sasig

REPORT

Adopting AI: crucial security considerations for organisations

In association with

sosafe

Part of Nineteen cyber series

NATIONAL CYBER SECURITY SHOW

bigsasig
6 May 2025 | London

INTERNATIONAL CYBER EXPO
30 SEPT – 1 OCT 2025
OLYMPIA EVENTS, LONDON

sasig
Daily Webinars
Weekly Events

Introduction

Artificial intelligence (AI) is no longer science fiction, rather an everyday reality, with businesses of all shapes and sizes now taking advantage of at least one form and realising benefits as a result.

But as AI abounds, so too do the security considerations surrounding it. Will AI become an unwitting accomplice to cyber criminals, enabling them to more easily penetrate organisations? Or could AI implementations end up becoming the next legacy systems in an organisation's estate, remaining unpatched, untouchable and insecure for fear of disruption?

To learn more about how security professionals are dealing with AI's proliferation within their organisations, Tarquin Foliss OBE, Vice Chairman of The Security Awareness Special Interest Group (SASIG), and subject matter experts from security awareness scale-up, SoSafe, recently hosted an invitation-only roundtable on safe AI adoption and custodianship.

Focusing on governance, risk management and emerging challenges, the roundtable brought together a select group of senior cybersecurity professionals from a range of businesses and critical national infrastructure sectors.

The discussion – held under Chatham House Rule to encourage openness and the sharing of information by providing anonymity to those involved – moved beyond the usual discourse on AI's potential for cyber threats, instead emphasising ownership, ethical considerations and corporate readiness for AI adoption.

Here are some of the insights from the discussion.

AI on the boardroom agenda

Andrew Rose, Chief Security Officer at SoSafe, kicked the discussion off by asking how high the subject of AI risk is on boardroom agendas?

While AI's potential to enhance business processes is widely acknowledged, the associated risk and security implications often lag behind in corporate discussions, indeed these risks often only become visible with time and experience. Some organisations have established dedicated AI governance teams, while others have simply implemented policies restricting the use of generative AI tools to mitigate data security risks.

"The board is becoming curious about AI now, which is a really positive thing," one participant noted, referencing efforts to introduce policies and structured discussions at the highest levels of leadership.

Others pointed out that some boards remain overly focused on the benefits of AI—efficiency, automation and innovation—without adequately weighing the potential downsides such as bias, misinformation and cyber threats. "I'm sure AI functionality is high on the agenda, but how well understood and how highly rated are the related threats and risks?" one participant questioned. Thus there is an argument for developing AI literacy at the executive level to ensure informed decision-making.

The role of governance

The conversation then moved on to the subject of AI governance, with SoSafe's Andrew asking: "If organisations have a suitable data governance strategy in place, do they need a separate AI policy?"

As one participant outlined: "There's definitely an argument to have an AI strategy/policy to act as an educational tool — particularly for the board and executive committee to drive an understanding of existing 'AI', such as Machine Learning, and generative AI, its uses and risks.

This was a suggestion supported by several of those present, with another participant noting that in their role as a CISO, they rely heavily on acceptable use policies and education, as well as implemented guard rails to better manage and mitigate AI related risk.

Another concern was data security, particularly regarding how AI models interact with sensitive information. Some organisations have implemented strict policies limiting AI's access to data, while others have struggled with how to enforce governance. As one participant noted, "Trying to secure third-party supply chains is criminally difficult, and with AI, it's even more so."

This was echoed by another individual who explained how their organisation had to remove access to some AI solutions to prevent a potential data breach and they discovered sensitive company data being pushed out in the wild.

"We've proactively blocked access to external generative AI solutions, like ChatGPT. That's because when we actually examined people's usage of such sites, we found vast quantities of our data being shared, some of which was confidential and restricted," they said.

The participant went on to say that their organisation is now rolling out Microsoft Copilot in a controlled way internally, so as to provide business users with a suitable alternative, one that doesn't risk a daily data breach.

