zühlke
empowering ideas

Securing Artificial Intelligence

A CISO perspective across industries

Zühlke August 2024

Executive summary

Zühlke has interviewed CISOs from small to large organisations across different industries to understand their concerns and ongoing efforts to secure the use of Artificial Intelligence (AI) in their respective organisations. The interviewees come from both the public and private sectors, including market-dominating organisations operating in the DACH region and beyond.

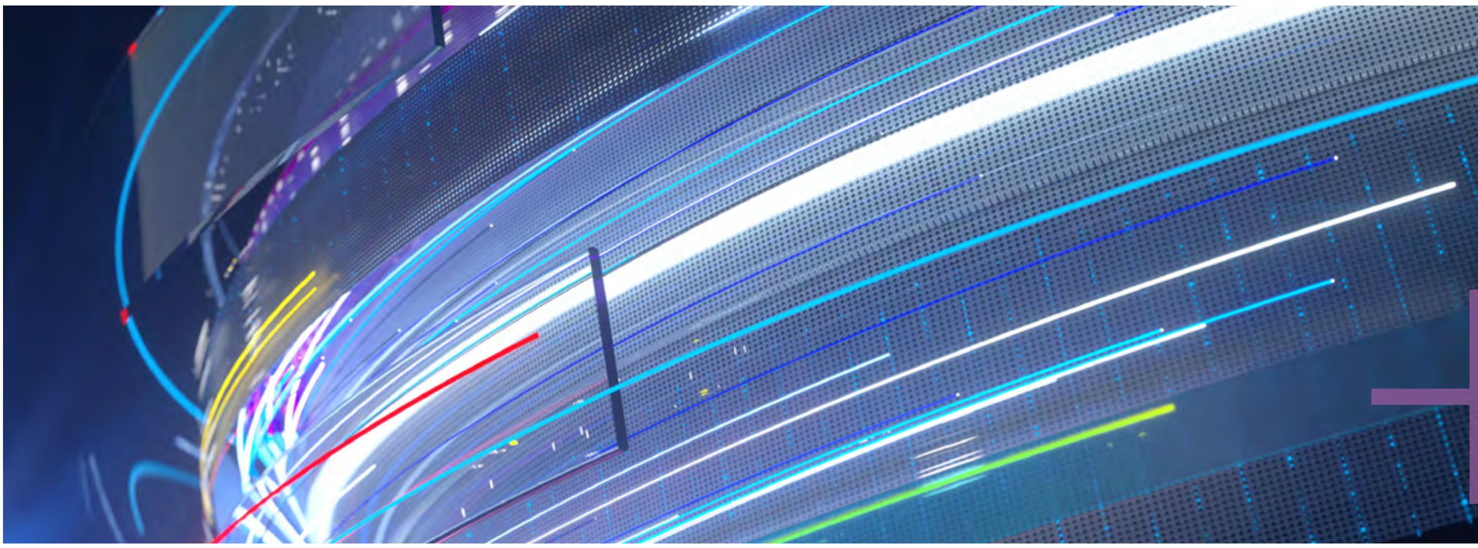
Five key findings were identified, with the following main concerns:

- Insufficient maturity of available standards and best-practice guidelines for securing AI systems.
- Pressure from business leaders to adopt AI as quickly as possible.
- Lack of control over complex AI SaaS solutions.
- Expected surge in attacks fuelled by the rapid adoption of AI by cybercriminals.
- Burdening business processes with legal over-regulation, eventually benefiting the attackers.

To counter these concerns, professionals are actively engaging with AI to gain experience, and organisations are investing in the education of specialists and the general employee base. Across industries, community and knowledge sharing structures have now been established. More measures for countering the concerns are listed in the recommendations section on page 16.

About this study

Focus	CISOs & security professionals
Sectors	Banking, insurance, industrial sector, medical device and healthcare, pharmaceutical industry and life science, commerce and retail, energy and utilities, government and public sector, transportation, and leading cloud providers
Countries	Germany, Austria, Switzerland
Publication date	July 2024
Survey mode	22 personal 45-minute interviews with prepared questions



