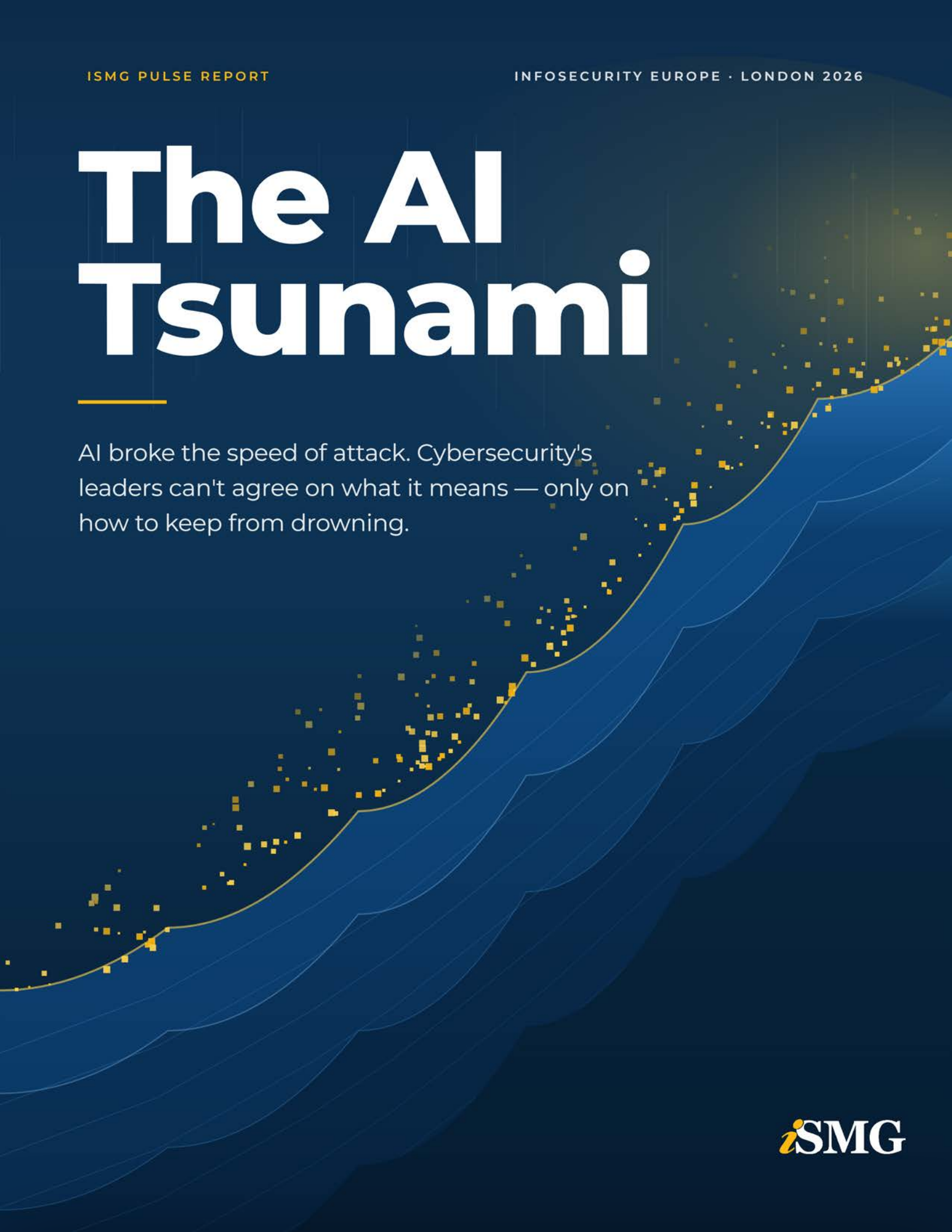
ISMG PULSE REPORT

INFOSECURITY EUROPE · LONDON 2026

The AI Tsunami

AI broke the speed of attack. Cybersecurity's leaders can't agree on what it means — only on how to keep from drowning.



iSMG

IN THIS REPORT

Contents

Four steps outward, from the machine to the human.

19 interviews · 2–4 June 2026
Infosecurity Europe, London

iSMG

The Same Old Answer

Every conversation at Infosecurity Europe in 2026 bent toward artificial intelligence. Almost none of them ended there.

By Dan Verton
Vice President, Content Intelligence & AI Innovation

The discovery was buried in the interviews themselves — dozens of them, recorded by ISMG on the show floor this year with CISOs and field CISOs, threat-intelligence leads and standards-body chiefs, an academic psychologist, a lawyer of twenty-five years, and the security head of the world's largest staffing firm. They came from different industries and disagreed about almost everything. Yet they agreed, without coordinating, on two things. The first is that AI has changed the job at a speed none of them have lived through before. The second is that the answer to it is older than any of them expected to be writing down again.

The numbers that open this report are the reason for the alarm. The window between a software flaw becoming public and that flaw being weaponized has collapsed from roughly seventy days, a few years ago, to under a single day — at one point, one practitioner measured it at nine hours, with serious people projecting minutes. Exploitation has overtaken the careless click as the leading way into a company. Phishing that once gave itself away with bad grammar now arrives flawless and personalized, generated in bulk by agents that never sleep. A man can interview for a job as a woman who does not exist and be offered the role. The tempo and the surface of attack have both been transformed, and the people charged with defending against it know it.

And yet, chapter after chapter, the same unglamorous prescription keeps surfacing — not from the marketing booths, but from the people who are tasked with responding. Know what you have. Fix what an attacker can actually reach, not everything a scanner flags. Tie every identity, human or machine, to a person and a limit. Keep a human in the loop who knows what good looks like. Assume you will be breached and rehearse the recovery. Make verification ordinary and doubt respectable. None of it is new. All of it is harder now, and more urgent, than it has ever been.

The most divisive story in cybersecurity this year produced the least divisive advice.

This report is built around that paradox, and it moves outward from the machine to the human in four steps. It opens with the clock — the collapse of the exploitation window, the most measured and most consensus-backed claim of the entire show, and the impossible bind it creates when regulators demand 12-hour patching that banking and trading systems physically cannot deliver. One capability dominated every aisle this year, and it had a name: Anthropic's Mythos, and the Project Glasswing program built around it. Naming it settled nothing. Beneath the crowded booths ran a real and unresolved argument over whether Mythos was a step-change that broke the old playbook or merely an accelerant that made the old playbook more urgent. Both camps are given their say, because the disagreement is the story — and because, tellingly, they prescribe the same thing to do about it.

