STRATEGIES FOR BUILDING A SECURE FUTURE IN THE AI ERA

Who Is Responsible for AI Risk, and What Controls Are Needed to Ensure AI Deployment Does Not Increase Security Vulnerabilities?

“The speed, scale and sophistication of cyberattacks have completely transformed. What used to be about 2.3 million attacks occurring on a daily basis back in 2024 has now increased to about 8.95 million – almost a 300% increase.”

— Haider Pasha, Chief Security Officer for EMEA and LATAM, Palo Alto Networks

Industry experts Haider Pasha, CSO for EMEA and LATAM at Palo Alto Networks; Rick Doten, CISO at Centene Corporation; Florian Jörgens, CISO at Vorwerk Gruppe; and Chuck Markarian, CISO at PACCAR, discussed enabling AI deployment and innovation while implementing controls and guardrails that maintain security.

In this video interview with Information Security Media Group, Pasha, Doten, Jörgens and Markarian also discussed:

- How adversaries are using AI to accelerate reconnaissance, phishing and deepfake-driven social engineering;
- Best practice approaches to establish guardrails and controls for responsible AI adoption without slowing innovation;
- Strategies to achieve a balance of investment between traditional controls and defenses against AI-enabled threats;
- Who owns AI risk?

Pasha is responsible for designing cross-regional threat detection and response architectures grounded in AI and cloud-native security models.

Doten oversees cybersecurity across Centene’s enterprise and Medicaid lines. He authored the CIS Controls v8, and he lectures publicly on cloud, vendor risk and leadership in security.

Jörgens guides information security strategy spanning product, manufacturing and supply chain, and is also an author and lecturer on awareness and human-centric security culture.

Markarian has run global security programs at PACCAR since 2005, steering strategy, operations, access, forensics and resilience across its IT and OT environments.

> “The biggest risk in AI is that there is no strategy and there’s only commands. We have peers whose leadership is saying, ‘Figure out how to use it because I’m told that we’re going to fall behind if we don’t.”
>
> — Rick Doten, CISO, Centene Corporation

ADVERSARIES USING AI TO ACCELERATE ATTACKS

TONY MORBIN: How are adversaries using AI to accelerate reconnaissance, phishing and deepfake-driven social engineering? How much is theory, and how much of it do you see in practice?

RICK DOTEN: They’re using it like all of us are: to learn about targets for enumeration, to take the things they learn and ask, “Now that I have this architecture, what can I do with it?” They’re using it to code, create exploits or script out different commands, and they’re using it to assist them mostly the way we would in vibe coding or in learning or checking perplexity. “What is this kind of thing and that kind of thing?”

If nothing else, it’s making them more efficient. With agents, they can now distribute and have multiple things going on at once. It’s taking one person who can now do the work of what a team of distributed people would be able to do.

FLORIAN JÖRGENS: We see that the adversaries are moving away from mass campaigns toward more hyper-personalized AI-driven attacks. It’s not that they are trying to attack a big number of people. They are now focusing on specific people – more into spear phishing and using AI to scan social media channels to get as much information as possible to concentrate on specific people. Meaning you as an attacker, you no longer need a nation-state budget to run these specific attacks.

CHUCK MARKARIAN: The quality and speed at which they’re able to generate these things is increasing. The effectiveness is going up. It’s harder to detect, and it’s coming faster.

HAIDER PASHA: Unit 42 is an arm of Palo Alto Networks that focuses on the tactics, techniques and procedures of attackers and offers incident response services and proactive services. Some of the latest results that we’re seeing from Unit 42’s research is that the speed, scale and sophistication of cyberattacks have completely transformed. What used to be about 2.3 million attacks occurring on a daily basis back in 2024 has now increased to about 8.95 million – almost a 300% increase – because they’re using AI tools to get faster.

The other interesting fact is we’re seeing the time to a successful breach, which is the KPI of the attackers, drop. They’re getting better at what they call mean time to compromise and mean time to exfiltrate. What used to be on average about 44 days a couple of years ago, in some of the most recent attacks,

> **“Adversaries are moving away from mass campaigns toward more hyper-personalized AI-driven attacks. You don’t need a nation-state budget to run these specific attacks anymore.”**
>
> — Florian Jörgens, CISO, Vorwerk Gruppe

we’ve seen this come down to below one hour. Our own Unit 42 team has done some tests that show that you can bring it down using AI to about 25 minutes. The speed, scale and sophistication that’s changed over the course of the last couple of years with AI is incredible.