However, another participant explained their approach. "When ChatGPT suddenly emerged, companies were blocking it left, right, and centre… but by the time these things enter the public domain, the horse has already bolted." This was in reference to the fact that, once business users have experienced a nascent technology, attempting to remove access to it will undoubtedly be met with at best discontent and at worst defiance. Staff may look to circumvent restrictions and continue to use banned technologies, which raises questions in terms of security and policy adherence.

The regulatory landscape

SoSafe's Andrew then raised the subject of AI regulation and asked what good AI regulation looks like?

Several participants acknowledged that governments worldwide are beginning to develop regulatory frameworks for AI, with the European Union leading the charge through its AI Act. While participants noted the EU's efforts to develop regulations focused on the ethical and social implications of AI, they expressed concerns about the technical complexity of such regulations. Concerns were raised about whether this type of regulation can keep pace with the rapid evolution of AI capabilities.

However, there remains a significant gap in global coordination. The group noted that the differing priorities from country to country when it comes to AI regulation create challenges for multinational organisations. This fragmentation means that businesses must navigate a complex landscape of compliance requirements, adapting their AI strategies based on regional legal constraints.

There were also comments about the global nature of technology, and that some nations could develop AI faster in an unregulated environment, thereby gaining competitive advantage. Henry Ford did not take time to build a safer car, he mass produced a functional vehicle, and safety and security followed later – that's a model that many technologies have followed. But is it a terribly unwise path for a technology as powerful as AI?

The elephant in the room was America's stance on AI regulation, with several participants saying it was their understanding that the United States wasn't placing much focus on regulating AI, prioritising speed over safety. With China adopting a similar approach, this left European countries stuck in the middle. Can regulation go hand in hand with rapid innovation? The US, for a variety of reasons, has decided not. Meanwhile, China will use regulation to suit its purpose. This is about the ethical development of technology.

Ethical considerations for AI

The participants discussed several key ethical issues around the use of AI. One major concern was the accountability and responsibility for decisions made by AI systems. There was uncertainty around who would be held accountable - the AI developers, the product owners, or the end users. As one participant noted: "It's not just about blocking access to AI; it's about educating employees on its responsible use." It was considered essential that each AI process had an accountable owner, and that the absence of such was a clear indication of glaring gaps in governance and oversight.

Another ethical issue raised was the potential for AI to be used to create deepfakes and disinformation campaigns. Participants noted the difficulty in regulating this, as social media platforms are often reluctant to take down content, even when it is known to be false or misleading. There were concerns about the impact this could have on elections, public discourse and corporate reputations. "Organisations must be mindful of how AI is used to generate content and ensure there are safeguards in place to verify accuracy," one participant noted. Easier said than done, but having a pre-prepared process to address deepfake attacks on your brand was seen as essential.

The use of AI in sensitive areas like journalism and national security also raised ethical red flags. One participant from a media company highlighted how they need to rely on human intervention to ensure accuracy and impartiality in their reporting, rather than fully trusting AI-generated content. Similarly, the use of AI in military targeting was seen as highly problematic, with the potential for unintended civilian casualties. For critical services, the group did not think we were anywhere near ready to take our 'hands off the wheel'.

AI adoption & the pivotal role of data

When it comes to adopting AI, the importance of good data should not be overlooked. As SoSafe's Andrew highlighted: "Oftentimes, organisations' data is disorganised and inaccurate, held in unstructured locations with irregular labelling and classification. Yet it's this data that companies are inevitably using to train AI models. This increases the risk of toxic and biased output," he said.

Andrew also highlighted that, while much of the data being used to train AI models now has been created by humans, there's a good chance that the majority of it going forwards will have been created by AI. We'll then have a situation where AI is being trained on its or another AI solution's output. "How will innovation thrive once we get to this stage?" Andrew asked the group.

Data as a primary consideration ahead of AI adoption was a theme that several participants empathised with. One participant explained how their organisation's AI-powered productivity assistant inadvertently granted unauthorised access to sensitive files due to improper data tagging, reinforcing the need for faultless classification and permissions management.

"A lot of the success associated with AI comes down to how your data is structured and what sort of classification you have in place," one participant said.

