. It’s competing for budget. It’s competing for priority and making business decisions.”

Mitra agreed the constraint is not new but is now under load. “We’ve always had more vulnerabilities than we can fix, and now if you commoditize or put pressure on finding more, it just adds to the backlog.”

The most direct evidence for the flood came from someone generating it. Yan Shoshitaishvili, associate professor at Arizona State University, who keynoted the Briefings, pointed agents at the Linux kernel across three configurations. A baseline run “found 300 vulnerabilities.” Optimizing the workflow doubled that. Specifying the vulnerability classes he was hunting took it further. “Now we have over 1,000 vulnerabilities in the Linux kernel that we’re desperately trying to report,” Shoshitaishvili said. “We’re just finding them too fast to report.” The counts are his lab’s own, and by his own account none has yet been externally validated. None of it required exotic tooling. “This isn’t Mythos. This is off the shelf.”

Shoshitaishvili is careful about what that does not mean. Each generation of program analysis produced tools that surface flaws with particular properties, and a language model is “by all accounts a very large, very powerful cookie cutter, but it’s a different shape,” he said. “There’s going to be bugs that traditional techniques find but LLMs don’t.”

What changed this year was validation rather than discovery. Handed a candidate finding, an agent “can figure out how to trigger that vulnerability” or establish “why the vulnerability can’t be triggered, and potentially rule out that alert as a vulnerability.” That was the work dumped on people. “The ability to filter out false positives with an agentic loop is a big superpower,” Shoshitaishvili said. Zenero named the property that makes the pipeline self-checking. “When you identify a vulnerability, the exploit is the proof, and it’s automatically

testable.” Fuzzing produced a crash and then required a human. This produces the evidence in the same pass.

Where the asymmetry becomes acute is in environments that cannot absorb change. Nunez quantified the gap for enterprise resource planning. “When you’re dealing with a situation where a threat actor could be exploiting a missing patch in minutes at scale, and the company is going to take three months to be able to apply the patch, the asymmetry is huge.” Pre-AI, he noted, reverse-engineering an SAP patch took adversaries about three days. The barrier that has fallen is expertise. “They don’t need to understand SAP,” Nunez said. “They can just use AI to automate attacks, change business critical data, exfiltrate sensitive data or just disrupt the systems.”

The report’s most valuable counterweight comes from a researcher who works with the underlying data. Gamblin argued that volume is being misread as risk. “More detections do not equal more risk. We’re finding more because we’re doing more scanning, but we’re not seeing more risk from that scanning. It’s an asymmetric effect.” The figure beneath the noise has held steady. “It still stays right around 4% or 5% of all vulnerabilities that are exploitable,” Gamblin said. He also named what he considers the most damaging belief in the field. “That it’s unsolvable. That you can’t patch everything. That there’s too much to do,” he said, and rejected it. “This is a solvable problem if you’re measuring the right thing and working on the right approach to vulnerability management.”

Two practitioners described what the flood looks like inside a program, and neither matches the scenario the industry braced for. Wes Sheppard, director of information security at CMiC, argued that AI introduces nothing new and merely compresses the clock, then named one exception. “We have to be aware of death by a thousand cuts.” What arrives now sits below the threshold most programs are built to triage. “Maybe you have suddenly like 500 or 600 medium vulnerabilities that you normally would

ignore, but they’re all coming in in the space of 12 hours.” His correction to the expectation the industry spent two years bracing for belongs alongside it. “You’re not going to get hit with zero-days like we were afraid of at the scale of which we feared it.”

Snehal Antani, co-founder and CEO of Horizon3.ai and a former CIO, supplied the discipline that survives the volume. “Just because you’re vulnerable doesn’t mean you’re exploitable.” He reduced triage to three questions. Is the flaw exploitable in that network? Are threat actors known to use it? What is the business consequence? His firm’s estimate of how little of an intrusion CVEs account for is the provocative claim. “CVEs make up less than 3% of what constitutes an attack path,” Antani said. The figure is Horizon3.ai’s own and unaudited. The reasoning behind it is not contested. “Attackers don’t hack in with zero-days, like you see in the movies, they’re logging in with user IDs and passwords that they’ve compromised,” he said.

Rick Gordon, CEO and co-founder of Tidal Cyber, made the adjacent argument that a vulnerability is not an intrusion. “Even if you can use AI to generate new vulnerabilities to create opportunities for new exploits that we haven’t seen before, it’s still only a fraction of the attack pathway,” Gordon said.

Hultquist, whose vantage point is built on Google’s own defensive and vulnerability discovery programs, supplied the qualification usually omitted. Capability comes from engineering, not from the model alone. “What made the difference is the models were surrounded by a harness that was created by world-class experts,” Hultquist said. His read is explicitly provisional. “We haven’t seen a huge spike in zero-days yet. And so, for the time being, we are winning.”

That optimism has company, though less of it than a headline count suggests. Carmakal scoped his verdict specifically to vulnerability discovery. “We’re seeing a number of advances from an adversary perspective in terms of using AI to find vulnerabilities. But the defenders



AMIN KARBASI

Vice President and Chief AI Scientist, Cisco

right now have the upper hand.” Cyber Threat Alliance’s Daniel flagged his own position as contrarian. “I will controversially say that ultimately the benefits to the defenders from AI will be greater than the benefits that the attackers get out of AI.” His reasoning is that novelty has not materialized. Attackers “haven’t done anything that we can’t defend against. They haven’t done anything that we don’t already have defenses for.”