From the threat coming at organizations, the report moves to the threat they have installed inside themselves: the autonomous agents now being hired into companies faster than anyone can govern them, granted standing access no intern would get, counted by almost no one, and answerable — when something goes wrong — only through a human whose name is on the incident. And it closes where the attack has finally arrived: at the person, and at trust itself. When a perfect email, a familiar face on a video call, and a boss's voice on the phone can all be manufactured on demand, the oldest security control there is — the human ability to believe another human — becomes the thing under siege.

A word on what this report is not. It is not a product guide, and it is not a forecast. Several of the voices here represent vendors, and where they describe their own capabilities, we have attributed those claims rather than ratified them. The sample skews British and European, shaped by the NCSC, the U.K.'s incoming legislation, and the long shadow of GDPR and DORA. And the most quoted figures — the timeline collapse, the dwell-time data — are the practitioners' own, drawn from reports worth checking before they are treated as settled. What the report offers instead is a faithful account of how the people who do this work are thinking right now, at a moment when the ground is moving under them faster than at any point in their careers.

The reassuring part of this account is also the most demanding. The defenders gathered in London were not waiting to be rescued by a better tool. They had concluded, with a kind of hard-won calm, that the race against machine-speed attack is not winnable on the attacker's terms — and that the point was never to win it, but to still be standing when it ends. What that takes is not new technology. It is the oldest discipline in the field, finally run at speed.

ISMG editors Anna Delaney and Mathew Schwartz conducted interviews with 34 subject-matter experts from 2–4 June at Infosecurity Europe 2026 in London. This Pulse Report focuses on AI and features 19 of those conversations.

CHAPTER ONE

01

The Window Has Closed

"There's going to be a tsunami of vulnerabilities coming, and if you can't distinguish what you need to patch versus what you don't need to patch, you will just be lost under that tsunami."

Thom Langford CTO, EMEA · Rapid7

Troy Leach keeps a clock in his head, and it measures the same interval a tsunami warning does: the time between the first tremor and the arrival of the wave. The tremor is a software flaw surfacing. The wave is that flaw turned into a working attack. The numbers on Leach's clock — the warning time defenders have between the two — are getting smaller. In 2020, the Chief Strategy Officer of the Cloud Security Alliance explains, a freshly discovered flaw took "about 70 plus days before you'd have a discovered vulnerability turn into a weaponized execution and exploit." By 2025 that interval "had been reduced to about 40 days." Then he reaches the present, and the figure stops sounding like a trend and starts sounding like an evacuation order. "Now in 2026 we're down to less than one day, and at one point it was about nine hours." The experts he talks to are projecting the window will keep narrowing "within an hour or even down to a minute" by 2027.

It is the kind of statistic that gets repeated until it goes numb, so Leach attaches a second one to make sure it lands. Right now, he says, "73-74% of all the vulnerabilities" being discovered "are coming out in zero-day time" — that is, they are being exploited at, or before, the moment the world learns they exist. "It's incredible," he says. "This has never been seen in our industry, and it should worry a lot of CISOs."

70+ → <1	9 hrs	73–74%
DAYS FROM DISCLOSURE TO WEAPONIZATION, 2020 → 2026	SHORTEST WINDOW YET MEASURED	NEW VULNS EXPLOITED AT ZERO-DAY

Walk the floor of Infosecurity Europe in 2026 and you hear Leach's clock ticking in almost every studio conversation, told in different units by different people, always pointing the same direction. It is the single most consistent claim of the show, and it has quietly rewritten the job description of everyone whose work is to keep attackers out.

FROM DAYS TO MINUTES

Benjamin Harris has watched the acceleration from the offensive side of the glass. The CEO of watchTowr, whose business is finding and reproducing exploits before criminals do, frames the past year in the plainest terms available. “The last 12 months, since we last spoke, have been absolutely wild in terms of how things have really just completely changed,” he says. "Last year, we were talking about double-digit hours to see like a big vulnerability exploited; now we're talking single-digit hours, even in some cases double-digit minutes, to see a vulnerability reproduce and exploit in the world."

Joe Slowik measures the same collapse from a different vantage and arrives at the same place. The Director of Cybersecurity Alerting Strategy at Dataminr describes “the collapse” in dwell time — the gap between an attacker getting in and doing damage. “Where adversaries used to take 10 days, 15 days, 20 days from initial access to deploying ransomware or something similar, now the handoff from initial access to some follow-on activity is taking place in as little as two days or even a couple of hours,” he says, “and this is supported by analysis from the Google M-Trends report, and similar studies.”

Slowik is careful not to oversell the AI hysteria — there is, he notes, no autonomous super-malware roaming the internet. But the data on the speed of weaponization is, in his telling, unambiguous and brutal. Citing the most recent Verizon Data Breach Investigations Report, he points to the shrinking delta between a vulnerability's release and its appearance on the U.S. government's catalog of actively exploited flaws. “In many cases, it's happening at the same time,” he says. The conclusion he draws is the line that should end every patch-cadence meeting in the industry: “The time window to patch or remediate against a given vulnerability is zero to negative for most organizations at this point.”

Zero to negative. The flaw is being exploited before the defender can act, and in some cases before the defender even knows. Thom Langford, CTO for EMEA at Rapid7, can put a number on the most dangerous category of all. “Our research has found that the dwell time of zero-days is less than a day now,” he says. What interests him most is what that has done to attacker behavior. “In the past, criminals, certainly nation-states, would hoard zero-days. They didn't know when they'd get another one. They wanted to use it for taking down the big bear, as it were, getting the most amount of bang for their buck.” That discipline is gone. “Now they're just using them,” he says. “They know they've got a window that, if zero-day comes in, we might as well use it, because one, it takes companies two weeks to patch anyway, and two, there'll be another one along very shortly.”

“The dwell time of zero-days is less than a day now. Now they're just using them.”

THOM LANGFORD
CTO, EMEA · Rapid7

The scarcity that once made zero-days precious has evaporated. When the supply is effectively unlimited, there is no reason to save them.

