WHITEPAPER
AI and cybersecurity

AI and cybersecurity

A double-edged sword

In recent years, Artificial Intelligence (AI) use has surged. Generative AI, in particular, has stormed into the public and business consciousness in a manner few could have foreseen. Exciting possibilities are emerging, but critical security questions are arising. Many organisations adopt AI faster than they can update their threat landscape, security measures and policies.

On top of that, unhindered by regulatory and process constraints, cybercriminals rapidly embrace AI to boost their abilities to perform complex cyberattacks. And with data flows growing exponentially, exacerbated by the IoT and the explosion in generative AI use, cybersecurity teams face unprecedented challenges.

As a result, the demand for cybersecurity expertise is increasing, but the supply of qualified professionals is shrinking. This scarcity isn't limited to fresh talent; even experienced cybersecurity experts are under immense pressure, leading many to leave the industry. AI is critical in helping cybersecurity teams rise to this dynamic challenge.

AI has already transformed cybersecurity areas like malware detection and safe browsing. AI-driven systems can swiftly identify and quarantine suspicious files, protecting networks and data. With nearly seven in 10 organisations planning to use generative AI for cyber defence[4] the technology looks set to revolutionise the field further, enabling better threat detection and greatly enhancing team productivity.

$60.6 billion projected size of global AI in cybersecurity market by 2028[1]

329 million terabytes of data generated each day[2]

+40% faster phishing attacks powered by generative AI tools[3]

The world is how we shape it

sopra steria

AI-driven risks

Understanding cybercrime 2.0

AI can help cybercriminals scale up, augment their skills, operate in multiple languages, and attack in new and unpredictable ways. With AI in their arsenal, criminals can increase attack volumes and scan for more vulnerabilities. They can also create new vulnerabilities by manipulating AI models and using deepfakes to trick people into making mistakes.

There are three categories of AI-driven cybercrime. The first is AI-powered attacks. These are attacks where cybercriminals use AI technology to their advantage. The second is AI theft. These attacks target intellectual property like AI models and training sets within organisations. The third is attacking AI. This entails manipulating AI systems to skew their results or operations.

AI-powered attacks

Cybercriminals exploit AI's capabilities to manipulate people and technology for their nefarious ends.

Increasingly realistic (spear) phishing

The days of generic phishing emails riddled with spelling errors are gone. Thanks to large language model technology, criminals combine public and stolen data to generate sophisticated spear-phishing campaigns.

These campaigns flawlessly impersonate colleagues, acquaintances, and trusted institutions to target specific individuals and organisations. Phishing is still the leading infection vector in all cyberattacks, identified in 41% of incidents.[5]

Deep fakes

Adding new dimensions to phishing threats, deep fake technology is becoming increasingly sophisticated. Algorithms ‘learn’ the patterns and characteristics of a person’s voice, looks, and personal style and mimic them quickly in new content.[6]

So, imagine that 'your colleague' calls you and asks you to grant them access to a sensitive folder, or make an urgent payment. Microsoft’s VALL-E AI software, for example, needs just a three-second audio clip to clone your voice.[7] ‘Novel social engineering’ attacks, which include ‘CEO fraud’ rose by 135% in the first two months of 2023.[8]

Malicious code and automated malware

Cybercriminals use AI to spread infected software and generate ever-harder-to-detect malware. Researchers created an AI-generated malware called BlackMamba, which bypassed cybersecurity technologies, including industry-leading endpoint detection and response tools.[9] The extent to which such threats exist in the wild is difficult to determine. Yet, the example underscores the profound transformation of the threat landscape.

AI theft

With 73% of organisations having ‘hundreds or thousands of models,’ AI theft is a growing concern.[10] Unauthorised acquisition and misuse of these assets can jeopardise competitiveness, intellectual property, and security.