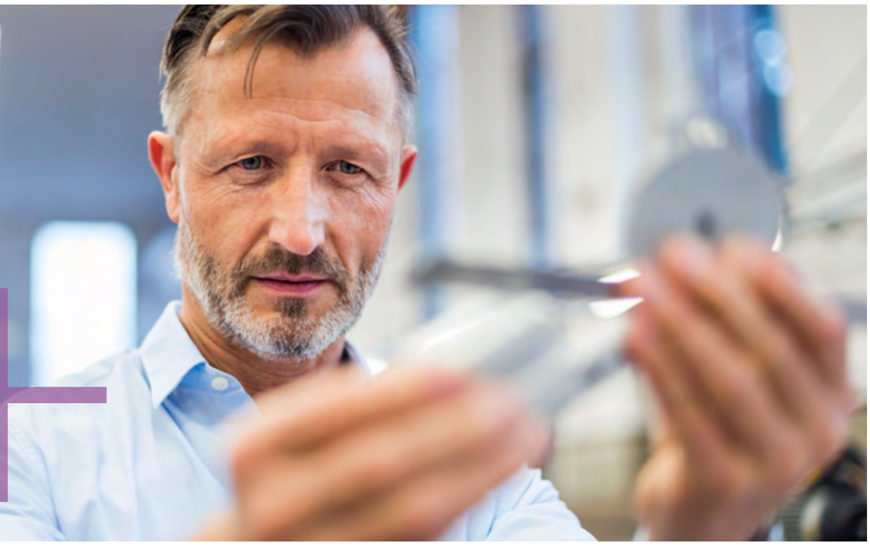
A study across industries

Zühlke has interviewed CISOs from small to large organisations across different industries from both the public and private sectors, including market-dominating organisations in the DACH region and beyond.

The driving questions were:

- When confronted with the risks of AI, what are the major concerns of the experts responsible for the security of their organisation?
- What is the maturity level of the organisations when it comes to the security of AI?
- What are the strategies applied to proceed?

This report focuses on securing AI systems from a CISO perspective and on a governance level, while only touching briefly on the discussion around AI-powered defence and the development of secure AI. We give an overview of the survey methodology and the issues, discussion scope and definitions identified, before covering the seven key findings in detail. Finally, we conclude the report by providing several recommendations.



Motivation: the rise of AI

Artificial Intelligence (AI) systems are progressively being used beyond their original realm of data science and are now quickly finding their way into domains where their power to automate is poised to increase efficiency and efficacy. AI systems, either locally deployed, as a third-party service, or in the hands of cybercriminals, are likely to change the cybersecurity landscape fundamentally.