DOTEN: One quick callback is that because they’re using AI to write some of these exploits, we can’t fingerprint them like we used to, where there was some way that they would write a script or certain things that we could trigger to them. Now, it becomes more anonymous. From the attribution side, it gets tougher.

BALANCING TRADITIONAL CONTROLS WITH AI-ENABLED DEFENSES

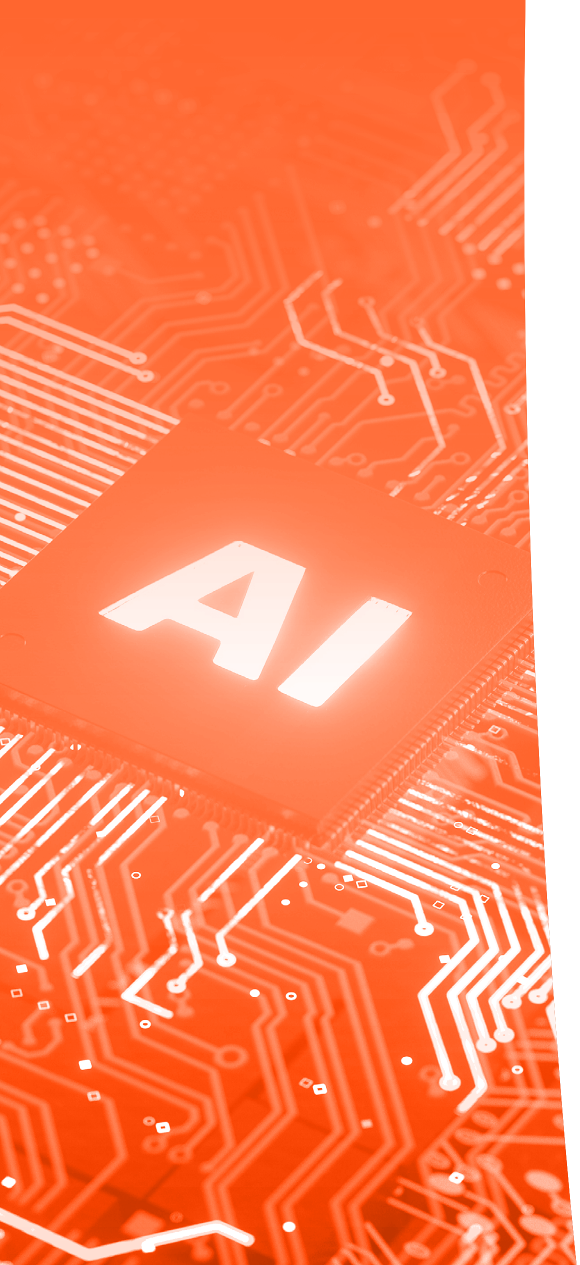
MORBIN: As adversaries are using AI to enhance what they used to do, but doing it more accurately at mass and more authentically with things like deepfakes, how do we balance our investments between the traditional controls that we still need to continue with and defenses against new AI-enabled threats, such as prompt injections or AI agents not doing MFA? How do we tackle that?

JÖRGENS: From my perspective, you don’t throw away the old playbook because this old technology is still working. Image security, identity protection, endpoint detection – these things still block the majority of attacks. We are shifting toward more resilience against new attack surfaces. For me, it’s some kind of evolution but not a replacement. What we need to focus on now are things like new layers – such as deepfake, phishing detection, anomaly analyzers and zero trust – but you can’t get rid of the old security layers.

DOTEN: I agree. Ninety percent of the controls are still going to apply. We still need identity. We still need data protection. We still need configuration management and vulnerability management. It increases the attack surface when we go into agentic AI and more complex architectures with MCP and API-based connections, data access, managing those credentials. It requires more comprehensiveness to do all the basics that we still need to apply.

MARKARIAN: If you look at what you can do today with AI internally in your systems, you’re going to use AI to find information for you. It’s going to find information that you didn’t even know you had access to. That goes back to: do we have good data controls in place? Are we managing our data properly? It’s exposing a weakness that’s existed for a

> "Because AI can be more precise and take the time to find information, every single one of us who turned on Copilot found that our SharePoints were not very secure."
>
> — Chuck Markarian,
> CISO, PACCAR



long time. It's going to drive spend in those areas for some of us to go up even higher because it's visible now and it wasn't as visible before.