"As we were rolling out Copilot with a small user base, we found that people were able to prompt it to return information held in another user's SharePoint drive. It was retrieving classified and confidential information from another person's space because that data had not been appropriately tagged," they added.

"We then had to undertake an enormous project to boost our DLP [data loss prevention] and information classification across the organisation to enable us to get our data — both structured and unstructured — in a state where we were actually ready for adoption."

One issue was noted in that many developers allow the AI access to all data and do not recompile each request using the permissions of the user asking the question. This enables users to ask innovative questions to bypass safety controls and access sensitive data.

Thomas Owen, Chief Information Security Officer at SoSafe raised another key concern: data poisoning— the risk that malicious actors could manipulate training datasets to produce skewed AI outputs. As one participant responded: "We already have data integrity issues without AI. My biggest risk is not AI itself but how my third parties develop those products," highlighting the challenge of securing AI-driven processes managed by external vendors.

This underlines the pivotal need for organisations to have human checking built into their processes, to ensure accuracy of AI outputs. However, as SoSafe's Andrew pointed out: "What happens when people get lazy? It's too easy for someone to check an AI model's output 10 times and assume that the 11th time and 12th time and so on will be correct too."

Christine Siu, VP of Product at SoSafe, agreed, stating that "AI is a means to an end at the end of the day. Therefore, you must be comfortable with the output you get from the AI, and be ultimately responsible for it."

Concerns were also raised about AI becoming legacy technology too quickly, particularly as organisations embed AI into core operations. Participants debated the potential for AI-driven decision-making models to become too entrenched to modify, leading to security blind spots.

As SoSafe's Andrew highlighted: "You get AI embedded into your processes and then there is a risk that you don't touch it because it's working and it's giving you the right output. If you then go and change the data set, add new inputs, update the decision model, or even change the platform, you risk losing the reliability you've built."

The reliability of interconnected AIs was also raised as a concern, specifically whether organisations would be able to understand the decision complexity they would have created. As one AI feeds off the output from another AI's results, the potential for catastrophically poor outcomes escalates. The good practice lessons learned about the value of network segregation in critical systems, to minimise the chance of systemic failure, seem to apply here.

The human element

Finally, the discussion focused on the human element of AI and the importance of educating and informing individuals about both the benefits and drawbacks of using AI.

There's frequently a perception among workers that AI might end up replacing them, which ensures its implementation is often poorly received. One participant's advice was to take an incremental approach to AI and build trust over time, both for the benefit of the users, and the resilience of the process. However, another individual pointed out that AI is actually creating lots of roles and opportunities for careers.

Someone else outlined the importance of highlighting the different types of AI that are out there and the different problems they can help solve. "If people don't fully understand an AI tool, or any tool for that matter, they can go off down the wrong track and waste a significant amount of time," they said.

"Before an organisation looks to adopt AI, they first need to fully understand their own business and the outcomes they want to drive," another added.

SoSafe's Andrew suggested that AI awareness might form part of an organisation's new user training, ensuring that people new to the company start off on the right foot when it comes to utilising AI.

One participant highlighted that their organisation had already adopted such an approach. "We have incorporated AI training into our organisation, which outlines both the ways in which you can use AI, as well as the ways you shouldn't use it," they said. "We also have usage policies in place so we can hold people to account when they don't use AI correctly."

Several present questioned what AI's impact on human creativity would be. "AI has been developed in order to enhance creativity, and it's going to do that in some aspects, like with the ability for technology to do very specific, cutting edge activities. But on the human aspect, we're in danger of becoming less creative, and that's not just in terms of critical thinking, but we're also going to become less able to be individual and unique in our thoughts," one noted.

The SASIG's Tarquin followed this up by touching on skills, highlighting that we need to determine what we'll need going forwards and strive to keep people engaged in the overall process. "Otherwise, we risk losing the capacity to develop and grow the experts of tomorrow," he said.

AI in cybersecurity: opportunities and risks

The roundtable discussion highlighted the growing role of AI in cybersecurity, with organisations leveraging AI-driven tools to enhance threat detection, automate responses and strengthen defenses. However, the discussion also revealed concerns about adversarial AI, data security and the challenges of integrating AI securely within existing infrastructures.