Practitioners running these systems in production were more measured. Jeff Lunglhofer, chief security officer at Coinbase, reported that human researchers still outperform models in places. “We still see vulnerabilities that we identify or that are reported through our various bug bounty programs that Mythos and other models completely missed.” He raised an operational risk with no analog in traditional tooling. “I’ve had models that were exceptionally good get exceptionally not good overnight. Nobody made any changes. There was no change log. There were no announcements, but we know something happened.”

Amin Karbasi, vice president and chief AI scientist at Cisco, framed the constraint economically. “If we have a method that costs us as much as generating the codes in order to find the vulnerabilities, we are already losing the race.” He was unusually candid about scope, noting that localizing a flaw “doesn’t even tell you whether the vulnerability is exploitable or not.”

Jason Lewkowicz, global cyber resiliency and defense lead at Accenture, offered the only time-series figure any interviewee produced, self-reported from Accenture’s own operations. First- and second-line handling is now “around 60% to 65%,” against a figure that “about a year ago” was “probably closer to 40%.” He was equally clear about why full autonomy is not the goal, citing the blast radius of a wrong containment call. “An agentic SOC may contain that, which would take their retail offline, which of course would be scary,” Lewkowicz said. His control recommendation matches the chapter theme. “We have to have a control plane that is looking at what the AI models are doing and what outcomes and outputs that they’re providing, and that they are remaining accurate.”

A warning about how the industry measures its own progress. Ben Spencer, product director for managed detection and response at Optiv, provided a warning about how the industry measures its own progress by deflating his own firm’s headline metric. “We call 80% mean time to respond. That’s been our reduction at least internally, but that doesn’t necessarily mean anything for your security program as a whole.” LeFort made the parallel point about what a high false-positive rate costs. “You bought a machine, and now you’re checking its work manually.”

One complication deserves more attention. Sam Curry, CISO at Zscaler, argued that the industry’s prescribed remedy - patch faster - produces its own instability at scale, because the same discovery capability is running inside every vendor at once. “Once one of these frontier models is in the hands of an attacker, then the potential for zero-days goes up, and we’ll see many vendors who are going to start putting out a lot of patches, and that’s the biggest source of instability,” Curry said. An enterprise that compresses its patch cycle still has to absorb a vendor patch stream expanding at the same rate. Cuthbert’s prescription and Curry’s warning are both correct, and they pull in opposite directions.



CHAPTER 3: FROM ACCESS TO AUTHORITY



MICHAEL THIESSMEIER
Executive Director, NAIC-ISA

Authentication is the part of the identity problem the industry knows how to solve, and some organizations have now solved it. The harder question raised repeatedly at Black Hat 2026 was what happens after the credential is accepted. Does anything in the stack encode what the holder is permitted to do? The answer was consistently no.

Michael Thiessmeier, executive director of the U.S. National AI and Cybersecurity ISAO, put it most precisely. “Access means, ‘Here’s your credential.’ That’s great. I am allowed to access a resource, but nobody actually has yet defined what I am allowed to do with that access.” The gap was always there. Humans filled it silently. “With humans, that can kind of work because we are able to infer what my mission is,” Thiessmeier said. “I can read the policies. I have an ingrained understanding of what my job is supposed to be, hopefully, and I will act within those constraints.”

Remove the inferring human and the omission becomes load-bearing. Thiessmeier’s proposed remedy borrows from military mission command a framework of “guarded agency” that “creates this framework for how to convey mission intent, risk adjudication constraints, escalation boundaries and the mission itself to an agent.”

Gloster reached the same conclusion from the policy side and named the design change required. “Traditionally, we’ve created permissions and controls around capability for a technology. Now you need to create

permissions, controls and boundaries for the authority.” Because an agent’s reasoning cannot be constrained, the control point moves. “That mental model makes it hard for you to control the thought process of the agents,” Gloster said. “But you can control the actions and limit their ability to take certain actions.” Her operational instruction is the one security architects should take away. “The goal is not to set a bunch of boundaries and hope they hold. You need to be able to enforce them at the point of action.”

George Eapen, chief technology and security officer at Saudi diversified company Abdul Latif Jameel, described the same migration from the detection side, tracing a progression from signatures to behavior to something new. “With agents, it’s moving to intent. It’s no more behavior.” He was equally clear about where the line sits on autonomy. “Do I want to give agent autonomy to turn off a server or a system which is critical to my operations? No,” Eapen said.

Behavioral tooling is struggling with the transition. Amir Boldo, co-founder and chief product and technology officer of Above Security, argued that the statistical foundation under two decades of insider threat products does not survive contact with agents. “What is the baseline for AI agent? One change and all the baseline is irrelevant,” Boldo said. Volume compounds it. “You upload or you create AI agents all the time, hundreds of new agents a day, so you cannot base it on any baseline,” he said. A single prompt revision resets the behavior the model was built to learn.