Model inversion attacks on training data

Training data is the lifeblood of AI models. In model inversion attacks, cybercriminals exploit vulnerabilities and weaknesses in AI infrastructure to reconstruct the samples used to train synthetic models. Consequences can be dire, compromising the integrity of AI systems and laying bare the foundation upon which AI-driven insights and decision-making rely.[11]

Machine learning model theft

Criminals gaining unauthorised access to machine learning models can alter their functioning, steal intellectual property, and erode trust in their reliability. Even apparently well-protected systems may not be immune to theft. In ‘black box' test simulations, researchers could replicate the target systems with relatively low resources. This is a growing concern as more organisations seek to monetise their data and AI models to create new revenue streams.[12]

A shifting landscape

In the past, we fortified our digital 'castles,' built tall walls, and deployed guards to protect our data and systems. However, with the rapid growth of data and AI, today's cybersecurity landscape is multidimensional. Cybersecurity resembles an air traffic controller's role more than soldiers defending a castle. In this evolving landscape, cybersecurity's role isn't just defence; it's about ensuring the safe passage of data to its intended destination.

Attacking AI

Criminals manipulate AI system vulnerabilities and compromise their functionality for malicious uses. Two in five organisations report having faced AI breaches, one in four of which were malicious.[13]

Prompt injection

Criminals use AI to sneak harmful prompts into AI systems to generate malicious content or act unpredictably. For instance, security researchers tricked some of the most well-known, openly available chatbots into behaving destructively, such as requesting users’ bank account details. Such attacks use concealed information, including hidden instructions on web pages, to steer AI systems off-course.[14]

Data poisoning

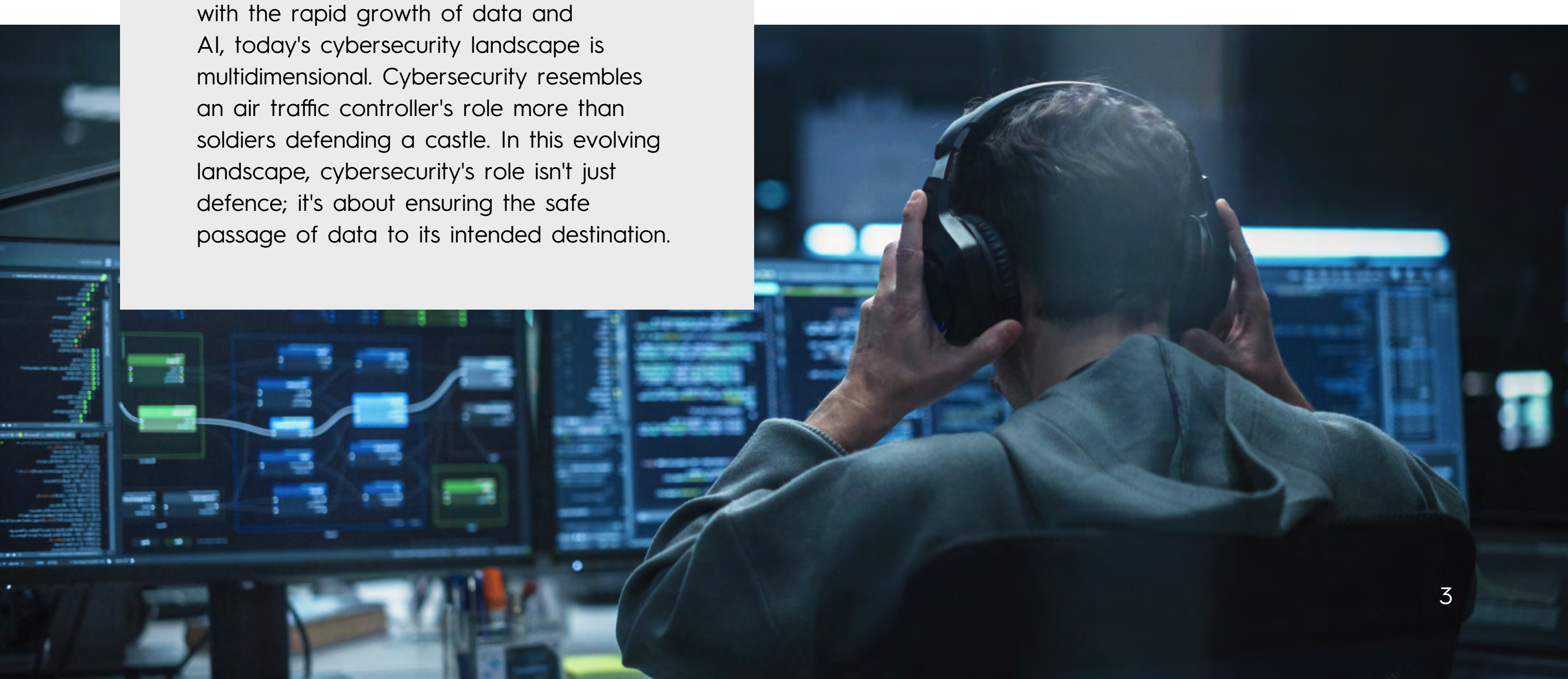
Data poisoning is an increasingly prevalent threat that targets AI-based technologies like autocomplete, chatbots, and spam filters.[15] It involves injecting tainted data to compromise systems’ reliability and yield misleading results, from skewed product reviews to disinformation.