DOTEN: It is highlighting weaknesses and gaps we already had because it can be more precise and take the time to find. Every single one of us who turned on Copilot found that our SharePoints were not very secure.

MARKARIAN: As you go beyond that, some of the new threats, deepfakes are going to get better and harder to detect. But prompt injection, or invisible prompt injection, you get an email now that's got a prompt in it, but it's white on white - white text on white background. You don't even see it, but Copilot does. Copilot can execute those prompts as it's scanning and going through, looking for ways to summarize your emails. I don't know what the tools are going to be yet to detect some of that stuff and take action and fix it.

PASHA: I was at the Gartner event in London, presenting a whole topic around securing agentic AI. A lot of the customers that came up talked about one particular slide that I had up there, which talked about the expanding attack surface due to AI. Traditionally as CISOs, we've been securing web application data and infrastructure. This was our core in terms of what we would focus on.

But since the adoption of large language models and now with agentic AI, you have to deal with the LLM world. You have to deal with the infrastructure when it comes to the models that are being used as part of that development life cycle. You have to focus on the action, and now you have to focus on the memory - specifically where the short- and long-term memories are. With the expanded attack surface, the recommendation that I normally explain is you need at least 70% of that foundational security.

You need about 20% of what I would call as "AI-enhanced detection and response" that brings down your mean time to detect, mean time to response capability. Maybe 10% on some sort of experimental AI security technology that could help fix a blind spot. Maybe that exists somewhere with agentic AI as an example. It is important to have that mindset in terms of being able to build on top of the foundation that you have already established with zero trust and other controls – it's important then to leverage AI to get faster and be able to scale.

WHO OWNS AI RISK?

MORBIN: Who owns the AI risk in the enterprise? We're looking at everything from the use case owner to governance, legal, HR and the CEO. There's a lot of people with a stake, but who's accountable? We're not holding agentic AI accountable because you can't put agentic AI in prison. Who owns the risk?

DOTEN: It's no different than it's ever been. Like with data protection and data governance, you have data owners, and we have the application owners. We have the business owners. It goes back to the business, depending on what your business does, and the steering committees will involve all the people that you talked about. But depending on what the business does, it could be under data governance. If there's a data officer, they may own AI. In my previous company, that's the way that was. But we are a large, well-funded company. For others, it might be a compliance issue – it might be a technical issue under IT. It varies, but I don't view it as a new thing.

MARKARIAN: The business wants to move forward fast with AI. They want to utilize AI. We want to get this down into the hands of every individual, every user to utilize it and to enhance processes, speed things up, have a presence for our investors to tell them, "Here's how we're using AI, here's how it's helping the business."

What's incumbent upon us in the CISO role is to not be blocking that, but to work with the business to help them enable it, to help them understand the associated risks with it, and to have that kind of educational or training awareness on what proper use should look like. Bring in the people who help develop some of that policy around it and ensure that that is part of our overall process in how we use AI.

JÖRGENS: For me, it's more some kind of shared responsibility. It's not a single function. As a CISO, you own the security dimension, but not the business risk of AI alone. It's more like a shared accountability model for CIOs – for the technical integration. Organizations have legal and compliance for regulatory aspects and the CISO for security and resilience. It's nearly touching different departments and areas.

MARKARIAN: If I'm not bringing up the risks, I don't know that anybody else is going to bring up the risks. It's helping make that risk visible to them so they can accept and take action on it.

PASHA: You're a risk translator or an AI risk translator. That's a very important point. I fully agree. Many of the successful implementations of AI that we've seen have two traits. Number one, the success comes from being able to build specific use cases against a specific problem using a specific set of tools, not trying to sprinkle AI on top of everything. Number two, where you have some form of an AI governance committee that's been set up with the roles that you guys talked about – CISO, CTO, chief risk officer – owning the shared responsibility.

This is the same discussion we've been having with cloud computing for the last 20 years. It's all over again talking about AI and data and governance and all the other stuff. It's like history repeating itself. But we've got some best practices that we've used from the cloud. We need to ensure we apply it from a governance perspective here on AI.