Several participants shared that their organisations are already utilising AI for cybersecurity, either through in-house development or via third-party vendors. One noted that their organisation employs AI to “detect attacks in their networks,” while another highlighted the reliance on AI-powered security tools provided by major partners. Despite this growing adoption, there was general agreement that AI remains a supplementary tool rather than a standalone solution.

However, AI is not just being used for defensive measures, with attackers also utilising AI to enhance their methods and effectiveness. Participants warned of AI-generated phishing campaigns and AI-driven automation in hacking attempts. “AI isn’t just a defensive tool; attackers are using it to create more sophisticated attacks,” SoSafe’s Andrew emphasised. This has led organisations to re-evaluate their cybersecurity strategies to stay ahead of AI-enhanced threats.

Final thoughts

The extremely insightful roundtable discussion underscored the need for a proactive approach to AI governance. It proved essential that organisations must align AI adoption with robust data governance, security strategies and ethical frameworks - and they need to regularly reiterate these controls to stay alongside the development of both AI capabilities, and the emergent risk factors.

While AI presents transformative opportunities, businesses must navigate its risks carefully to ensure long-term sustainability and security. The discussion highlighted the dual nature of AI—it is both a powerful enabler and a potential risk factor. Striking the right balance between innovation and responsibility, and keeping on top of the emerging risks, will be key to harnessing AI's full potential in the years to come.

Essential advice

Governance

1. Governance is critical - set that up early with guidelines for adoption, usage, & tracking of AI dependencies.
2. Identify the smallest number of 'cannot fail' areas and focus most of the AI oversight on those.
3. The code of ethics, privacy requirements etc need to be constantly reviewed and updated.
4. Regulation to limit social harm does not necessarily equal constraint on innovation or value generation, and any push back on this is suspicious or an indication of a lack of understanding.
5. Track & manage the complexity you'll be creating when AIs start to interact.

Adoption

1. Ensure AIs have accountability and human checking built into their processes - if you cannot identify a meaningfully accountable and enabled party, you've gone too far.
2. Take an incremental approach - optimise and augment, build trust at one level before moving on; speed brings risk, and you'll move faster this way.
3. Manage your data - Invest in data classification, infrastructure and governance early!

Human

1. Teach people about the benefits and drawbacks of AI, make them AI literate and aware and aware of the potential risks of AI, and the attacks that can be enabled with AI.
2. Give confidence to the people within your organization who want to lead in this space.
3. Consider the human impact. We live within the confines of the information available to us, consider how AI will change, lessen or expand that bubble for your staff.

About SASIG

The Security Awareness Special Interest Group (www.thesasig.com) was founded in 2004 as a safe, trusted environment for those interested in or responsible for the cyber resilience of their organisation to meet, share knowledge and expertise, and exchange experiences without fear of a hidden agenda or sales pitch. Membership now numbers c11,000 representing 4,500+ organisations from government, the private sector, law enforcement, academia and non-profits – mainly UK-centric but increasingly global.

SASIG hosts a variety of webinars each week, several in-person events each month and our flagship Big SASIG conference each year (www.bigsasig.com), all free for members to attend.

About SoSafe

SoSafe is Europe's leading security awareness training and human risk management platform. It assists organisations in reducing human risk by applying behavioural science to enhance employees' security instincts. With a focus on delivering effective, engaging, and measurable training, SoSafe enables businesses to mitigate risks, foster a culture of security, and defend against emerging cyber threats.

Learn More

To discover how SoSafe can help your organisation build a stronger human firewall and enhance its cybersecurity resilience, visit their website at sosafe-awareness.com sasig

REPORT

Adopting AI: crucial security considerations for organisations

In association with

sosafe

Part of Nineteen cyber series

NATIONAL CYBER SECURITY SHOW 8-10 APRIL 2025

bigsasig 8 May 2025 | London

ICE INTERNATIONAL CYBER EXPO 30 SEPT – 1 OCT 2025 OLYMPIA EVENTS, LONDON

sasig Daily Webinars Weekly Events