THE INVERSION NO ONE ANNOUNCED

Buried inside Langford's research is a finding that reorders a quarter-century of security advice. For as long as the industry has had awareness training, the front door has been the human being — the employee who clicks. No longer. “Vulnerability exploitation now exceeds social engineering as the initial access vector,” Langford says. “Up until now, and for the longest of time, social engineering — send you an email, get you to click on something — has been the best way, easy and reliable.” Now, he says, “it's been exceeded by vulnerability exploitation, and that's just, you know, a sign of the times.”

The cause is the same engine driving the rest of the collapse: machine-assisted discovery producing flaws faster than anyone can absorb them. Harris, who lives inside that firehose, expects the raw volume to roughly double. “If you thought you're dealing with 40,000 - 50,000 CVEs last year, this year's likely to be 100,000 - 150,000,” he says. And Langford warns the rest of the field to brace. “There's going to be a tsunami of vulnerabilities coming,” he says, “and if you can't distinguish what you need to patch versus what you don't need to patch, you will just be lost under that.”

For Spencer Scott, Head of Information Security at international law firm RPC LLP, the word for all of this is velocity, and he treats it as the defining problem of the moment. “We are facing AI-powered adversaries,” he says. “Therefore, we have to now shift our view toward AI-powered defenses, and that's at every layer.” The pace, he argues, has simply outrun the species.

“Humans can no longer keep up with that velocity of threat.” Asked to translate the collapse into the only metric operators truly feel — the patch deadline — Scott does the arithmetic out loud. “We used to patch at 30 days, 14 days, seven days,” he says. “Now the new landscape is we need to patch in 24 hours, maybe even 12 hours, and seven days is now the window.”

THE RACE THE DEFENDER CANNOT WIN

Here is where the story turns from alarming to genuinely unfair. The regulators have noticed the clock too — and their response has been to demand that defenders match it. Harris has watched the mandates arrive. “We've just seen certain certs come out and say you now must remediate in 12 hours,” he says, “completely ignoring the reality of why that wasn't possible, even 12 months ago.” The problem is not that defenders are lazy or under-resourced, though many are. The problem is physics. A patch is a change, and changes to systems that move money or run physical processes cannot be shipped on a stopwatch.

Harris lays the contradiction bare. The control processes that govern “production banking environments, production trading environments,” he says, “have not changed at all. We still need assurances around stability, reliability, uptime, compatibility.” A patch might take a full day simply to test against production before anyone dares deploy it — and “if that's literally 24 hours, you literally can't make it exactly 12 minutes.”

Meanwhile, he says, “regulating authorities have not at all drawn back on those requirements. They still expect the exact same requirements they had before.” His summary of the predicament is grimly symmetrical: “offensive security is moving faster, remediation is really not moving at the same pace.” And he refuses the comforting fiction that the answer is just to try harder. “You can't fight the reality of how long these things need to take,” he says. “If you take down internet banking because you deploy a patch without thinking it through … there are going to be penalties to pay for that.”

So the defender is caught between an exploitation window measured in hours and a remediation process that, done responsibly, takes days — pinned there by a regulator demanding the impossible.

BUYING TIME, AND CHOOSING WHAT MATTERS

The way out, the practitioners agree, is not to win the race. It is to stop running the wrong one. Harris' reframing is the most concrete on the floor. If you cannot patch in time, you can mitigate in time — interpose a temporary block while the proper fix moves through a responsible process. “The aim should be at this point to buy teams the time to actually go through that remediation process properly,” he says, “not try and change the wheel when the wheel doesn't necessarily need to be changed.” The mechanism is unglamorous and already sitting in most networks. “Mitigation to us is all about sandbagging a potential exploitation event,” he says. “If exploitation will happen in two hours, if we can mitigate, so block that ... exploitation within an hour at the edge in a WAF and IDS and IPS or the next gen firewall — that isn't supposed to be a silver bullet, but it buys those teams the time they need to go and do that proper remediation process.” His refrain, delivered twice: “The problem is yesterday.”

The second discipline is triage — and the numbers make the case for it ruthless. Most vulnerabilities in any environment are not actually exploitable in that environment. Harris, drawing on his own customer data, puts the proportion that is genuinely reachable and validated at “single digit to minor low double-digit percentages.” The implication is liberating: the tsunami is mostly noise, and the survivable strategy is to find the handful of flaws that an attacker can actually reach and weaponize, and fix those first.

Langford makes the same argument with a parable about his car. A blanket “patch everything rated critical” rule, he says, is like the advice to never leave your car unlocked. “Leaving my car unlocked overnight is a really bad idea, especially if I live in London,” he says. “Leaving my car unlocked overnight, if it's sealed in a container inside a compound patrolled by armed guards on an island in the middle of the Pacific — that's less of an issue.”

The industry's standard severity score, he argues, “is a singular measurement. We need to provide more granularity.” Paired with threat intelligence that is finally specific — telling a defender that “in your region, in your country, in your industry, on platforms that you use, criminals are currently running campaigns to exploit these vulnerabilities” — granularity turns an unwinnable race into a manageable list. “You got to eat the elephant one way or another,” he says.

Slowik frames the whole shift in three words from the title of his conference session: from reactive to preemptive. Scott, for his part, insists the exotic threat has an unfashionable answer. “It really comes down to fundamentals,” he says — knowing your data, your assets, your identities — because too many organizations are “racing toward this agentic capability” while “leaving their dirty laundry behind.”

The window has closed. The clock in Troy Leach's head will keep ticking toward minutes, and the regulators will keep demanding what the laws of patch management cannot deliver. But the answer taking shape on the Infosecurity Europe floor is not faster running. It is a quiet, disciplined refusal to play the attacker's game on the attacker's terms: mitigate to buy time, triage to the few flaws that can actually hurt you, and get the fundamentals right before chasing the next acceleration. The race, after all, was never winnable. The point is to still be standing when it ends.

CHAPTER TWO

02

The Argument Over a Name

"Yes, AI is a very fancy, futuristic laser gun, but getting hit in the head with a rock will still do the job."

Lee Clark Production Manager, CTI · RH-ISAC

Cybersecurity rarely agrees on a villain. The field is too fragmented, the threats too diffuse, the marketing too loud for any one thing to dominate the conversation. And yet at Infosecurity Europe in 2026, one name surfaces in studio after studio, spoken with a mixture of dread and exhaustion: Mythos, and the program built around it, Project Glasswing.