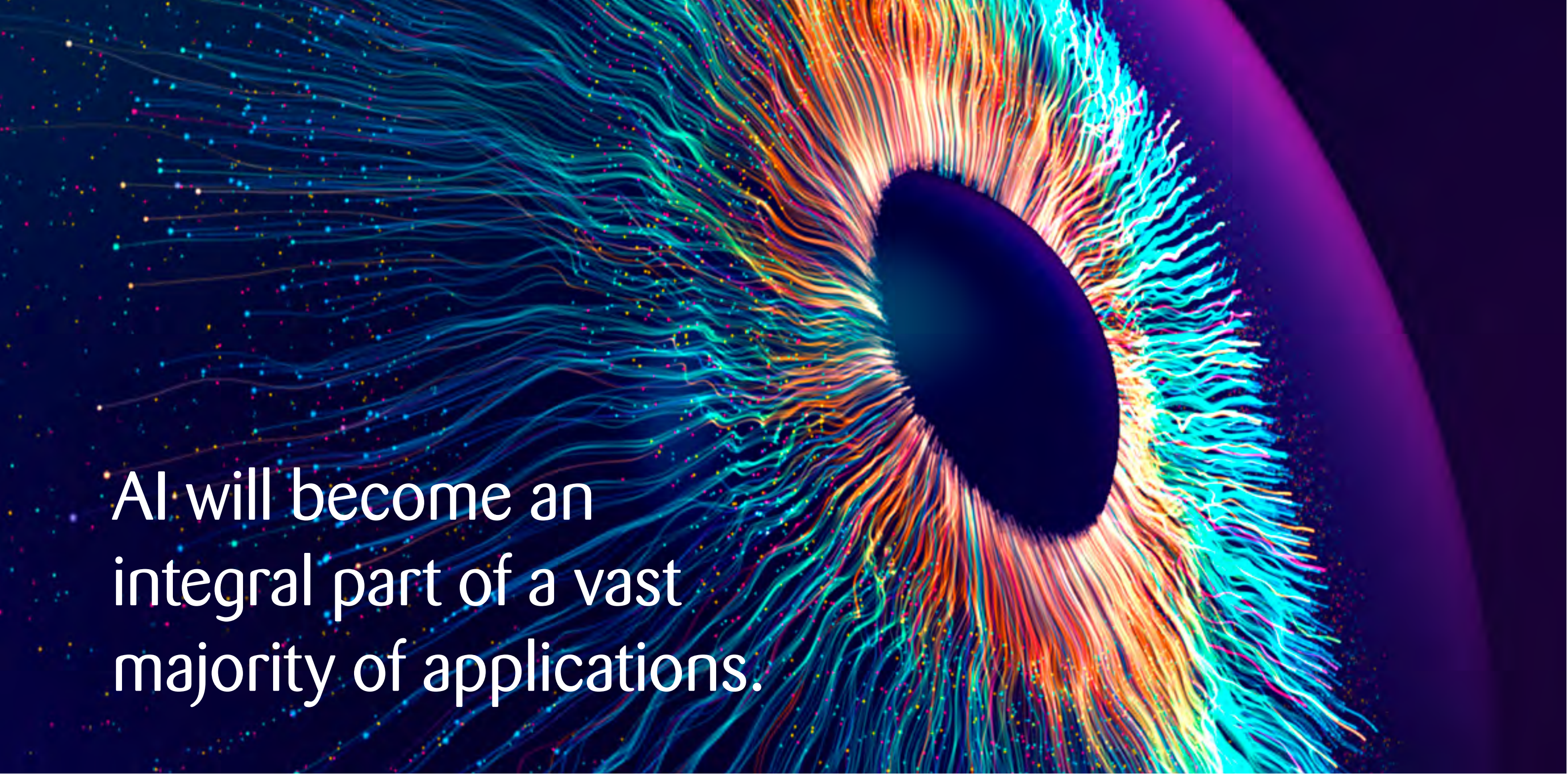
The methodology

The study was conducted in 2024 in the form of anonymous 45-minute interviews with CISOs and experienced security experts, covering their view on the effect of AI on the cybersecurity risk landscape, through the position of their organisation, to the forthcoming regulations. We interviewed experts from several sectors including banking, insurance, industrial sector,medical device and healthcare, pharmaceutical industry and life science, commerce and retail, energy and utilities, government and public sector, transportation, and leading cloud providers.

Just a new technology?

Five key findings have been identified, with the following main concerns:

- The insufficient maturity of available standards and best-practice guidelines for securing AI systems.
- The push from business executives to adopt AI as quickly as possible.
- The lack of control over complex AI-SaaS solutions.
- The expected surge in attacks fuelled by the rapid adoption of AI by cybercriminals.
- The risk of burdening business processes with legal over-regulation, eventually benefitting attackers.



AI will become an integral part of a vast majority of applications.

While most interview partners agreed that AI is **'just a new technology'** that is relatively easy to manage, they admit that the speed at which it is being introduced is quickly outpacing their ability to keep up with the increased risk posed by such tools. They acknowledge that new concrete threats are already emerging, such as authentic looking spear phishing attacks enabled by AI. The balance between defender and attacker is likely to shift towards the attacker, at least in the short term.

Scope: focus on risk minimisation

Before considering the results in more detail, we revisit the term Artificial Intelligence, abbreviated to AI. Nowadays, AI has mostly become a marketing term, encompassing aspects from classical machine learning through to the personification of the idealised human being. There is consensus that the dominance of such automation methods is increasing at a very high pace, and that AI will become an integral part of a vast majority of applications.

The study participants separated the discussion on the **adoption of AI for cybersecurity** into various areas, of which we focus on **risk minimisation**.

This includes the following topics:

1. Defence against AI-powered attacks.
2. Securing AI systems once deployed.
3. Defence against careless use of AI systems. [1]

Other AI topics in cybersecurity that do not fit under the umbrella of risk minimisation, and that are not therefore covered in this report, are a) the secure development of AI and b) the use of AI for defensive purposes. The first concerns all security aspects pertaining to company-internal development of AI tools. These include properties such as robustness, bias, diversity, etc., which are mostly seen as the responsibility of the relevant developers and less of the CISO. The second relates to AI-powered software to improve cybersecurity defence, such as automated anomaly detection, which usually consists of products or services bought from external suppliers.

By comparison, the Swiss Cyber Defense Campus has published a collection of essays by academic experts on LLMs in cybersecurity [2] which covers LLMs in more detail. The Swiss federal institute of technology (ETH Zürich), in cooperation with Swissmem and Netx Industries, has examined the general state of AI in Swiss industry. [3]

[1] This topic does not cover fully autonomous systems
[2] Large Language Models in Cybersecurity – Threats, Exposure and Mitigation, edited by Kucharavy et al., Springer (2024)
[3] von Dzengelevski et al., The state of AI in the Swiss tech industry, ETH Zürich (2024)

Key findings

The following sections elaborate on the five key findings that emerged from this study.

Evolving cyber-risk landscape

With the increasing adoption of AI, the **cyber-risk landscape is changing**. Most interview partners agreed on this, but not on how exactly this change is happening.

With a strong agreement of 95%, most participants supported the statement that business will embrace the adoption of AI, regardless of its security. AI introduces probabilistic errors and, thereby, blurs business processes, resulting in increased financial and reputation risks. In part, this is because, due to their high complexity, we tend to accept AI systems as somewhat trusted black boxes.

The second most frequent concern is the increased risk of **leaking business secrets** by having employees posting them carelessly in AI-SaaS solutions such as ChatGPT. These tools will be used to create documents and plans leading to important business decisions, with severe risks for both loss of strategic business secrets and susceptibility to unnoticed manipulation of such documents.