GUARDRAILS WITHOUT SLOWING INNOVATION

MORBIN: How do you establish guardrails for responsible AI adoption without slowing innovation?

DOTEN: The biggest risk in AI is that there is no strategy and there are only commands. I have a lot of peers whose leadership is saying, "Figure out how to use it because I'm told that we're going to fall behind if we don't." Or the worst example I have is a friend of mine who's leadership said, "At the end of the year, we're going to cut 10% of your staff. Figure out how to replace that with AI."

We have that MIT article from a couple of weeks ago that said 90% of AI implementations fail. But we also had 90% of websites fail in 2000, but web didn't go away. The good ones worked. What makes a successful implementation includes having a strategy, understanding the risk, understanding specifically what you do in the capabilities of it and having a good program for it. That's the biggest risk – I'm seeing a lot of people implement things incorrectly. They don't understand its uses. They're trying to route everything through MCP. There are lots of these little things that they don't know. That to me is the biggest challenge. I don't see it as a limitation. Our job is to make it and our business is to support it.

MARKARIAN: You have a lot of users who are kind of learning, investigating, trying to understand what they can do with it. To your point, Rick, it's very disjointed. It's not focused around a strategy and an approach. Over time it has to evolve to that, and it will evolve to that. But it's so early. We're still doing a lot of learning.

JÖRGENS: You need a clear set of rules – for example, no sensitive data in public AI model. It's documented. Mandatory bias check and so on. And if possible, some kind of sandbox environment where business units can experiment with different AIs using anonymized data.

SECURING THIRD-PARTY AI MODELS

MORBIN: Anonymized data, particularly in medical fields, comes up a lot. How are you preventing sensitive data or proprietary data from leakage and poisoning and tampering? For third-party AI models, how are you providing DLP for AI?

DOTEN: There's a wealth of LLM firewalls. This is ironic – one that keeps coming up, but one that has been two and a half years old and we've figured it out. It's whether everyone's trying to not like the answer to it and say, "How do we block it? How do we control it?" A lot of these LLM firewalls have been acquired in the last two and a half months. But this is a well-understood thing. The larger piece is the stuff that's internal because when setting up the governance, we need to define the types of AI we're using. It's not just a public LLM. It could be an internal model. It could be internal SharePoint. It could be embedded into the SASE tools we're using. It could be embedded into the protection tools we're using. It could be embedded into back-office tools, and it could be agentic.

Each of those have different things we need to account for – the data they access, the types of accounts they use, the ownership of it that we've been talking about. Everyone's still worried about sensitive data going up to the model. That's an older problem. We have all these other ones that have moved on as it's evolved, and we get stuck on that, not to mention our third parties who are using AI and we're sharing data with. It's the same thing. What clouds are my third parties using? Are we scrutinizing them? What AI models are they using? Are we scrutinizing them as well?

MARKARIAN: A lot of our traditional tools are SIEM tools, including your SOC. They've been using AI to some degree for a long time. It's embedded in there.

PASHA: We almost have to differentiate between AI for security and the security for AI discussion. There's a security piece, which everyone's been doing for quite some time. Palo Alto has been there for more than a decade using machine learning and other forms. AI wasn't invented when generative AI came out a couple of years ago. What's interesting today is securing AI. That capability requires not necessarily just the tools.

The tools are important. Things like protecting the data, protecting the applications, protecting the models, protecting the agents now and identifying the agents – all of those are new attack surfaces that we need to think about. But for me, before we talk about the technology, there needs to be a level of strategy and also a level of policy and process that needs to go before that. What kind of AI risk assessments am I able to do? What type of secure AI development life cycle have I been able to build? What type of red teaming can I do? Can I do it in real time?

The third-party model intake stuff that you were talking about is critical. To understand the wider ecosystem is important. When you think about the guardrail policies, all of those come before you start thinking about technology. Yes, it's important, but if you can get those initial points tucked away, the technology is easier to implement.

DOTEN: Technology's always been the easy part. Technology supports a process. In the SDLC, if you have a development team and they're using AI-assisted development, we're now realizing that it's not just third-party code. AI does not code like humans do. It will not just change one thing. It changes everything. It's also not very tight, and it will overdo things, and it may delete things if you don't be on top of it.