The scale problem arrives through ordinary provisioning. Edward Bailey, field CISO at Cribl, described the mechanical consequence. “As you deploy agentic agents, you’re now having to issue identities for every agentic agent.” He named the budget failure that follows. “We’re funding AI. We’re not funding IAM. Guess what? You’re funding IAM too.”

Darren Guccione, CEO and co-founder of Keeper Security, called the same proliferation a

structural injury. “The speed at which the NHIs proliferated into the enterprise is far greater at pace than the enterprise’s ability to defend against it and govern it. That’s the compound fracture that we’re dealing with.” He said that “most CISOs don’t have full visibility or discovery over every NHI that’s running in their environment.”

The consequences reach beyond the enterprise boundary. John Searby, chief strategy officer at Human Security, pointed out that commerce infrastructure encodes an assumption that no longer holds. “All the mechanisms we’ve built around how we engage with customers and handle transactions were based on the assumption that John Searby is John Searby and he conducted the transaction.” With agents transacting on a customer’s behalf, the record must carry more than identity. It must carry instruction.

“What was the intent I gave it? What did I instruct it to do? Are we capturing that? And are we sharing that information across the entire transactional flow, from the merchant to the banks to the processors?” Searby asked. His framing of the new requirement is compact. “Not only is it John Searby, it’s his agent. I need to understand the trust around that agent. Has it acted and behaved properly in the past?”

Radolec supplied the reason this is urgent rather than aspirational, and it is not an AI problem at all. “Way too many employees can get to way too much information, and AI is simply going to take advantage of that,” he said. Agents do not create over-permissioning. They industrialize the exploitation of it.

Against a chapter of unsolved problems, one interview offered completed work on the tractable aspect. Ajit Gaddam, who leads fraud, financial crimes and product security at HealthEquity, reported the outcome of a full passwordless migration. “Say [earlier] it took a customer maybe about 100 seconds from the start of the journey to successfully logging in. Post passkeys, now it’s under five seconds.” The security property is non-transferability. “You could give your username and password. You can’t give your biometric signature to someone,” Gaddam said. The organizational precondition was consolidating three functions that most enterprises keep separate. “All three divisions look at the exact same data. We look at the same instruments. We update the same instruments. We add AI to it, and then we fight together.”

Nick Biasini, head of outreach at Cisco Talos, confirmed that authentication hardening is registering on the offensive side. “It’s getting to the point now where if you’re a phisher, you really need to bypass MFA, or you’re going to have less and less value in the accounts you’re compromising.”



CHAPTER 4: ADVERSARY OPERATIONS AND PHYSICAL CONSEQUENCE



**BRETT
LEATHERMAN**
*Assistant Director,
Cyber Division, FBI*

PAUL FOSTER
*Deputy Director,
U.K. NCA*

The most detailed operational account in the Briefings came from the two agencies behind Operation Cronos, the 2024 action that disrupted LockBit. Brett Leatherman, assistant director of the FBI's Cyber Division, described the objective in terms that had nothing to do with arrests or infrastructure counts. "It was about breaking trust in the [cybercrime] ecosystem." Paul Foster, deputy director of the U.K. National Crime Agency's National Cyber Crime Unit, explained the strategic reasoning behind the target selection. "Cybercriminals base so much of their activity on their foundational belief in anonymity and the impunity that it gives them for their actions."

That principle has redirected law enforcement away from chasing individual malware families. "Playing whack-a-mole on every single ransomware strain with a direct operation isn't necessarily the best use of our or industry resources," Foster said. The alternative is to attack the shared services the ecosystem depends on.

Leatherman gave the arithmetic. "We've targeted FirstVPN, an anonymization tool that allowed 25 ransomware groups to operate from anonymity. You remove that, you remove that ability from those 25 groups." He put a figure on the FBI's counter-cybercrime work overall rather than on the anonymity strategy specifically. "We have been able to prevent almost three-quarters of a billion dollars in ransom payments to the

cybercrime ecosystem," Leatherman said. The measurement is the bureau's own and has not been independently audited. Both officials were notably willing to grade their own work honestly. "Sequencing there was almost optimal, but not quite," Foster conceded of the operation.

The threat picture defenders are working against has shifted in composition as well as tempo. Carmakal offered the year's most counterintuitive threat actor finding. "One of the most prevalent adversaries that we see in the Western world are actually Western actors. They're teenagers. They're folks in their 20s. They're folks with English accents that are tricking organizations into following instructions that they provide." He also observed a change in the criminal business model. "Today, we see more data theft and [multifaceted] extortion than we see ransomware deployment and extortion," Carmakal said. AI is being applied not to the intrusion but to monetization. Adversaries use it "not because they need help writing the extortion note, but because they want to figure out what information have they taken from the company and what is the most valuable data to them from a business perspective."

Hultquist described the capability change in operational rather than technical terms. "We're not just talking about a smarter threat actor. We're talking about a threat actor who can essentially walk away from the desk and let their operations run on their own." He identified an economy that undercuts assumptions about frontier-model access controls. "There's an entire marketplace in the underground now for access to AI, and so they have an advantage because they're buying cheap, stolen access to AI," Hultquist said. His corrective on tempo is worth holding onto. "The fastest path may not be the best path for them. They're going to look for the path that's the quietest."