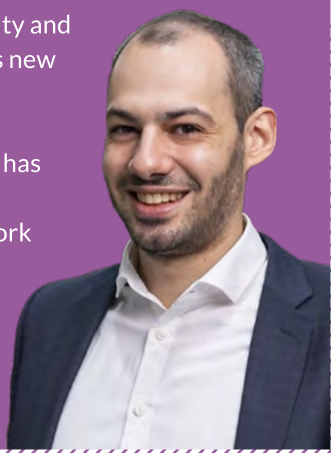
The public sector players interviewed also mentioned the risks of leaking official secrets.

Business will embrace the adoption of AI, regardless of security.

Expert opinion
Dr sc. ETH David M. Sommer

While the security threats emerging from AI are not fundamentally different from those posed by existing technologies, AI systems amplify the responsibilities of IT engineers to include even more socio-economic aspects than before. These AI systems are not only mostly trained on data originating from humans, but they also interact with humans, and can even impersonate humans. In addition, some AI systems introduce probabilistic error rates to previously deterministic processes, which means the procedures must be able to handle these errors.

Therefore, established risk management frameworks such as NIST or ISO are extending their scope with AI-specific additions to improve the reliability and trustworthiness of this new dimension of human computer interaction. In specific terms, NIST has developed an AI Risk Management Framework (AI RMF) and ISO has published its own guidance on risk management for AI, ISO/IEC 23894.



Expert opinion
Dr Raphael Reischuk, Partner & Group Head Cybersecurity, Co-Founder and Board Member of the Swiss National Test Institute for Cybersecurity NTC

Along with the risks identified by the interview partners, the future of AI will see more risks emerging. A few additional issues that might soon have an impact are:

Data poisoning / manipulation of business processes: more training data of AI models is likely to be poisoned, and the data entered to deployed models might be further manipulated. Since these models are set to process large amounts of data and derive business intel from it, this poses companies with the risk of executing faulty business decisions, resulting in significant damage, possible business bankruptcy, and loss of life.

Supply chain attacks / dependencies: as the few dominant AI models are increasingly integrated into most digital solutions, reliance on these solutions poses the risk of a single point of failure causing a cascading effect in the event of a compromise or security flaw, such as the recent compromise of the xz-utils library used as a module in a wide range of projects.[4]

Model leakage / IP loss: training large models consumes huge amounts of resources, time, and money. As a result, large models are typically not built by individual organisations themselves, but rather sourced in a SaaS style. This, however, puts the SaaS provider at risk of losing their intellectual property when hackers steal the models by sending large numbers of cleverly crafted queries.[5]

Concern about increased **dependency on third-party providers**, such as AI-SaaS services, is prevalent among the interview partners. Smaller organisations are highly dependent on the security procedures of their tool providers but are faced with insufficient transparency to evaluate the risks relating to availability, reliability, and data leakage, and to exercise accountability. Currently, most AI products are seen as being closer to proof-of-concepts than enterprise-ready, which leads to inevitable risks when deployed. The ability to adjust the security position lies with the provider, not with the organisation supplied.

In general, participants are missing the tools to evaluate and determine the quality of AI products from a security perspective.

New risks are emerging

As well as the known issues that are merely given a new dimension due to AI, new risks are also emerging. Prominent examples include intense **spear phishing** campaigns, which come with voice and video chats that are indistinguishable from authentic ones. These large-scale attacks are expected to become more frequent, thus requiring additional verification steps to secure payment and information release.[6]

A technique called **prompt injection**, which manipulates AI systems with the goal of releasing protected information to unauthorised users, will become a pressing issue with the introduction of **generative AI**. Systems such as chat bots might have access to sensitive information and could be tricked into revealing it in a way similar to social engineering performed on human beings. The Open Web Application Security Project (OWASP) has even dedicated a Top 10 list to the risks of Large-Language-Models[7] (LLMs).

[4] CVE-2024-3094 Detail (NVD)
[5] OWASP Top 10 for LLM Applications; LLM10: Model Theft
[6] Consider using Passkeys with FIDO, see Zühlke's blogpost 'Passwords are dead – long live passkeys'
[7] See OWASP Top 10 for LLM

Attackers have an advantage as they do not follow regulatory and ethical AI requirements.

It is evident that AI-driven automation will significantly empower attackers, ranging from script kiddies through cybercriminals and state players. This will lead to an increase in vulnerability exploitation rates, attack frequency, and facilitate lateral movement. On the defense side, AI systems will assist with task automation, analysis support, and anomaly detection, excelling in security log analysis.