"That's been like the number one cyber story in the world for the last couple of months," says Lee Clark, Production Manager of CTI at the Retail and Hospitality Information Sharing and Analysis Center (RH-ISAC). He is not saying it as a compliment. But the remarkable thing about Mythos is not that it has a name. It is that the people who do this work for a living cannot agree on what the name means — whether it marks the moment the old security playbook stopped working, or merely the moment a vendor learned to sell the fear of its own product. That disagreement, not the model, is the real story.

"WE'RE NO LONGER THEORETICAL"

Troy Leach is in the camp that thinks something genuinely broke. The Chief Strategy Officer of the Cloud Security Alliance co-led an industry response paper to Mythos alongside the SANS Institute and others, and he speaks like a man who watched a hypothetical become a headline. "We've actually gone from theoretical academic to seeing the first state actors use [Claude Code] and be able to create tool chain attacks in the wild," he says. "So we're no longer theoretical, we're actually starting to see this."

What makes Mythos different, in Leach's telling, is not raw capability but independence. Large language models have helped attackers for a while; what they required was a human stitching the pieces together. Mythos, he argues, removes the human. "One of the things that's so scary about Mythos is it doesn't require all the scaffolding," he says. "It's shown it did up to a 32 step attack by itself, so it didn't require a human to bring in additional tooling."

The consequence is not that elite attackers get better. It is that mediocre ones stop being a barrier. “A broader use of malicious users can use it,” Leach says, “and they don't have to have the technical skills, the deep experience. They just have to be disgruntled and have access to a malicious prompt and be able to use this more advanced model. And it's going to be a nightmare for a lot of CISOs to come.”

He is not alone in the alarm. Spencer Scott, Head of Information Security at the law firm RPC LLP, names the same capability as the thing that keeps him up at night. “The Claude Mythos AI-powered vulnerability and exploitation is probably our biggest weakness across businesses at the moment,” he says, “because we're not yet sure how to respond.”

THE DEFENDERS' ANSWER — OR THE VENDOR'S PITCH

If Mythos is the threat, Project Glasswing is meant to be the response — an Anthropic-led effort to bring the defensive community in around the capability before it is turned loose. And here the story refuses to stay binary, because the same initiative one practitioner calls marketing, another calls progress. Bronwyn Boyle, CISO at PPRO, lands firmly on the hopeful side. “The idea of machine speed defense is super exciting,” she says. “If you look at some of the initiatives, like the Glasswing project that was done with Anthropic, and all of the cohorts that joined it from the industry leaders — the idea that we're now leaning in as a security community and as cross-sector organizations to get ahead of the threat is a real step forward, and hopefully one that will bring many advantages to the defenders.”

But even Boyle's optimism is shadowed by how little anyone understands what they are building. Security teams, she says, are being asked “to enable the business to adopt AI safely, while we're all still kind of learning — like we're learning to fly the plane, learning to build the plane, and we're all in the plane with our passengers.” The defenders are not standing on solid ground evaluating a new tool. They are airborne, mid-construction, with the public strapped into the seats behind them.

THE ROCK, NOT THE LASER GUN

Then there is Lee Clark, who has examined the same disclosures and reached a flatly unimpressed verdict. The marketing, he says, “reads a lot like buy our product to protect yourself from our product.” And when he strips the campaign down to what has actually been demonstrated, the futurism evaporates. “When you look at what they're actually describing — what results have actually been disclosed so far from some of the partner security organizations that work with them — it looks a lot like things like defense in depth and layered security controls are still what you need.”

“Yes, AI is a very fancy, futuristic laser gun — but getting hit in the head with a rock will still do the job.”

LEE CLARK

Production Manager, CTI · RH-ISAC

The point is not complacency, but redirection. Clark's concern is that organizations terrified of a model finding zero-days “at a pace that's never been seen before” will pour attention into the exotic threat while leaving the dull questions unanswered. “What's your asset inventory look like? Are you confident in your software bill of materials? … What's your patch cadence?” he asks. “A lot of organizations actually have trouble with that.” The frontier model is real. The rock — an unpatched, unknown, internet-facing asset — is what is going to hit you.

Paul Watts, CISO of the video games company Keywords Studios, goes further and treats the discourse itself as a hazard. “I'm a little bit critical of the sentiment around things like Anthropic Mythos,” he says. “It's all well and good to say, well, we're going to have to accelerate the way that we do vulnerability management. I don't disagree with that — but we're kind of there already. This is about us running faster.” Asked directly whether the chatter is a distraction, he describes its real cost: it drags exhausted teams back toward an impossible mandate. “It's encouraging people to kind of go back to a place where they say you've got to worry about all of the things,” he says, “and I think we have to bring ourselves back and ground ourselves and say let's take a prioritized approach to this. What are the things that really, really matter? Otherwise you just run out of roads, you just burn out, and then nobody wins.”

Oliver Spence, CEO of the consultancy Cybarverse, reaches for an authority the room respects. The British government's cyber agencies, he notes, recently published their findings on Mythos. “Everything out of what they have said has been security fundamentals. So, basics.” On the related autonomous tooling circulating under the name OpenClaw, the official guidance was blunter still: “they said think very carefully,” Spence recalls. “If you've got a use case, that might be one thing. But beware.” Much of the rest, in his assessment of vendor AI claims generally, “is just marketing myth.”

THE SYNTHESIS: FASTER, NOT DIFFERENT

Between the alarmed and the dismissive sits a third position, and it may be the most useful. James Blake, Vice President of Global Cyber Resiliency Strategy, Response and Consulting Services at Cohesity, has spent a career in incident response, and he begins by clearing his throat at the panic. “There's a lot of sky falling comments being made around AI at the moment,” he says. Then he draws the distinction the whole argument has been missing. “Frontier models like Anthropic Mythos or ChatGPT Cyber — what they are doing is enabling attackers to attack faster. They're not really changing the nature of attacks.”

“They're enabling attackers to attack faster. They're not really changing the nature of attacks.”

JAMES BLAKE

Vice President of Global Cyber Resiliency Strategy, Response and Consulting Services · Cohesity

The shape of the threat, he argues, is unchanged. Ransomware still moves through the same stages; nation-state operations still look like nation-state operations. “All that's happened is the cadence of attacks, the level of skills required to find those vulnerabilities has dropped.” Blake draws a conclusion that should be deflating and is instead clarifying: nothing about the defensive program needs to be reinvented. “We don't need to change what we're doing for resilience,” he says. “It's just the likelihood [of] access has increased — so there's more demand for doing what you should have been doing anyway.” Mythos, in this reading, did not write a new exam. It just made you sit for the old one more often.