Nefarious models

The very tools that can be employed for good can also be abused. Models can be meticulously fine-tuned for specific tasks or domains, opening up the possibility for malevolent applications. The misuse of large language models and the emergence of new models intentionally designed for malevolent purposes, such as FraudGPT and WormGPT, may erode trust in digital communication and sow social discord.[16]

Compromised data confidentiality

AI systems may inadvertently expose confidential information. Complex solutions involving multiple stakeholders and third-party technologies amplify these risks. Even with anonymisation, everything from political affiliations[17] to sexual identities[18] has been released.



Seizing the opportunities

Driving cybersecurity preparedness

While the threats of cybercrime 2.0 are alarming, AI's adoption among cybercriminals resembles that of society in general.[19] Barriers to entry are still relatively high for criminals. Acquiring malicious code and automated hardware still requires substantial time, effort, expertise, and significant financial resources.

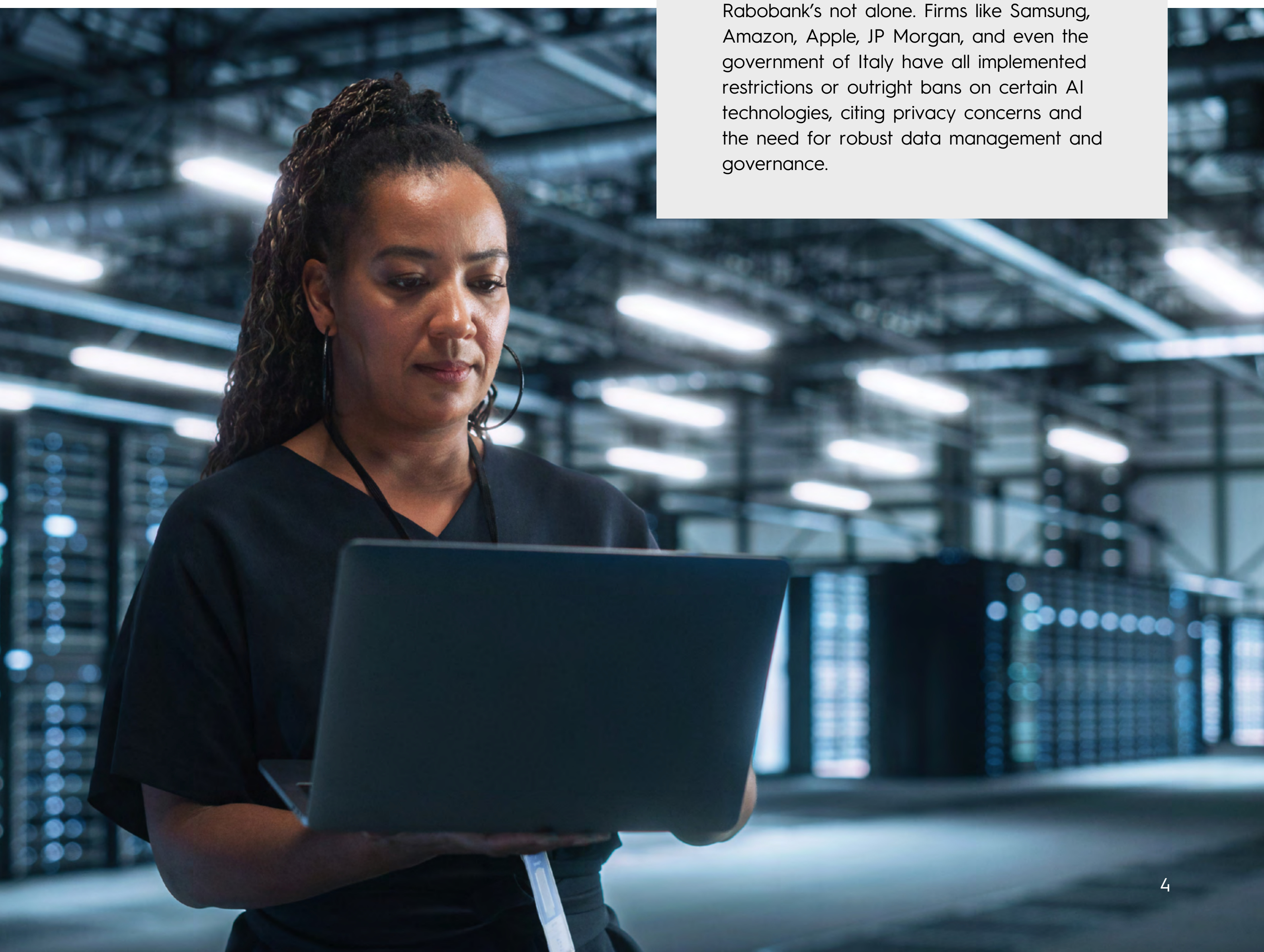
Boosting human efficiency

In the future, however, AI-driven tools may empower even those with basic programming skills to engage in ever-more sophisticated attacks. Adopting AI in cybersecurity is one crucial way leaders can prepare for an uncertain future.

Another advantage of embracing AI in cybersecurity is mitigating skilled staffing challenges. Research shows that using AI and automation extensively in cybersecurity can reduce human effort while simultaneously enabling teams to process more data for deeper real-time insights. It helps cybersecurity teams combat the 33% of wasted time on false positives by deploying AI-based automation for routine tasks, freeing cybersecurity experts to focus on exceptions.

Data governance as the foundation