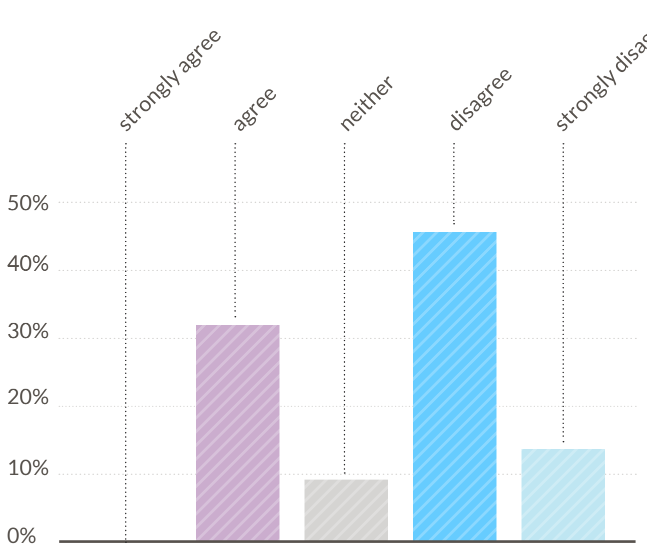
Our participants agreed that the **attacker-defender benefit ratio** will tilt towards the attacker in the short term until defence tools are available. Only then might the balance shift back a little. Cybercriminals are feared to have a significant advantage, as they do not need to follow the multiple regulatory and ethical requirements imposed on organisations deploying AI systems. On the question of the state of the balance in the long-term, opinions were inconclusively spread over all options. There is, however, agreement that the 'machine against machine' battle will significantly increase.

Are we prepared for AI risks?

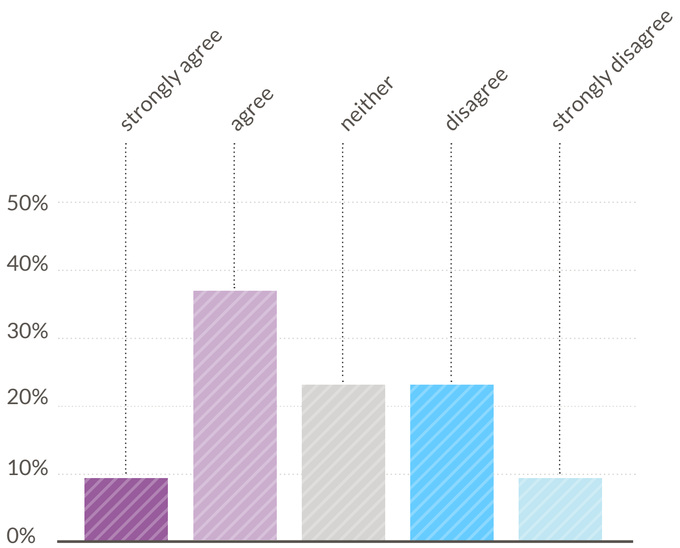
Among the participants, it is widely believed that AI adoption will severely impact the cyber risk mitigation landscape. This raises fundamental questions for everyone in the security industry, especially CISOs.

How well **are we prepared** for these new security challenges posed by AI? With respect to their own organisation, 44% of the participants (strongly) agree that their own company is well-prepared while 32% (strongly) disagree.

Security challenges resulting fom AI deployment and use



Security professionals are well prepared.



My company is well prepared.

Surprisingly, when asked whether the **industry as a whole is ready** (rather than just considering their own organisations), the overall majority of the answers were more pessimistic: only 28% of the participants (strongly) agree on this, whereas 61% (strongly) disagree. The most frequently mentioned concern was that some companies will be left behind by the sheer pace or impact of AI technologies.

How we are securing AI now

The potential of AI, both positive and negative, has been acknowledged by security professionals. The open questions are what its true long-term impact will be and which mitigation measures are already necessary.

Most participants reported that they had already **collected first experiences with AI technologies** on small, limited use-cases, potentially in isolated sandboxes. Often, motivated engineers and security experts are provided with a small number of resources to play around and gain experience. A few of the organisations interviewed are building up competence centres for AI-related security questions, often in collaboration with their data-scientist departments. Organisations with data analysis as part of their business models are creating AI governance boards or tasks-forces.