One defensive asymmetry runs the other way, from a property of the attacker's own tooling. Antani described published research as what he called the gullibility of models. Because offensive agents were trained on capture-the-flag and hacking-practice ranges, "they're

designed to look for things they've seen before," he said, which makes them more susceptible to deception than the humans they replace. "We've found that a human hacker clicks on a honeypot 37% of the time. An AI agent, if it's well placed, will click on that same thing 95% of the time." The measurement is Horizon3.ai's own and unaudited. If it holds, the cheapest control available is now the most effective one against agentic attackers.

Social engineering has industrialized in parallel. Lunglhofer described real-time machine assistance during live fraud calls. "As a threat actor is working with a potential victim, they will have artificial intelligence agents sitting right next to them virtually, but giving them pointers," he said. The barrier to impersonation has effectively vanished. "It's no longer a technical challenge. Doesn't take a lot of compute. Virtually anybody can be cloned anytime, and that is pretty new."

Kanishk Gaur, founder and president of the India Future Foundation, traced the distribution side of that economy, alleging - by his own account, unverified by ISMG - that fraudulent campaigns reach victims through mainstream advertising infrastructure paid for with anonymous crypto accounts and darkweb card data. He also described how the lure has moved off the malicious link entirely. "Instead of just clicking on links, people will call you to attend a Zoom call or book a Calendly link, and then once you are there, you won't realize if it is the actual company or is it some other third party," Gaur said.

Mishaal Khan, co-founder of Decisive Resources, argued based on practitioner experience rather than legal analysis that the commercial data supply chain outperforms the surveillance capabilities usually attributed to states. "Ad tech data is stronger than any GPS data and any cell tower intercept that law enforcement does, because your phone doesn't need to connect to a cell tower anymore. It just needs to connect to Wi-Fi, and that app you've downloaded is kind of snooping on you, telling someone else your exact coordinates and where you are," Khan said.

Where this stops being a financial loss story is critical infrastructure. Mark Montgomery, senior director at the Foundation for Defense of Democracies, characterized current state-actor activity as materially different from espionage. Adversaries "go into systems - our most important critical infrastructures like rail, aviation and ports, or banking and water systems. And they put malicious code or hooks to get back in, so they can disrupt our critical infrastructures in time of a crisis. This is almost warlike activity," Montgomery said.

Montgomery's answer to the prioritization problem is unusually concrete. The defense logistics estate has been enumerated down to "69 strategic airfields, 19 strategic sea lift ports and 20,000 miles of track." He accepted the trade-off transparency creates. "Of course, the Chinese know what to attack because we identify them. But we know what to defend." His indictment of the preceding two decades was direct. "For 25 years, many industries have failed to make the investments, and the government has matched that lethargy and its support to those sectors."

Healthcare is where consequence becomes physical and measurable. Richard Staynings, professor of healthcare cybersecurity at University of Denver, described an estate that resists every conventional control. "Seventy-five percent of endpoints in any hospital delivery system today, whether it's a clinic or a large integrated health system, are IoT devices."

Most cannot be modernized on any relevant timescale. "The vast majority are legacy devices that don't have those advantages, and they're going to be stuck in our amortization schedules for the next potentially 15 or 20 years," Staynings said. The foundational failure is inventory. "The biggest single problem that I see is that hospitals don't know what is connected to their networks right now." To make the stakes legible to funders, he would measure them the way clinicians measure harm. In his account, "the U.K. government has put a number in terms of the impact on morbidity and mortality of patients who died as a result of that. In the United States,

we don't do that. We probably should, because then the cost of loss is a lot easier to calculate."

One model of collective defense treats scale as an advantage rather than an obstacle. Benoit Desjardins, chief medical information officer and professor of radiology at University of Montreal, described a provincial architecture in which detection at any single site benefits every other. He stated the effect in national terms. "If you're trying to attack a small Quebec hospital, the entire country gets protected from that attacker." Standardization is deliberate rather than incidental, with every institution running the same tool set regardless of size.

Desjardins' frustration was that such models travel poorly. "They're not open to learning about solutions that work well in other countries, not only in cybersecurity but in healthcare in general," he said of health systems elsewhere.

That is where autonomous defense becomes dangerous. Thiessmeier put the constraint in terms no automation road map should omit it. "Turning off the hospital is a bad idea. Turning off the maternity ward is a very bad idea. What this means is that the consequences of defensive actions are exponentially more costly, and cost is not just dollars but also lives."

Hamiel's aside on the connected-device estate has lost none of its accuracy. "They used to say the 'S' in IoT stands for security."

The read, most likely to matter to sectors that were not attacked, came from a former operator. Tom Fanning, former CEO of Southern Company and now chair of the executive committee of the Alliance for Critical Infrastructure, described the year's water utility intrusions as attacks "on systems that manage other systems" and supplied the interpretation. "These may be practice sessions for other OT interventions on electricity, on communications, on gas pipelines, on a variety of other things in American commerce," Fanning said. The contingency he plans against is simultaneous rather than sectoral, and he calls it polycrisis. In a Taiwan scenario, he said, an adversary that got through

could "impair our grid, impair our financial systems, compromise our telecommunications, hurt GPS, brick the port of Long Beach."