As other firms raced to deploy AI solutions, Rabobank set a powerful example in data governance. They took the bold step of temporarily pausing AI development, recognising that harnessing AI's potential securely requires a strong data foundation.[20] Rabobank's not alone. Firms like Samsung, Amazon, Apple, JP Morgan, and even the government of Italy have all implemented restrictions or outright bans on certain AI technologies, citing privacy concerns and the need for robust data management and governance.



> "In today's tech-driven world, cybersecurity is about so much more than advanced tools. It's about fostering a culture that values human insight, technology, AI, and responsible data governance. A holistic view is critical to empowering our teams to protect our digital assets today and in the future."
>
> **Mohamed Kassttar**, Chief Information Security Officer

How to use AI in cybersecurity

Nurture governance, safeguard IP, and empower people

AI-powered defences are revolutionising the fight against cybercrime. But the synergy between technology and engaged workforces truly fortifies your defences. As you embrace AI in cybersecurity, focus on three key elements: asserting strong governance, safeguarding intellectual property, and raising awareness among your workforce.

Establish strong AI governance

Maintain control over AI within your organisation through robust governance.

Set clear guidelines
Formulate company-wide policies, procedures, and guidelines governing the use of AI models, training data, and ethical considerations. Consider AI deployments' intended outcomes as well as potential security consequences.

Prioritise effective data management
Implement effective data management, spanning data collection, storage, and monitoring. Collect and retain only essential data (data minimisation) and anonymise all sensitive information.

Create a diverse AI governance board
AI systems may make unfair and inaccurate decisions due to biased training data or algorithms, diverting attention from real security issues. An AI governance board with diverse expertise can mitigate potential discriminatory outcomes and unintended consequences.