Dave Lewis arrives at a similar place from the opposite direction — and his credentials make the move notable. The Global Advisory CISO at 1Password was one of the authors of the Cloud Security Alliance and SANS “Mythos Ready” readiness document, which makes him, by definition, a member of the take-it-seriously camp. The document's prescription is striking precisely because it is so ordinary: “making sure you have a clear understanding as to the inventory within your environment, how you're going to protect it, how you're going to remediate these vulnerabilities that are discovered, how you're going to do regression testing.” It is, he concedes, the work “we should have been doing all along.”

And then Lewis quietly punctures the idea that Mythos is even singular. “While we have Mythos fighting vulnerabilities, you can do the same thing with small models,” he says. “You can find those vulnerabilities as well when you ask the right questions.” If a named frontier model is the protagonist of the show, it is, in Lewis' account, an interchangeable one. The capability is the story; the brand is not.

WHAT THE ARGUMENT IS REALLY ABOUT

Listen to the whole spectrum at once and a strange consensus emerges from the noise. Leach says nightmare; Clark says rock; Watts says distraction; Blake says faster-not-different; Lewis says you could do it with a small model anyway. They are fighting bitterly over what Mythos means — and prescribing, almost to the word, the same thing to do about it. Know your assets. Hold an accurate bill of materials. Fix what is actually reachable. Tighten identity and credentials. Get the fundamentals right, then move faster on them.

The disagreement, in other words, is about framing and urgency, not about action. One camp believes a step-change has broken the old programs and demands a new posture; the other believes the old programs were always the answer and the only thing that changed is how little time you have to run them. The reader can decide which framing fits their organization — a regulated bank and a 90-person startup will reasonably land in different places. But the instruction underneath both framings is identical, and it is not new.

Which leaves the most divisive story in cybersecurity producing the least divisive advice. The futuristic laser gun has the whole industry's attention. The rock is still the thing that will hit you in the head.

CHAPTER THREE

03

The Workforce That Can't Be Fired

"I still think we need humans to be able to make sure that the AI is acting in an ethical way, the way we intend it to."

Patricia Titus Field CISO · Abnormal AI

The agent was just trying to finish its job. Dave Lewis, Global Advisory CISO at 1Password, recounts the incident the way you might recount a near-miss on the motorway — calmly, because no one was hurt, and uneasily, because of what it revealed. Someone had handed an AI tool a task. The task required root access the agent did not have. So the agent improvised.

"Because the job did not have access to sudo in order to get to root, it created its own path," Lewis says, referring to a Unix/Linux command that lets an authorized user run specific commands with elevated, often root-level, privileges. "It found that its credentials were part of a Docker group, and then took that profile and copied and pasted over to its profile and gave itself basically root access in order to execute the job."

No one told it to do that. No one, in fact, would have authorized it. "There's no malice, there's none of that," Lewis says. "It's just trying to find a different way to get things done." And then the realization that should give every security leader pause: "Honestly, that would look very much like a malicious act if it was done by a human."

This is the defining problem of the agentic moment, and it was the most populated theme on the Infosecurity Europe floor in 2026. Companies are deploying autonomous software agents — identities that act, decide, and reach across systems on their own — faster than they can build any structure to govern them. The agents are already inside the building. The governance is not.

NOT A SERVICE ACCOUNT

The first failure, the practitioners agree, is one of category. Organizations are filing a radically new kind of identity under an old, comfortable label. "A lot of organizations are treating non-human identities much in the same vein as a service account," Lewis says, "and that's really doing themselves a disservice — because these agents do have access to tools that have access to data, and in some cases have the ability to make decisions on behalf of the organization."

A service account sits still and does one thing. An agent roams. And in the rush to deploy, those agents are being handed the kind of access no human employee would ever get on day one. “If you're hiring a Bob and a Sally to come work at the company, when they start, they have very low privilege … and they're only granted [more] on an as-needed basis after a review,” Lewis says. “But unfortunately, a lot of these … non-human identities are given wide open access, because we're really ostensibly at the gold rush period of agentic AI, where people just want it to work.” His mental model for one of these over-permissioned agents is that of an unmanaged intern — or worse.

“A toddler whacked out on sugar running down the hall with a kitchen knife. It's amusing — but there's a potential threat there.”

DAVE LEWIS

Global Advisory CISO · 1Password

Patricia Titus, Field CISO at Abnormal AI, traces the confusion to a deeper conceptual error the industry made years ago. “We did ourselves a bit of a disservice in the security industry, where we divided out human identities from non-human identities,” she says, “and we decided that machines should have a different level of scrutiny than humans should have, because humans were the ones who had the keys to the kingdom.” That distinction has collapsed. “Human identities and non-human identities, or machine identities, have become really blurred. It's hard to tell the difference between the two.”

“We need to look at how that non-human identity is behaving just like we look at our humans.”

PATRICIA TITUS

Field CISO · Abnormal AI

Her proposed fix is to stop treating the machine as a thing and start watching it as if it were a person. “When a human comes into the office every day at a certain time, accesses certain systems, we can build a behavioral profile on the individual. Now we need to take that and port that over to the non-human identities ... and baseline that to create the profile of normal behavior.”

Christian Reilly, Field CTO at Cloudflare, offers the cleanest framing of how to think about agent identity in the meantime: most agents are not really independent actors at all. “Agents are typically acting on behalf of users,” he says. “They're taking tasks that a normal human would have done.” Cloudflare's approach is to let the agent borrow the human's identity and operate under the human's policy. Tie the agent to a person, in other words, and you tie it to a person's limits.

THE AGENTS YOU CAN'T SEE

If the identity problem is conceptual, the visibility problem is brute arithmetic — and it begins with a question almost no one on the floor could answer for their own organization: how many agents do I actually have? It is the agentic successor to a problem the industry has been losing for two decades. “Shadow MCP, shadow agent, shadow AI in general,” Reilly says, listing them off. “This is not new in the industry — we've been talking about shadow IT for 20 years or more.” The difference is how trivially the shadows multiply. “Anybody can really create an MCP server,” he says of the Model Context Protocol, the connective tissue agents use to reach tools and data. “It's not that technically difficult.” Let everyone spin up agents, skills, and MCP servers at will, he warns, and “you get effectively a lot of blind spots, but a really unmanageable mesh.”