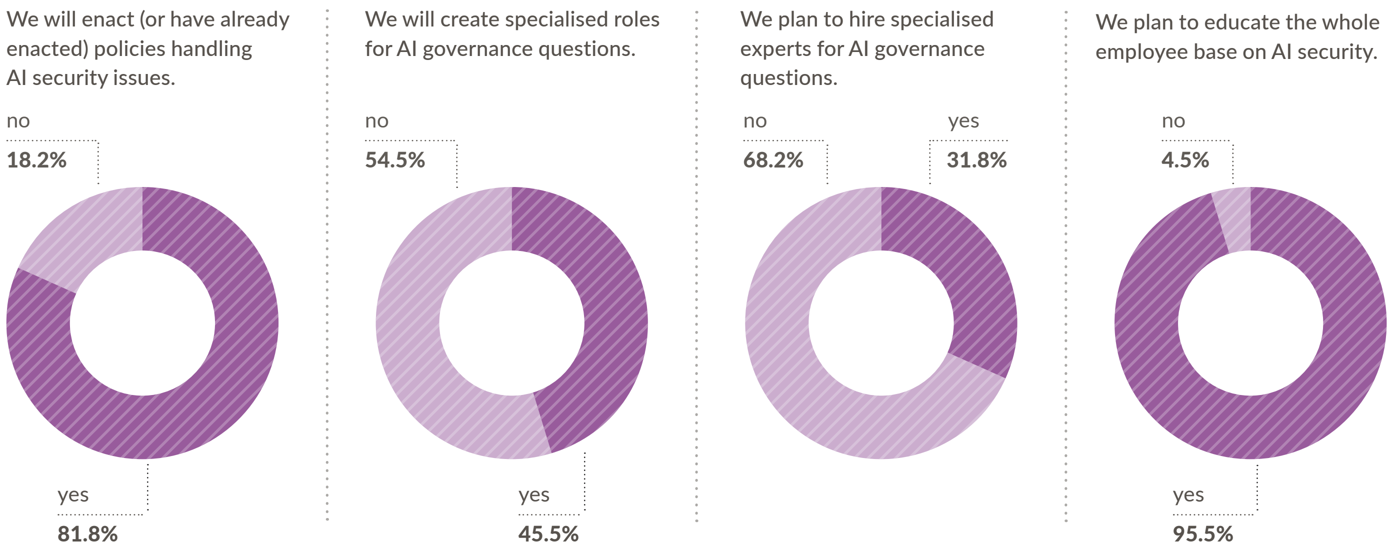
Expert opinion
Pascal Kocher, Security Operations Centre (SOC) expert

We can already see the attacker-defender benefit ratio shifting. With the help of AI, concrete examples of new types of cyber incidents include sextortion with AI-generated artificial images, convincing looking and sounding scam calls, and investment fraud in the name of celebrities.[8] The UK's National Cyber Security Centre (NCSC) predicts an uneven impact of AI on cyber treats.[9] Threat actors, including state and non-state, but also skilled and less skilled, are already using AI's capability to uplift in reconnaissance and social engineering. Novice cyber criminals, hackers-for-hire, and hacktivists will obtain unauthorised access and information more effectively. Overall, the UK's NCSC concludes, this is likely to lead to an increase in global ransomware threat.

> For risk management, it is essential to understand what AI is capable of and what it is not.

[8] See Swiss NCSC semi-annual report 2023/II
[9] See NCSC's article 'The near-term impact of AI on the cyber threat'

In the next 24 months...



AI-focused **third-party** services that are not hosted on-premises are of great concern. The fear of data leakage, manipulation, and unavailability grows with increasing dependence on services of this kind. Our participants agreed that blocking them would just lead to employees using the free versions of these services perhaps on their phones, in a phenomenon known as **shadow AI**. This would only render the issue uncontrollable. Instead, some organisations are creating transparent layers in front of these services when used by employees to carry out checks for data loss prevention (DLP) or to collect answers, in order to understand the origin of business decisions based on AI-facilited document writing.

Almost all the organisations interviewed have already prepared short **guides or pamphlets** for their general employee base.

When asked about **planned investments** relating to the security of AI adoption, participants most often mentioned education and awareness programmes for employees which mainly focus on understanding the risks of hallucination, data leakage, and regulatory requirements. In our study, 95% of companies also plan to **educate their whole employee base** on AI security.

Security specialists require an advanced understanding of data and AI, if necessary, through specialised training focussing on AI life-cycle management and third-party risks. Understanding what AI is capable of and what it is not is key.

In terms of education, enabling the middle and upper management to understand the risks and impacts of AI was mentioned by multiple participants.[10] In the absence of a sufficient understanding, organisations might take minor executive decisions that could result in major (negative) consequences because of considerations they may have overlooked.

Finally, many participants mentioned their hope in the forthcoming investments into AI-powered security tools which aim to gain an advantage over the attackers.

[10] Zühlke offers a course on Generative Artificial Intelligence for Business and on Creating business value from Data & AI for Data-driven Companies Executives

Changing responsibilities

The potential of AI to amplify or transform business processes is intriguing. While some organisations may remain cautious, others may follow the market push to start using AI technologies, and simply accept the risk. In this case, security departments are often only consulted late in the day and are unable to stop the adoption process.

At 71%, the vast majority of the participants agreed that **CISOs are shifting from control owners to risk decision facilitators**. However, it was also emphasised that this shift is not necessarily driven by AI, but by a natural tendency that had already been triggered by the fast pace of digitalisation and has now been amplified again by the urge to adopt AI. In the end, the business needs to drive the decisions. Others argue that CISOs, as control owners of the compliance processes, are the result of a healthy security position.