Whereas a couple of years ago, or maybe last year when we started doing more AI-assisted coding, and Claude Code and GPT-5 have become better at it, we're realizing this is not going to match the process we had to review things and do code reviews. We have to watch them more closely, and that's been a big challenge and risk for a lot of folks.

AI'S ROLE IN SOC WORKFLOWS AND AUTOMATION

MORBIN: What is the evolving role of AI in SOC workflows, enrichment triage and autonomous remediation? We've been taking a fairly negative aspect of AI. On the positive side, how are we benefiting from it?

JÖRGENS: AI is very strong in volume reduction. For example, you can correlate logs faster than any analyst ever. You can use it to do some kind of work which will not need an expert. It's for summarizing all the details and the information, and an expert will have a look at the results itself.

MARKARIAN: I have a peer who took their incident response playbook for tier 1 and tier 2 and – AI makes the decisions for each process along the way, and then the tier 3 analysts can look at it as necessary. We're not at that point yet, but I definitely see that evolving to it.

DOTEN: I've talked to a lot of startups that are doing that same kind of thing and trying to be autonomous and be able to do it. Again, it's still supporting the same process and it's writing tickets, and humans are defining whether it's an incident or not and things like that. One of the areas that I'm enthusiastic about is automated remediation. We have the tens of thousands of vulnerabilities that we get every single month. We have more than enough things for finding problems. We need help fixing problems. Humans are still fixing problems where we've automated the finding of problems.

We've always been afraid from 20 years ago of IPs taking out network segments of letting machines fix things. But because they can do basic things, let it do all the stuff that IT would need to do anyway. "Is this real? What is the scope of it? Let me make a fix. Let me test a fix. Is it going to break anything? What are the dependencies? Is it going to affect things?" Then have it create the ticket and say, "If I do this, all this stuff will happen. Let me know what to do." There's a lot of opportunity to finally burn down all those vulnerabilities that we have sitting there in the workflow.

PASHA: It's interesting. I've been at Palo Alto for about seven years, and I came from a couple of other big vendors in the industry doing cybersecurity as a specialization and the traditional model of the SOC. I've built a couple of security operation centers myself. I used to walk in and we would see a room filled with people – typically a 30-, 50-, even 100-person security operation center. I came to Palo Alto Networks and walked into the SOC very excited thinking, "Let's look at the people. Let me talk to them." There are literally eight people in the SOC, and I'm scratching my head, going, "Maybe I made the wrong decision here."

When you start to learn about automated remediation, automated threat hunting now, it makes a significant difference to the day-to-day job description of the SOC analyst. You may not need an L1, L2, L3 structure anymore. You may just get junior and senior analysts that do the same job, but you don't need people with eyes on screens 24/7 anymore. You can do with seven or eight people, although the SOC is automated completely.

To share some of our own numbers, our internal SOC is not a managed service; it's our own internal SOC at Palo Alto. We see upward of about 90 billion events on a daily basis, and it all trickles down using AI and machine learning. But ultimately, per analyst, you're only looking at perhaps one incident a day that an analyst has to go in and look at and take action against.

That's a significant cost saving in terms of the time and effort that the analyst has to take. It brings down the mean time to detect and response to about seven minutes and about 60 seconds in order to respond to some of the highest threats that we see. It's there, and a lot of our customers have moved in that direction. But what's important is that you're using AI to help change the job description a little bit, and you're still continuing to retain the talent to do the more important things. This is critical for us. We also don't have tier 1 analysts anymore.

WHERE HUMAN JUDGMENT MUST REMAIN

MORBIN: Where must human judgment remain in the loop? What are those seven and eight people doing?

PASHA: For us, it's the high-impact decisions. Business operations, customer data or regulatory compliance must have some level of human oversight. This is ultimately where the SOC comes in. But again, 90 billion events – the majority of them don't affect most of those things. We are very good at parsing through the data from beforehand. We don't have as many false positives as you would typically find in a traditional security operation center. We're also able to automate threat hunting now. The human for us comes into the loop before we do any threat attribution. It's important for the humans to come in and use their own intuition and their own contextual understanding to apply on whatever AI has found using the automation that it's used.

DOTEN: They need the humans for context, and it's for any type of thing. I have a couple of white papers on AI ethics models, and AI ethics is not the morals and values that we're thinking of, but it's more like, "Is it doing what I expect it to do and did it drift from that?" I want to know when it drifts so I can realign it. Humans need to be the ones to do that because AI will get biased in itself and think that everything is okay. I've written papers on that specific example as well.