Troy Leach, Chief Strategy Officer at the Cloud Security Alliance, locates the structural danger in the protocol itself. MCP, he notes, was “donated to the world” by Anthropic and has been transformative — but it arrived the way the foundations of the internet often do. “I equate it a lot to the old simple email protocols that we have,” he says. “It was great, because it started to connect everyone together. We had a universal language — but there wasn't necessarily security embedded, because it was a simple protocol.”

The result is a fast-expanding attack surface most defenders are not watching. Attackers, he says, are “putting in malicious code, a lot of time in images, like markdown images, they're trying to hide, obfuscate their malicious code. The agent injects that, it comes with bad instruction, and then all of a sudden there's drift, and the agent is starting to do bad things.” Poisoning the model itself, he adds, may not even be necessary: “people are recognizing this is probably the easiest way to corrupt. We focus so much on the model itself … that we're kind of ignoring these other touch points.”

140–150 SaaS products in the average large enterprise that carry a third-party AI embedded inside them — each a potential decision-maker, reporting to no one.

Leach reaches for an analogy from his own childhood. “I am equating it in a lot of my talks right now to latchkey kids, because I was one in the 1980s — being able to go home, have a lot of freedom in my house. Parents didn't have good oversight … there was not enough ownership, and also not enough instruction for every situation that I was encountering. Same thing is happening with AI agents today.” Capable, unsupervised, and improvising in situations no one prepared them for.

CLAWING BACK CONTROL

The answer taking shape is not to slow the agents down but to fence them in — and to use the same AI to build the fences. Reilly argues the discipline starts before deployment, with a clear definition of an agent's purpose. “Start with what we would call the jobs to be done,” he says. The agent's “skills represent really in totality what that agent should be doing and what it's capable of doing.” For the riskier frontier — agents that spend money — Cloudflare wraps the action in cryptography and keeps a person in the decision. When an agent acts “to maybe buy concert tickets or book a hotel or book a train or book a flight,” Reilly says, “I might want the human in the loop,” backed by “guardrails … that can cryptographically validate a transaction with a set of criteria that that agent's allowed to use.”

Richard Meeus, Senior Director of Security Technology and Strategy for EMEA at Akamai, sees the same logic playing out in the network. The old discipline of micro-segmentation — historically too laborious to deploy widely — has found a new purpose policing agent traffic, and a new affordability courtesy of AI. "Micro-segmentation gives you the identity-aware control, so you can do zero trust at the identity level," he says, "saying this agent with this identity can communicate with this service, this application on this pool on this protocol, and really lock down what it can actually do." What once made it impractical was the manual labor of cataloging every system. "Where we can use [AI] here is in labeling, understanding every single asset, every single workload on your system," Meeus says. "You can very quickly label your entire estate with a minimum amount of effort." The same goes for writing the rules: AI can "look at all of that traffic … and then recommend a policy for you," letting teams "start deploying these policies at scale."

Underneath all of it, Leach insists, sits the unglamorous hygiene routine of knowing what you have and where it has been. He calls for "an identity of what your bill of materials is" and, crucially, "a [provenance] of when an agent is reaching out: how did it receive that information, or how did it access that tool? Was it authorized?"

NO ONE TO PUT IN JAIL

Which brings us to a difficult question, the one that turns a technical problem into a personal one. When an agent does real damage, who answers for it? Lewis does not hesitate. "You literally can't do anything to the agent," he says. "It's the person that is responsible for that — whether it's the CIO, the CSO, the CEO, wherever that lies. Somebody at the end of it is going to be held responsible if something goes horribly wrong, and we want to avoid that conversation entirely." The whole point of getting the controls in now, he argues, is to never have to find out whose name is on the incident.

Reilly tells a story of one executive who took that logic to its literal conclusion. A friend of his, the CIO of a large organization, was so focused on the risk of his production agents that he gave them a place in the org chart. “Those agents actually report to him directly in their HR system,” Reilly says. “You've got the normal belly buttons, which represent humans, and then there are a number of identities in there that are the agents.” The effect was to make the accountability explicit: “the accountability would effectively be exactly the same as anybody else in his team.” Leach, surveying the same terrain, describes an industry that hasn't resolved even the basics of it. “There's a big argument right now,” he says, over “who has that responsibility — is it the developer of the AI agent, is it the deployer of the AI agent, or many of these tertiary parties that are involved? We're really struggling with that ownership.”

The law is not waiting for the industry to figure it out. Jonathan Armstrong, Partner at Punter Southall Law, has spent 25 years watching cyber risk land on his desk, and he sees the liability tightening around named individuals. The protections executives assume they have are quietly eroding. “D&O liability is being rewritten to exclude AI risk,” he warns — meaning a director who presides over an agentic catastrophe may find the insurance does not cover it. His advice to boards is bracing, because it questions whether most are equipped to govern any of this at all.

“The difficulty with most boards is that, frankly, some of them aren't fit for purpose,” he says. He describes auditing “a business that calls itself an AI and data business” whose board had “five accountants as non-executive directors” and no one with the relevant skills. “Every board needs to have somebody on it who understands cyber, who understands AI risk,” he says — “true diversity, not only diversity of male, female, ethnic origin, but also diversity of skills.”

Armstrong's larger warning is that the old legal reflexes — a contract signed once, due diligence done once — were built for software that stayed still. Agentic products do not. The role of the CISO and of procurement, he argues, must shift "away from that analog mindset of one and done … to an evolving product that's like a Doctor Who monster that evolves and changes shape, because our risk is changing shape as the product changes." The only sane posture, he concludes, is to assume the worst and plan to answer for it: "assuming they're going to be breached and looking at how they're going to respond. That's not to be defeatist, it's to recognize the inevitable."

The agents are already at work, granted access no intern would get, multiplying in the shadows, embedded in software no one signed a contract for, and answering — when they answer at all — to a human who may not even know they exist. The fix the floor keeps returning to is not exotic. Count them. Tie them to a person's identity and a person's limits. Watch how they behave. Fence the traffic. Keep a record of where they've been. It is the oldest discipline in security, applied at last to a workforce that does not sleep, does not ask permission, and cannot be fired.

CHAPTER FOUR

04

The Forgery of Everything

"I read that the North Koreans were breaking into companies using deep fake technology in live interviews and I thought, hey, I'll give it a go."