Overall, AI technologies will remain a crucial factor for security officers. Despite this, 55% agree that creating a specialist **Chief AI Officer (CAIO)** position is not the right approach. In total, less than half of the CISOs interviewed (45%) expect their organisation to create dedicated roles for AI security and even fewer (32%) expect them to hire relevant experts. Especially in SMEs, AI security will be handled by existing roles. Extending existing capabilities instead of setting up new avoids creating new silos.

For larger organisations, however, it may prove beneficial to have specialists in this evolving area. Some participants view the departments handling risk management and compliance as better suited than security departments for the build-up of AI knowledge. Regardless of which role or group oversees AI security in the end, it is essential that they keep an eye on further developments of AI.

Expert opinion
Felix Huber, expert in AI risk management

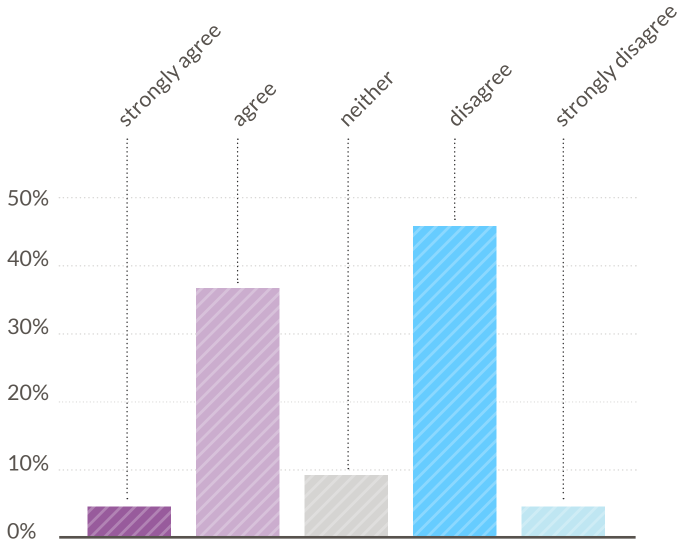
Although the interviewees believed the available standards and best-practice guidelines for securing AI systems to be insufficient, there are notable best practice recommendations and ongoing standardisation efforts regarding AI risk and AI management.

Standards and frameworks
The NIST AI Risk Management Framework focuses on the risks associated with development, deployment and use of AI. ISO/IEC 23894 (Guidance on AI risk management) extends ISO/IEC 31000 (Risk management) with AI specifics. ISO/IEC 42001 (AI management system) aims to establish AI governance.

Best practices
The OWASP AI Exchange collects AI security controls and mitigation measures. The MITRE ATLAS is a knowledge base for understanding and mitigating AI threats.

In addition to these, the ENISA is also exploring options on giving guidance with the ENISA Multilayer Framework for Good Cybersecurity Practices.

Regulatory requirements will suffocate AI



Policies, standards and best practices

In our study, 80% planned to have **policies** regarding AI security issues in place in the next 24 months, often for procurement and operations purposes. This indicates that organisations already see the need to regulate AI and are proactively defining methods of handling the security risks of AI internally. These first policies for this new type of technology are based on initial experiences and knowledge available in the public domain. [11] As we learn more about the field, the policies will need to be frequently revised over time. Large numbers of participants also plan on educating the whole employee base, which says a great deal about the importance of the topic.

On the one hand, **standards and best practices** for securing AI have been identified as necessary starting points for security measures. On the other hand, the lack of practical experience is currently impeding their broad application, which is especially true for best practices. The many standards currently emerging seem to have agreed on the ethical principles, but reliable recommendations are not yet available. The ecosystem will require some time until all the different approaches converge.

Regulation and fears

Current and future **regulation of AI** [12] is a major concern raised by the respondents. Most of them agree that AI shows significant positive potential, but the negative impact needs to be limited by means of regulations. Most interviewees agreed that the legislative process for detailed regulation is too slow to keep pace with the current technological development. Based on previous experience with data protection regulations and ongoing efforts, especially regarding EU legislation, over-regulation has often been cited as a burden on security because of the additional requirements imposed. This might especially affect small organisations and start-ups that cannot provide the resources for potential mandatory processes, forcing them to either default or leave the business.

[11] See NIST AI RMF or ISO/IEC 23894:2023 in combination with ISO 31000:2018 for AI risk management and OWASP AI Exchange and MITRE ATLAS for collections of best practices, threats, vulnerabilities and controls

[12] Some of the ongoing efforts include the EU's AI Act, China's Generative AI (GAI) Measures, US' AI Bill of Rights, UK's Pro-Innovative Approach to AI Regulation, and Singapore's National AI Strategy (NAIS)

There is a desire for regulatory provisions in order to facilitate quality evaluation of AI products.