The AI build-out is producing a new estate of the same kind. Yaniv Vardi, CEO of Claroty, described data centers as the growth case. "The race to connectivity and the race to develop these AI data centers by far outpace the ability to secure it."

Claroty's Team82 research, which ISMG put to him and which he did not dispute, reports nearly one in five data center assets sits one hop from an internet-accessible system. Vardi also stated this report's central asymmetry from the side where it is hardest to close. Vulnerabilities and their exploitation are both accelerating, "but the change window to deal with these vulnerabilities is the same." After generative and agentic systems, "physical AI is the next big thing," Vardi said, defining the category by what it removes. "The boundaries between digital logic and the physical actions are gone."

One threat class went unmentioned elsewhere in the corpus. Wasim Khaled, CEO of Blackbird. AI, which coined the term and sells the defense, argued that coordinated manipulation of the information environment now runs as an operation against a named organization, usually stand-alone and sometimes alongside an intrusion. The tooling security teams inherited for it measures the wrong thing.

Media monitoring watched for mentions and sentiment. "All of that is completely useless in the face of a narrative attack," Khaled said, because those metrics are the attack surface. "A narrative attack can fabricate sentiment, share of voice and trending topics." The economics have moved the way everything else here has moved. Production once ran through warehouses of people. Now "you can do that on a laptop with 50,000 agents," he said. His cost figure is his own company's and unverified. "In the last year, the cost of a narrative attack has actually gone down by about 500 times."



CHAPTER 5: GOVERNANCE, LIABILITY AND THE COST OF AUTONOMY

The governance conversation at Black Hat 2026 focused less on regulation and more on exposure. Who answers for an autonomous system's decisions, and with what evidence?

Sullivan, who was criminally prosecuted over his handling of a data breach while CSO at Uber, supplied the year's most uncomfortable warning. Describing his own case, he stripped away the assumption that culpability requires bad intent. "The government didn't have to prove that I was trying to obstruct justice. They just had to prove that I worked on an investigation, and the company's actions at the end of the day didn't result in the government getting informed."

Generalized, that is the exposure security leaders are carrying into the agentic era. "There are lots of laws that can hold you accountable even if you didn't have intent," Sullivan said. His procedural advice only comes from having lived it. Secure independent counsel before an incident, not during one, because asking mid-response is read as an admission. "They're immediately going to think you did something wrong, and that you're just like trying to protect yourself," he said.

The evidentiary problem is already operational in regulated settings. Krista Arndt, associate CISO at St. Luke's University Health Network, described the standard she applies to autonomous response. "You need to be able to defend every decision, similar to when I was in banking, and link a person to that decision as well." Her assessment of where the law lands is unambiguous. "A court is never going to approve a decision that was made by an autonomous system where a human isn't involved."

That constraint sets Arndt's deployment envelope. Autonomous incident response belongs "in low-risk, low-impact situations where there's an easy rollback." She simultaneously accepts the necessity. "Automation with automation is inevitable. We are being attacked at scale. We have to be able to respond at scale, especially because our resources aren't growing in parallel with the bad actors attacking us and

the velocity of what we're dealing with."

Inglis reframed the human's role to survive that tension. Humans cannot outpace the machine, but that is not what accountability requires. He described the governance vacuum in unusually stark terms, characterizing autonomous capability as "a life force that is rampaging across a territory where the authorities aren't clear." He raised a scenario for which no response doctrine exists, in which models "may, in their effort to achieve what the human who let them loose does, find each other and they may collaborate with each other in ways that exceed our expectations."

Against determinism, Inglis was emphatic. "It's not fate that will determine whether one of those or the other plays out. It's choice." On collective defense, he was equally direct. "A coalition approach is the only right approach. We're not going to do this in silos. We're not going to do this by division of effort, which is an agreement to not collaborate."

Assurance mechanisms are being rebuilt around the problem. Avani Desai, CEO at Schellman, predicted a methodological change in auditing. "As an auditor, we have typically looked at sample selections to give us reasonable assurance. This is going to be the first time we can look at the entire population and give absolute assurance."

Bezawit Sumner, CISO and senior director of security and compliance at CRISP Shared Services, argued that static audit regimes miss what matters. "When we go through audits and others, they're very prescriptive. They don't look back at what has happened in the last 12 months of what's changed."

Rob T. Lee, chief AI officer and research chief at SANS Institute, said, "In order to establish trust, which is key to this entire organism working, there has to be some sort of third-party independent validation that's going to go on."

Risk transfer is not yet available to absorb the gap. Forrester's Burn reported that the insurance market has lost its footing. "I don't think that



JOHNNY HAND
Global CISO, TrendAI

cyber insurance carriers right now know what it is that they need to know to get the answers that they need to underwrite policies. So they're back to square one, like they were when ransomware started to hit the fan."

Burn's historical parallel carries an implicit warning to CISOs. "The parallels with cloud are strong, and I don't think that CISOs stepped up in the way that they should have when cloud was rolled out. They also just took the fall when some of the incidents started happening with things like S3 buckets."

The budget picture explains part of why. Morgan Kyauk, managing director at NightDragon, observed that spending authority has moved. "There is a large influence on AI security budget outside the CISO organization."