There are elements where we need to kind of watch it. There's a part within the workflow that the human makes the decision - is this an event and do we need to go through this when you make a business decision to do something about it? But there are elements also in the business applications, even with the agents or whatever we're implementing, is it doing what I expect it to do, and do I have telemetry to know if it starts to drift to be able to correct it? And then have that workflow to correct it.

PASHA: It's important to know that this is what the state looks like today. We may end up changing this. We may give AI the capability of taking some business operational decisions in the future. This is where having not just the AI council but having a trust council comes into effect as well is crucial, where they can talk about the ethics and the guardrails and all sorts of other things, where eventually AI will be empowered to take action for some of the more high-impact decisions.

JÖRGENS: The humans will still need to take all the responsibility regarding the impact for customers and partners. What we cannot forget is deciding whether an anomalous behavior is malicious or just a rare business exception, which happens sometimes or very often depending on the progress. In the end, we still need humans.

REGULATORY FRAMEWORKS SHAPING AI SECURITY

MORBIN: What are your predictions specifically on regulatory frameworks? How are they going to shape AI security? We've got very different approaches in Europe and the U.S.

DOTEN: We're pretty much having to educate our regulators just as finance did back when there was two-factor authentication being imposed in 2005. I know my friends who did that. Europe, Asia and some countries are looking at it. Australia's looking at it. Some different states are doing different things. A lot of it is around bias and discrimination of the models, which is a place to start. We're going to see a lot of back and forth, and a lot of it's going to be defined by each industry to understand what is possible.

REGULATION AS THE ACCELERATOR FOR TRUST

MORBIN: To what extent is regulation acting as the brakes that allows us to go faster, and to what extent is it hindering innovation?

DOTEN: Everyone keeps bringing that up, and I don't see how that impacts it at all. Formula One cars still need to have brakes. It's like taking the brakes off is not going to make them go faster. In fact, having brakes is what makes them go fast – the ability to control them.

PASHA: I'm based out of Europe or EMEA, and we love our regulations in EMEA. We've got the NIS2 framework. We have GDPR from a data privacy perspective. We now have the EU AI Act as well, which will become global standards for AI risk management and security controls, and it will be adopted worldwide in other places.

My main prediction is there will be certain sectors that will be affected sooner than others – specifically financial services. Financial services, by this year as well as going into 2026, will have very specific AI security requirements because they've been early adopters of leveraging AI for fraud management for more than a decade now. As they start to develop agentic, as they start to use LLMs and all sorts of other functions with the customer data that they hold, which is extremely sensitive, they're going to have probably some of the strictest controls. That's probably my first prediction.

The second, which is mostly inspired from my visit to the Gartner event, is that we will start to see more mandatory AI security audits for critical infrastructure over the course of the next couple of years because we are starting to see early stages of critical infrastructure wanting to leverage some level of AI automation. The mandates and security audits behind that is going to be critical.

AI should not be a black box. From a regulatory perspective, the organizations will need to prepare for what I call explainable AI requirements, where security decisions made by AI must be auditable. They must be justifiable. You must be able to explain them. In certain cases, in a couple of years in the EU especially, you need to be able to do it in real time, not three days later or five days later. You need to be able to show that you've got that capability. From my perspective, those would be some of the predictions.

DOTEN: Regulators don't like black boxes. The interpretability of the models and the transparency of the implementation of those models is something that's going to be important. I agree that we're going to get to the point where we need to be able to offer them in real time to prove that they're not drifting into a biased or discriminatory state. Because we want consistency out of them, which is what they are looking at.

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.

(800) 944-0401 • sales@ismg.io

BANKINFO SECURITY® CUINFO SECURITY® Just for Credit Unions GOVINFO SECURITY® HEALTHCAREINFO SECURITY®

infoRisk® TODAY CAREERSINFO SECURITY® Data Breach. Prevention. Response. Notification. TODAY CyberEd.io

CIO.inc DeviceSecurity.io PaymentSecurity.io FraudToday.io

CYBER THEORY CyberEdBoard Xtra mile LIFECYCLE MARKETING GREYHEAD

iSMG
INFORMATION SECURITY
MEDIA GROUP

902 Carnegie Center • Princeton, NJ • 08540 • www.ismg.io