Navigate regulatory complexity
Assess AI systems for compliance with evolving regulations, such as the GDPR and the upcoming AI Act in the EU. Allocate adequate resources to pivot dynamically to new rules.

Build awareness and training

Make it a priority to help everyone in your organisation understand AI's role in security.

Augment and empower cybersecurity teams
Equip cybersecurity staff with real-time insights around anomaly detection and pattern recognition. Explore the possibilities of deploying chatbot-like interfaces to empower your teams in conducting remediation actions. Remember to prioritise privacy evaluations when implementing these tools.

Level-up awareness
Empower staff with the knowledge and tools to recognise and counteract emerging AI-driven threats, including sophisticated ones like deepfakes. Implement targeted training backed by consistent communication and engagement initiatives.

Empower your cybersecurity team
Offer continuous training opportunities to your cybersecurity professionals. Help them build skills in using AI-based cybersecurity tools and responding to AI-driven cybersecurity threats.

Protect your systems and intellectual property

Your organisation's defence strategy must be as dynamic as the threats it faces.

Incorporate AI into product development
AI tools can potentially spot vulnerabilities much faster than humans. Embracing AI-driven cybersecurity at the product development stage enhances code quality, reducing the risk of security gaps. Generative AI can improve testing by simulating potential threats like malware.

Harden infrastructure to evolving threats
Go beyond multi-factor authentication and leverage AI for user and entity behaviour analytics (UEBA) to enhance pattern detection, especially for suspicious login attempts. AI can proactively manage vulnerabilities and respond to incidents in real time, minimising the need for human intervention and thus reducing incident response times.

Strengthen threat intelligence
Invest in robust AI-powered threat intelligence capabilities. Combine third-party tools and industry-specific resources with expert and end-user awareness. Consider integrating AI for incident triage, log analysis, and threat detection tasks.

AXA prioritises data security in AI

Global insurance leader AXA launched a secure, internal generative AI service developed in collaboration with Microsoft. The AXA Secure GPT platform prioritises data security and privacy, ensuring employees can harness AI's transformative capabilities without compromising confidentiality. AXA's commitment to responsible innovation joins several other forward-thinking companies on this secure path to AI integration.[21]

"AI is a powerful cybersecurity tool but should not replace human expertise and vigilance. We will always need smart people in a field dealing with deception, deceit, and highly nuanced real-world scenarios. Over-reliance on technology, on the other hand, will always fall short."

Paul Verhaar, Practice Lead Artificial Intelligence

Securing the future

Fusing AI and human ingenuity

The convergence of AI and cybersecurity offers exciting opportunities. But balance and responsibility are essential. Prioritising accountability and ongoing monitoring is crucial for maximising the benefits of AI in cybersecurity while also minimising its possible drawbacks.

As we embrace AI-driven cybersecurity, we're not just adapting; we're building a more secure digital future. Combined with human insight, AI has unveiled tremendous potential, empowering us to proactively protect, precisely detect, and rapidly respond to threats.

In a world where data is gold, our mission extends beyond protecting organisations; it's about safeguarding what matters in our interconnected world.

Sources

[1] Markets & Markets, 2022: *Artificial Intelligence in Cybersecurity;* https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-security-market-220634996.html?utm_source=Globenews&utm_medium=referral&utm_campaign=paidpr

[2] Exploding Topics, 2023: *Amount of Data Created Daily;* https://explodingtopics.com/blog/data-generated-per-day

[3] SoSafe, 2023: *One in five people click on AI-generated phishing emails;* https://sosafe-awareness.com/company/press/one-in-five-people-click-on-ai-generated-phishing-emails-sosafe-data-reveals/

[4] PwC, 2023: *2024 Global Digital Trust Insights;* https://www.pwc.com/gx/en/issues/cybersecurity/global-digital-trust-insights.html

[5] IBM, 2023: *Security X-Force Threat Intelligence Index;* https://www.ibm.com/reports/threat-intelligence