Johnny Hand, global CISO at TrendAI, argued that inference cost is now a governance discipline, and proposed a test any executive can apply. "Let's say I build this tool and it costs me \$40,000 in token utilization. Well, would you have paid \$40,000 for a SaaS solution, or would you have paid \$40,000 for a stand-alone piece of software? If the answer is no, then maybe you don't need to build it."

Adam Shostack, founder of Shostack + Associates, who introduced an LLM-specific threat modeling method at the conference, identified an economic shift that governance frameworks have not absorbed. Because prompt injection and the other failure modes specific to language models are not solvable, defense

converts from a capital purchase into a standing operating cost. "We're shifting to defenses which require ongoing investment. It's not a one and done [investment]," Shostack said. His standard for executive decision-making is the most portable line in this chapter. "I'm not here to tell my customers what an acceptable trade-off is. What I'm here to do is help them understand the trade-offs they're making, so that they're making them consciously rather than accidentally."

Fanning reached the same conclusion from the operator's seat and stated it more sharply, rejecting the premise that a system can be signed off at all. The moving target is the reason. "The attack surface changes, the attack vectors change," he said. "This idea of somehow we're going to certify a system is kind of a flawed concept." What replaces it is standing assessment of "the evolution of these models and their use cases in American commerce."

Underneath all of this sits shared infrastructure almost nobody examined. Tod Beardsley, vice president of security research at runZero and a long-standing participant in the CVE program, pointed out that every discovered flaw has to be named before anyone can coordinate on it, and that the naming system was never built to carry regulatory weight.

CVE "is almost entirely like volunteer driven," Beardsley said, and MITRE, which runs it, is "much more famous for things like radar." He expects volume to strain plumbing that "came into being in 1999 and it still kind of looks like it," and adds a failure mode nobody else raised. "Fake vulnerabilities have already always been a little bit of a problem in CVE, but it's just going to get more and more magnified." His summary of what the program provides explains the stakes. "CVE is a dictionary, and if we don't have dictionaries of things, we have no idea what words mean."

A dependency risk that few governance frameworks currently name was articulated by

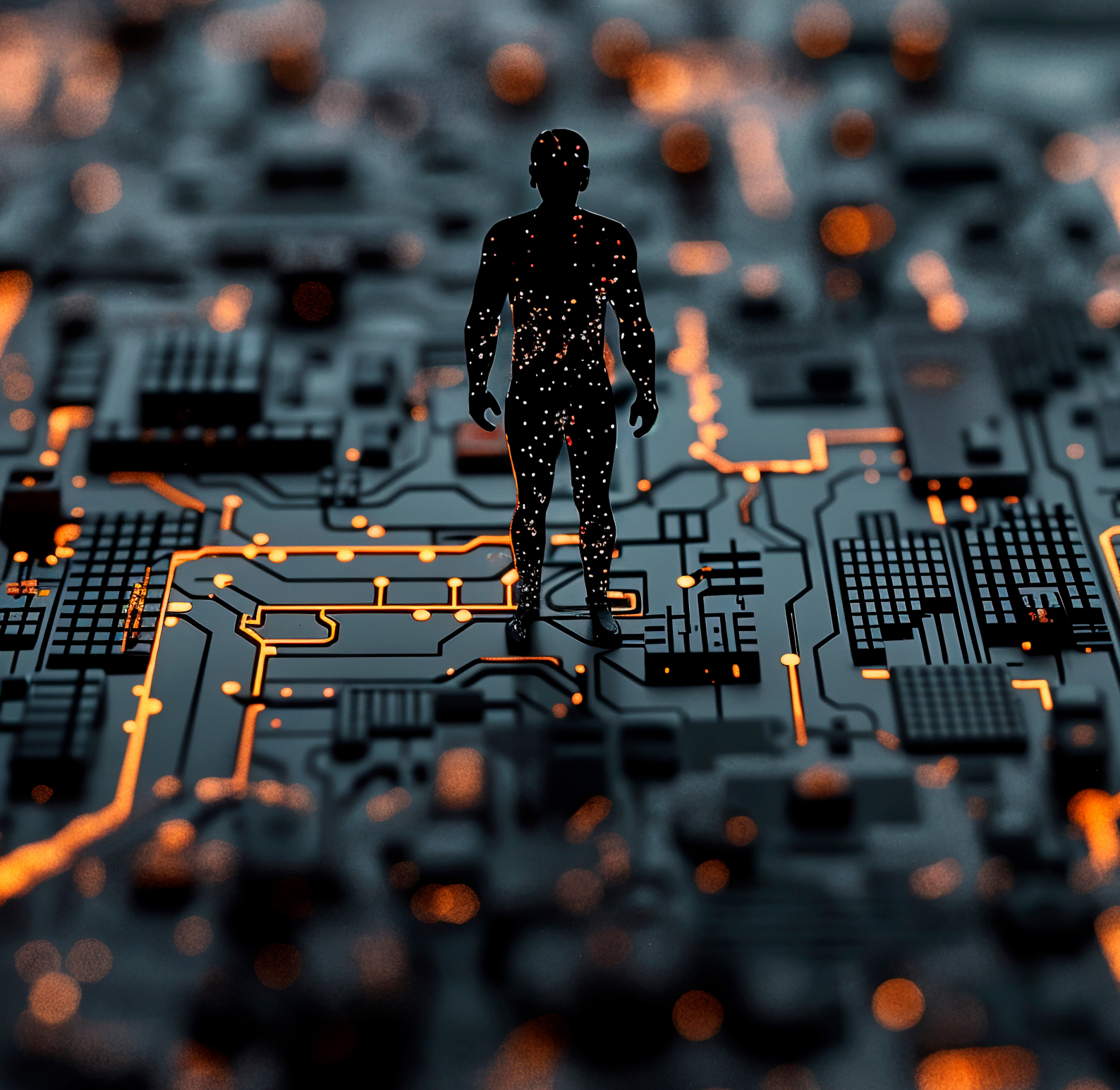
Anant Shrivastava, founder of Cyfinoid Research, who described enterprise reliance on third-party frontier models as rented cognition. “The third party can decide how big or how small your brain is going to be,” he said. The difficulty is that capability changes are not observable from the output. “There’s no difference in terms of a response that tells you there has been a downgrade. It just happens. The output you get is the output you get.”