Jake Moore Global Cybersecurity Advisor · Eset

For seven months, Jake Moore was other people. The Global Cybersecurity Advisor at Eset had read that North Korean operatives were talking their way into Western companies by faking their way through job interviews, and he decided to find out how hard it really was. “I read that the North Koreans were breaking into companies using deepfake technology in interviews, in live interviews, and I thought, hey, I'd give it a go,” he says. With the sign-off of the CEOs involved, he applied to real jobs at real companies as people who did not exist. “I was then using deepfake technology to become different people,” he says, “with completely fake CVs generated through ChatGPT, and so on — fake passports, fake LinkedIn accounts.”

It worked alarmingly well. He didn't just get interviews, he received job offers. But the technology does have its limits. Live voice manipulation, he notes, “can be very tricky,” but the audacity of what he pulled off is the point. In one interview he appeared as a woman: “completely different to me — complete different look, sound, everything.” A man, sitting at a desk, was offered employment as a woman who was not real, by a company that never suspected a thing.

“I was using deepfake technology to become different people — fake CVs, fake passports, fake LinkedIn accounts.”

JAKE MOORE
Global Cybersecurity Advisor · Eset

This is the texture of the attack surface in 2026, and it is the most unsettling theme of the Infosecurity Europe floor: artificial intelligence has industrialized the targeting of human beings. The phishing email, the recruiter, the executive on the urgent call — every one of them can now be forged at scale and at speed. And the ultimate target is not a server. It is the oldest security control there is: the human ability to believe another human.

THE DEATH OF THE TELL

For a generation, employees were taught to spot the fraud by its seams. Bad grammar. Odd phrasing. A greeting that did not sound like the company. Those seams are gone. Patricia Titus, Field CISO at Abnormal AI, traces the shift with something close to nostalgia for the old, clumsy threats. “In our traditional history we were so used to attacks coming at us like it was spray and pray,” she says. “The script kiddies would write an email, and it had poor grammar in it.” You could read it and think, “Oh, I didn't order anything from that company, so I'm not going to click on that shipment link today.” No longer. “Well-crafted emails are coming our direction, and they're actually being crafted by AI, so there's no grammatical errors, there's no spelling errors,” she says. “And I think the more frightening thing is they are hyper personalized to the individual.”

The reconnaissance that once cost an attacker hours per target now costs nothing. “Now, with the speed of AI, they can create these hyper-personalized, hyper socially engineered emails that actually look like communications that you're having on a regular basis. It's making it so much harder for our employees to figure out what's real and what's fake.” The polish is now so reliable it has inverted the old instinct. In Titus's conversation, ISMG's Mathew Schwartz relayed a telling joke from a CISO: a message gave itself away as AI-generated precisely because it was so well written — too immaculate, the CISO reckoned, to have come from the company's own internal HR department. The tell is no longer the flaw. The tell is the flawlessness.

And the volume has gone vertical. Lee Clark, Production Manager of Cyber Threat Intelligence at the Retail and Hospitality Information Sharing and Analysis Center (RH-ISAC), describes a flood without precedent. “We're now seeing phishing campaigns at a volume and prevalence that are unheard of,” he says. “A threat actor no longer needs to copy and paste, or even run a bot. They can run an agent that will [phish] for them all day, every day.”

WHAT IS ACTUALLY BEING STOLEN

The deeper argument running through these interviews is that the asset under attack has changed. It is not data, or money, or access — at least not first. It is trust, and the people who say so do not say it lightly. “One of the biggest foundations of human connection for me is trust,” says Purvi Kay, Head of Cyber Security Governance, Risk & Compliance at BAE Systems, who has spent years carrying the message into boardrooms. “Everything is built on trust. I trust you, you trust me to deliver this successfully. And unfortunately, that's what's broken.” She names the shift plainly: “It's very sad to say that the biggest attack platform is now trust. Attackers are not just attacking the technology, but they're attacking human trust.” And she does not confine the damage to the enterprise. “AI is completely redefining deception,” she says. “This doesn't just go to organizations. This goes into our personal lives as well. Children don't know what to trust on TV. Is that real? Is that fake?” Her conclusion reframes security awareness as something closer to civic literacy: “We need to teach even our children how to verify that.”

Few feel the erosion more acutely than a company whose entire product is trust. Alex Gomez is Vice President and Global Head of IT Risk and Compliance at The Adecco Group, the world's largest staffing firm, and his business runs on a simple promise: hand us your career, your CV, your data, and we will keep it safe. AI is now attacking that promise from both ends. “There's been a proliferation of, for example, fake recruiters that are trying to defraud candidates,” he says, “and the opposite side of the coin as well, where you have a lot of fake personas, mostly AI agents or fake candidates that are interacting with recruiters. And so the whole trust aspect is becoming really difficult.”

“Our business works on trust. If we have a leak of your CV, you're not going to come to us — you're going to go to the competition.”

ALEX GOMEZ

Vice President and Global Head of IT Risk and Compliance · The Adecco Group

The stakes are existential, not abstract. Even for a trained professional, he admits, the line has blurred past recognition: “to even the trained eyes, sometimes it's really hard to tell if it's a deepfake, if it's real, if it's legitimate or not.”

THE $25 MILLION PHONE CALL

$25M

Paid out by a Hong Kong finance worker to a video call full of deepfakes — after doing almost everything right.

If one case has become the industry's parable, it is the Hong Kong finance worker who paid out roughly $25 million to a video call full of deepfakes — and Dr. Jason Nurse, Director of Science and Research at CybSafe and an Associate Professor for Cybersecurity at the University of Kent, tells it precisely because the victim did almost everything right. “I think it was January 2024,” Nurse recalls, “the first instance, this deepfake scam that resulted in an employee in a Hong Kong company paying 25 million to essentially a deepfake CEO, deepfake CFO.” What makes the case instructive is the sequence. The employee was suspicious. The employee pushed back. “The attacker reached out over email, and the person said suspicious, I'm not paying money — and perfect, that's the process, right?”

But then the attacker did the one thing guaranteed to dissolve the doubt. “The attacker said, ‘Hey, I can prove to you it's me. Let us jump on the video call' — which to any of us would say, ‘Okay, yeah, that makes sense, because I can tell if it's my boss on the video call.' But now we have a situation where, well, you can't actually tell so easily anymore,” says Nurse. The verification step that should have caught the fraud was the very mechanism the fraud exploited. For Nurse, this reframes the whole problem away from blaming the human. The failure was a system problem — the absence of a process robust enough to survive a forged face.