[6] The Verge, 2023: *Here are some of the ways experts think AI might screw with us in the next five years;* https://www.theverge.com/2018/2/20/17032228/ai-artificial-intelligence-threat-report-malicious-uses

[7] Tech Monitor, 2023: *Microsoft's new VALL-E AI can clone your voice from a three-second audio clip;* https://techmonitor.ai/technology/ai-andautomation/vall-e-synthetic-voice-ai-microsoft

[8] Technology, 2023: *Scam email cyber attacks increase after rise of ChatGPT;* https://technologymagazine.com/articles/scam-email-cyberattacks-increase-after-rise-of-chatgpt

[9] Dark Reading, 2023: *AI-Powered 'BlackMamba' Keylogging Attack Evades Modern EDR Security;* https://www.darkreading.com/endpoint/aiblackmamba-keylogging-edr-security

[10] Gartner, 2023: *Don't Let Your AI Control You: Manage AI Trust, Risk and Security;* https://www.gartner.com/en/newsroom/press-releases/2023-06-07-gartner-security-and-risk-management-summit-national-harbor-day-3-highlights

[11] Mostly AI, 2023: *Model inversion attack;* https://mostly.ai/synthetic-datadictionary/model-inversion-attack

[12] Unite AI, 2023: *Stealing Machine Learning Models Through API Output;* https://www.unite.ai/stealing-machine-learning-models-through-api-output/

[13] Gartner 2022: *Gartner Identifies Top Five Trends in Privacy Through 2024;* https://www.gartner.com/en/newsroom/press-releases/2022-05-31-gartneridentifies-top-five-trends-in-privacy-through-2024

[14] Wired, 2023: *Generative AI's Biggest Security Flaw Is Not Easy to Fix;* https://www.wired.com/story/generative-ai-prompt-injection-hacking/

[15] Technology, 2022: *Data poisoning threatens to choke AI and machine learning;* https://technologymagazine.com/articles/data-poisoningthreatens-to-choke-ai-and-machine-learning

[16] Silicon Republic, 2023: *WormGPT and FraudGPT: The dark side of generative AI;* https://www.siliconrepublic.com/enterprise/wormgpt-fraudgptchatgpt-generative-ai-cyberattacks

[17] Techcrunch, 2019: *Researchers spotlight the lie of 'anonymous' data;* https://techcrunch.com/2019/07/24/researchers-spotlight-the-lie-ofanonymous-data

[18] Wired, 2010: *Netflix Cancels Contest After Concerns Are Raised About Privacy;* https://www.nytimes.com/2010/03/13/technology/13netflix.html

[19] Google Cloud, 2023: *Why AI: Can new tech help security solve toil, threat overload, and the talent gap?;* https://cloud.google.com/blog/ transform/why-ai-can-new-tech-help-security-solve-toil-threat-overloadand-talent-gap

[20] Management Scope, 2023: *Bart Leurs (Rabobank): 'We implemented a temporary AI stop;'* https://managementscope.nl/en/magazine/article/5337-bart-leurs-artificial-intelligence-risks

[21] AXA, 2023: *AXA offers secure Generative AI to employees;* https://www.axa.com/en/press/press-releases/axa-offers-securegenerative-ai-to-employees

Talk to us
Learn how we can enhance your cybersecurity processes and elevate your security posture, including through AI integration: www.soprasteria.com.

The world is how we shape it

Sopra Steria, a major Tech player in Europe with 56,000 employees in nearly 30 countries, is recognised for its consulting, digital services and software development. It helps its clients drive their digital transformation and obtain tangible and sustainable benefits. The Group provides end-to-end solutions to make large companies and organisations more competitive by combining in-depth knowledge of a wide range of business sectors and innovative technologies with a fully collaborative approach. Sopra Steria places people at the heart of everything it does and is committed to putting digital to work for its clients in order to build a positive future for all. In 2023, the Group generated revenues of €5.8 billion.

Sopra Steria (SOP) is listed on Euronext Paris (Compartment A) – ISIN: FR0000050809. For more information, visit us at www.soprasteria.com

sopra steria