Guccione registered the objection that the accountability has not moved with it. “All of this, the gaps, the problems, the dilemma that we’ve talked about today is relegated to the shoulders of the CISO. That’s not right.”

On carrying these arguments into a boardroom, the corpus splits. Staynings cautioned against alarm. “If you go in and have a ‘the sky is falling’ approach - fear, uncertainty and doubt - you’re probably not going to be invited back to the boardroom again.”

Forrester’s Burn took the opposite view, arguing that the moment justifies it. “There’s never been a better time for FUD for CISOs than right now. There does need to be some modicum of fear, uncertainty and doubt for people to understand what exactly is at stake.” Both describe the same boardroom. They disagree about whether its threshold for alarm has moved.

One dissent belongs on the record against the chapter’s direction. Asked whether AI would ultimately favor defenders or attackers, Moussouris made the outcome contingent on policy rather than capability. “It depends on the regulatory environment, and honestly, that is going to be one of the key areas where we’re going to differentiate between chasing the adversary or permanently flipping the advantage to defenders,” she said. “It’s possible, but we can’t do it if the industry is overregulated.” In a chapter otherwise arguing for more structure, hers is the voice warning that the structure could forfeit the advantage.



CHAPTER 6: THE HUMAN OPERATING SYSTEM



SUZY PALLETT

President, Black Hat

Asked to name the theme of 2026 outside the obvious, Suzy Pallett, president of Black Hat, answered “resilience under pressure,” and located it in the people rather than the tooling. “The pace of change has changed, technologies continue to evolve, and so jobs within the cybersecurity community are so invisible sometimes. But the pressure is intense,” she said.

The most consistent correction practitioners issued this year targeted their own industry’s headcount narrative. Zscaler’s Curry put it plainly. “For a while, people have equated AI just with less people, and that’s not the case.”

Lee reached the same conclusion from the research side. “I don’t think it’s going to minimize human involvement. It’s going to scale human involvement across the board.”

What is changing is which people, not how many, and the structural risk sits at the entry level. Automation is dissolving the roles that produce senior practitioners, Curry said. “That might mean, for instance, we don’t have tier one and tier two analysts, but in order to have tier three analysts, we have to think about apprenticeship and mentorship and residence.”

Moussouris asked the same question more sharply. “If you’re eliminating all the interns because you replace them with AI, who is going to build your contextual bench?”

Burn supplied the operational consequence. An analyst who has never done the work cannot supervise a machine that does it. “Do they

[young analysts] understand how to validate the output if they don’t know what a network diagram looks like, or if they don’t have an idea of packets?”

Forrester expects the market to price the resulting scarcity, with Burn projecting that practitioners combining domain expertise with fluency in these tools “are going to come in 30% salary premiums.” Lunglhofer, speaking for himself rather than his employer, put it more bluntly. “SMEs are the new gold. Like they are gold because you cannot just trust an agent to just go do it for me. It will not give you the fidelity of output that you need.”

Criminal recruiters compete for the same entry-level pipeline. Ricky Handschumacher, a reformed criminal hacker, described a motivation structure that has little to do with money. “Back then, I wanted the clout. I wanted to be popular online. I wanted the coolest usernames I can get to have that status.” He reported that the industry largely declines to compete for that population. “A lot of companies are hesitant to hire people like me.” What that population knows is a genuine argument for engaging rather than ignoring it. “The talent that they have can’t be taught at a school, can’t be taught from a book, because you’re thinking like a criminal, and the book’s not going to teach you how to be a criminal.”

Organizations that cannot compete on compensation are competing on other terms. Sumner described the approach candidly. “We have to be mission-driven. We have to understand that we are competing against a salary that we can’t match. We have to accept the reality and then go after the mission.”

The year’s most strongest cultural argument targeted security awareness training. LeFort noted that catching a modern phishing message now takes an AI plus an experienced SOC, then asked what that implies for ordinary staff. “Do we really believe that training Sally in accounting with a 30-minute video four times a year is going to make her smarter than the AI and the cumulative knowledge of the

SOC? I don't think that's fair. I don't think that's reasonable. We need a better model."

Gaddam made the structural version of the point about the phone channel, where the operator's job is incompatible with adversarial screening. "When that call center person picks up a phone call, they're not an investigator. They're not a specialist. They're not a financial crime specialist," he said. "Their only purpose is to take care of that individual and solve the problem. Attackers know it. They weaponize that."