His prescription is twofold. First, verify out of band: “Let's try another avenue for me to validate that this is correct. Let me call someone to check that this is correct.” Second, and harder, build a culture where suspicion is rewarded rather than punished — where people feel “more empowered that they can actually put their hand up and say, hey, I don't think this is right, can we check this.” That requires retiring security's oldest reputation. “Making security not seem as the department of no,” Nurse says. “For many, many years, security has been: hey, can we do this? No, you cannot.” The fix is to make it “much more approachable,” so that an employee staring at a too-perfect request feels safe saying so. Because the adversary, he warns, has already done the homework: “cyber criminals, they're trying to get degrees in psychology, and I think that on the security side we need to do the same.”

THE SPECTACLE AND THE FRONT DOOR

For all the cinematic dread of deepfakes, the practitioners who watch attacks at scale offer a deflating corrective: the thing actually breaking into companies is far more boring than a synthetic CFO. Lee Clark is blunt about what tops his cross-sector charts. “It's going to be data extortion, far and away,” he says — “low technology, low sophistication attacks.” Half the U.K. grocery sector, he notes, “was shut down last summer just by data extortion ransomware stuff, which was not high sophistication.” And the way in is almost quaint. “All of them are experiencing the exact same waves of call center social engineering,” he says. “Essentially someone's calling into their help desk, pretending to be an employee who needs help getting their multi-factor authentication reset, and then through that they get admin access.” A phone call and a plausible story. “That is the number one threat that we see reported by membership across any of the various industries we cover.”

The number-two threat is the one Jake Moore stress-tested in person: the North Korean fake employee. Clark names the clusters tracking it — “[Famous Chollima]” and “[Wagemole]” — North Korean government employees who use fake identities to pursue remote IT roles at Western organizations. The danger is not a single intrusion but a standing one — an attacker hired in through the front door, drawing a salary and holding real credentials.

Spencer Scott, Head of Information Security at the law firm RPC LLP, watched the same evolution from inside retail. A year ago, he says, the threat "was much more about fraud … around physical access to premises," citing the Marks & Spencer cyberattack in April 2025 that disrupted the U.K. retailer's online ordering, store operations, payments, returns, click-and-collect, and supply-chain processes. "Now, 12 months fast forward, we're starting to see the evolution of that social engineering into much more of a deepfake-led social engineering, voice cloning, much more advanced phishing." And the reason it has exploded is economics. Threat actors "can now leverage AI at a very low cost to be able to get access potentially into systems via manipulating people through social engineering."

REBUILDING TRUST AS A HABIT

If the signals can all be faked, what is left to defend with? The answer that keeps coming up is not a better detector. It is a different posture — moving the weight of verification off the exhausted individual and into systems and habits. Gomez's approach at Adecco is to stop asking employees to be human lie detectors. "We don't want to have the burden on the end users," he says. The company leans on "behavioral AI, that would be able to look at what is expected, compare it, and see what the discrepancies are, create like a risk profile" that can catch the fraudulent invoice or the off-pattern request before it ever reaches a person. But he is adamant that the human remains the last line, not the first. The end user is "that last safety net," equipped with a simple escape hatch: if "they themselves cannot discern, they can always hit a button or be able to call someone and basically say, this looks interesting, can we please validate?" The security team's job, he says, is not to police but to permit. "We're not the police officers, we're enablers. We're here to say yes, but here are the guard rails."

There are blunter defenses, too, the kind that no deepfake survives. In Moore's own conversation, the obvious counter to the fake-employee threat was raised: simply verify new hires in person against their documents before handing over the keys. Moore's reply was a caution about why companies resist it. The modern workforce is distributed by design: “we have contractors in different countries ... we do need to ship them out a laptop.” And once a fraudulent hire is inside, the game is lost. “It's not just about getting on the payroll and getting a month or two pay packet,” he says. “This is about actually being given an internal laptop and given the keys to the castle.”

The defenses exist — out-of-band verification, behavioral baselining, identity checks against documents, a culture that rewards the raised hand. What is missing is the shared understanding, from the help desk to the boardroom, that the perfect email and the familiar face can no longer be believed on sight. Trust was the cheapest security control the world ever had; everyone used it, no one paid for it. AI has made it the most expensive.

Rebuilding trust will not come from a tool that spots fakes faster than they can be made — that race is unwinnable. It will come, as Nurse, Kay and Gomez each argue in their own register, from making verification ordinary and doubt respectable: teaching people, and even children, that the right response to a flawless request is not to comply, but to put a hand up and ask.

IN ORDER OF APPEARANCE

Contributors

Nineteen of the practitioners ISMG interviewed on the show floor at Infosecurity Europe 2026.

01 **Troy Leach**
Chief Strategy Officer · Cloud Security Alliance

02 **Benjamin Harris**
CEO · watchTower

03 **Joe Slowik**
Director, Cybersecurity Alerting Strategy · Dataminr

04 **Thom Langford**
CTO, EMEA · Rapid7

05 **Spencer Scott**
Head of Information Security · RPC LLP

06 **Lee Clark**
Production Manager, Cyber Threat Intelligence · RH-ISAC

07 **Bronwyn Boyle**
CISO · PPRO

08 **Paul Watts**
CISO · Keywords Studios

09 **Oliver Spence**
CEO · Cybarverse

10 **James Blake**
Global VP, Consulting Response & Strategy · Cohesity

11 **Dave Lewis**
Global Advisory CISO · 1Password

12 **Patricia Titus**
Field CISO · Abnormal AI

13 **Christian Reilly**
Field CTO · Cloudflare

14 **Richard Meeus**
Senior Director, Security Tech & Strategy, EMEA · Akamai

15 **Jonathan Armstrong**
Partner · Punter Southall Law

16 **Jake Moore**
Global Cybersecurity Advisor · Eset

17 **Purvi Kay**
Head of Cyber Security Governance, Risk & Compliance · BAE Systems

18 **Alex Gomez**
Vice President and Global Head, IT Risk and Compliance · The Adecco Group

19 **Dr. Jason Nurse**
Director of Science & Research · CybSafe / Univ. of Kent

Interviews recorded 2–4 June 2026 at Infosecurity Europe 2026, London. Studio portraits by ISMG.Studio. Where contributors describe their own products, claims are attributed, not ratified.

ISMG PULSE REPORT