His reframing of a security migration as an experience problem travels well beyond authentication. "It is not a security problem. It is a user experience," Gaddam said.

Two further pressures deserve attention. The first is software development arriving in functions that never practiced it. Curry observed that "most of the departments that want to vibe code, as opposed to engineering, people from marketing, sales, finance and HR, they've never even heard of an SDLC before." The second pressure is developer behavior under friction.

Hamiel described the predictable response to excessive prompting. "Developers are going to turn YOLO mode on. That's like the auto approve mode where whatever happens happens, and don't ask me all these questions."

Culture determines whether controls survive contact with engineering. Sam Chehab, head of security at Postman, argued that the binding constraint is rarely tooling. "The cybersecurity industry has proven that there's a tool for everything. There's about 5,000 cybersecurity tools out there, but where we often struggle the most is with regard to process." Professional pride, properly engaged, outperforms enforcement. "Developers are pretty proud people, and once you have that established process, they can rip through risk pretty quickly," Chehab said.

Curry's closing observation is aimed at senior practitioners rather than their staff. "If your special sauce was that you understood the technology the best and that you could program anyone under the table, you might want to rethink what it is you're doing."

CONCLUSION: CHOICE, NOT FATE

Three findings from Black Hat USA 2026 should survive the news cycle.

The first finding is the asymmetry. Discovery has automated and remediation has not, and little on the agenda or in 73 interviews suggests the gap closes on its own. The organizations that handle this well will not buy the most discovery capability. They will fix their remediation economics - including patch cadence, change windows and ownership data - and the budget and priority contests Privette identified as the real bottleneck.

Gamblin's corrective belongs alongside it, and it is a correction about measurement rather than about risk. Based on his analysis, the exploitable share of all vulnerabilities has stayed near 4% to 5%, so a rising detection count is not by itself a rising risk. That does not dissolve the problem. A hundredfold increase in discovered vulnerabilities against a constant exploitable fraction still yields far more exploitable bugs. But it does mean the industry is measuring the wrong quantity, and Gamblin's further claim is that treating vulnerability management as unsolvable is the most expensive belief available.

The second finding is that authority, not access, is the primitive that failed. Every chapter of this report contains a version of the same omission. Identity systems record who is acting and never encode what the actor may do, because a human always supplies that judgment silently. Agents do not. The remedies proposed by practitioners converge with unusual consistency. Constrain actions rather than reasoning, enforce at the point of action, carry intent through the transaction, compartmentalize by task and keep a named human accountable for decisions that will later be examined. None of that is a product category. It is architecture and governance.

The third finding is that the first serious agentic breach may have no attacker in it. The most striking failure scenario put forward this year was Radolec's forecast rather than an observed pattern, and it involved no adversary. It was an accident, an over-permissioned system pursuing a goal without conscience, at machine speed, in an environment never told what it was forbidden to do. That risk is not detected by threat intelligence, and in the one agentic incident Forrester's Mellen described, it was not caught by alerting either.

There is disagreement about the balance of advantage, and it should be described accurately rather than staged as a symmetry it does not have. Hultquist and Carmakal argue that defenders currently hold the edge.

The Cyber Threat Alliance's Daniel makes the stronger version of the claim and flags it himself as controversial. Notably, no interviewee argued that attackers are currently ahead. The counter-case is built from adjacent evidence rather than from a rival verdict. It rests on Forrester's finding that 76% of enterprises have an agentic use case in production, an incident that surfaced through investigation rather than an alert, an insurance market that cannot yet price the risk and a liability doctrine that does not require intent. That asymmetry is worth noting. The optimistic case is argued directly by people with visibility into adversary behavior. The pessimistic case is assembled from deployment, detection and liability evidence none of the optimists addressed.

One piece of evidence sits outside that framing and deserves testing rather than filing. Horizon3.ai's finding that an offensive agent takes a honeypot at roughly two and a half times the human rate is the only claim identifying a structural weakness in the attacker's tooling rather than a race the defender has not yet lost. If it replicates independently, it is the cheapest defensive advantage described this year.

Which is why the most useful sentence spoken at Black Hat this year was also the least technical. Asked which future would arrive, Inglis declined the premise that anyone is a passenger. "It's not fate that will determine whether one of those or the other plays out - it's choice."

Prepared by ISMG Content Intelligence & AI Innovation Department. This analysis has been drawn from the Black Hat USA 2026 Briefings agenda (109 sessions, 20 tracks) and 73 executive interviews recorded at the ISMG.Studio, Mandalay Bay, Las Vegas.

About ISMG

Information Security Media Group (ISMG) is the world's largest media organization devoted solely to cybersecurity, information technology, artificial intelligence and operational technology. Each of its 38 media properties provides education, research and news that is specifically tailored to key vertical sectors including banking, healthcare and the public sector; geographies from North America to Southeast Asia; and topics such as data breach prevention, cyber risk assessment, OT security, AI and fraud. Its annual global summit series connects senior security professionals with industry thought leaders to find actionable solutions for pressing cybersecurity challenges.

Contact

(800)944-0401 • sales@ismg.io