Some participants expressed the desire that the industry should be given time to collect experience before attempting to impose new regulations. The assumption is that this would allow a balance to be found between ethics and openness and would allow room for innovative start-ups as well. For Switzerland, in which law-making processes are slower than in the EU and said to be better thought through, these concerns are also raised but less prevalent.

Around 40% of the respondents are worried that regulation will impede the smooth adoption of AI systems. In general, regulation will not be able to stop the rise of AI technology, which will balance this effect over time.

When inspecting **individual sectors**, strongly regulated industries like health, banking, and insurance seem to be more aware of the effort required to ensure compliance with the regulations, while other industries tend to underestimate it. This also depends on the individual organisation and cannot be overly generalised. Currently, all

organisations must focus on either building up expertise regarding AI regulation or staying up-to-date with the rapid developments.

When asked about their **expectations of governmental bodies and authorities**, the interviewees expressed not only concerns about regulation but also a desire for better knowledge sharing from three perspectives.

Firstly, this includes simple, government-issued advisories on how to achieve regulation conformity for smaller organisations and meaningful risks to include in the internal evaluation, which also considers the latest research and developments.

Secondly, the desire required and referenced should be free to access and use.

Thirdly, the desire for the governmental bodies to create and maintain trustworthy information sharing networks for the collection and distribution of up-to-date threat-intelligence data was expressed often, not only for AI specific issues. As most organisations have similar security concerns, it makes sense to tackle these on a collaborative basis.

Interestingly, some participants are seeking **quality evaluation of AI products to enable regulatory compliance**, giving smaller companies without specialist expertise in particular the necessary tool to make sourcing decisions on the adoption of third-party software, particularly SaaS-services. This corresponds to the aforementioned concerns about the lack of control over these services.

Recommendations

The following actions are recommended as the result of the study.

Do not rush

The topic of AI security is new and relevant. It shifts balances and introduces a few new risks, such as more effective fraud and spear phishing attacks by means of generative AI. However, most risks, such as data leakage, supply chain dependencies, and product immaturity, can already be handled today by improving established procedures.

Build expertise

Specialists must be aware of what AI is capable of and of what it is not. Motivated employees must be inspired to tackle the topic, especially regarding generative AI and prompt injections. AI products must be developed and deployed in a limited use-case with minimal impact in case of failure. Product providers may be asked for short-term licenses. Some organisations offer interesting courses. [13]

Create awareness

Employees must know where to be careful with AI and why. In particular, employees need to be aware of additional risks, such as fraud (authentic looking spear phishing), loss of confidential business data (by employees posting it in AI applications) and personal data (manipulated chat bots), but also forgery (using generative AI output). Short guides and pamphlets for employees are a helpful resource to explain the risks and provide guidelines on mitigating them. [14]

Adapt internal governance

With shifting balances, revisiting the organisational security governance processes might prove beneficial. For example, the next ISO 27001 issue is likely to contain regulations regarding AI. Zühlke's AI Risk framework might also be helpful. [15]

> Employees must know where to be careful with AI and why.

13 Zühlke Academy
14 The Institute of Risk Management (IRM) offers a Short Guide to Artificial Intelligence (AI) in Risk Management
15 Zühlke AI Risk Framework

Push for knowledge sharing

Joining AI security working groups, which are already established for most industries and sectors, is recommended. Security professionals should push for the establishment of government-orchestrated knowledge-sharing networks.

Follow recommendations

There is a growing number of high-quality recommendations. For example, it is worth consulting OWASP's dedicated AI Exchange platform, [16] as well as its specialised Top 10 lists for both ML [17] and LLM. [18] It is also worth consulting MITRE's ATLAS [19] knowledgebase. In the US, NIST offers an AI Risk Management Framework, [20] while in Europe, for example, Germany's Federal Office of Information Security (BSI) provides security guidelines for developers. [21]

Stay tuned

The tools and regulatory landscape available are rapidly expanding. Starting to maintain an overview early enough will allow security professionals to know when and how to act. This can be achieved by regularly checking the resources outlined in the previous paragraph and following new developments.

Review AI with a constructive scepticism

AI will likely bring a significant efficiency boost for all organisations, including criminal ones. It introduces probability into processes and operates with considerable error rates. For certain use-cases, such errors are acceptable, for others they are not. Approaching them with an open-minded and constructive mindset, while at the same time being critical and cautious, will certainly be worthwhile. A positive dialog about the potential and risks of AI will lead to a sound ecosystem of controls and will enable optimum use of AI.

16 OWASP AI Exchange
17 OWASP ML Security Top 10
18 OWASP Top 10 for LLM Applications
19 Adversarial Threat Landscape for Artificial-Intelligence Systems
20 AI Risk Management Framework
21 AI security concerns in a nutshell

Contact

Dr sc. David M. Sommer, Lead Security Consultant
LinkedIn

Sarah Röthlisberger, Senior Consulting Manager
LinkedIn

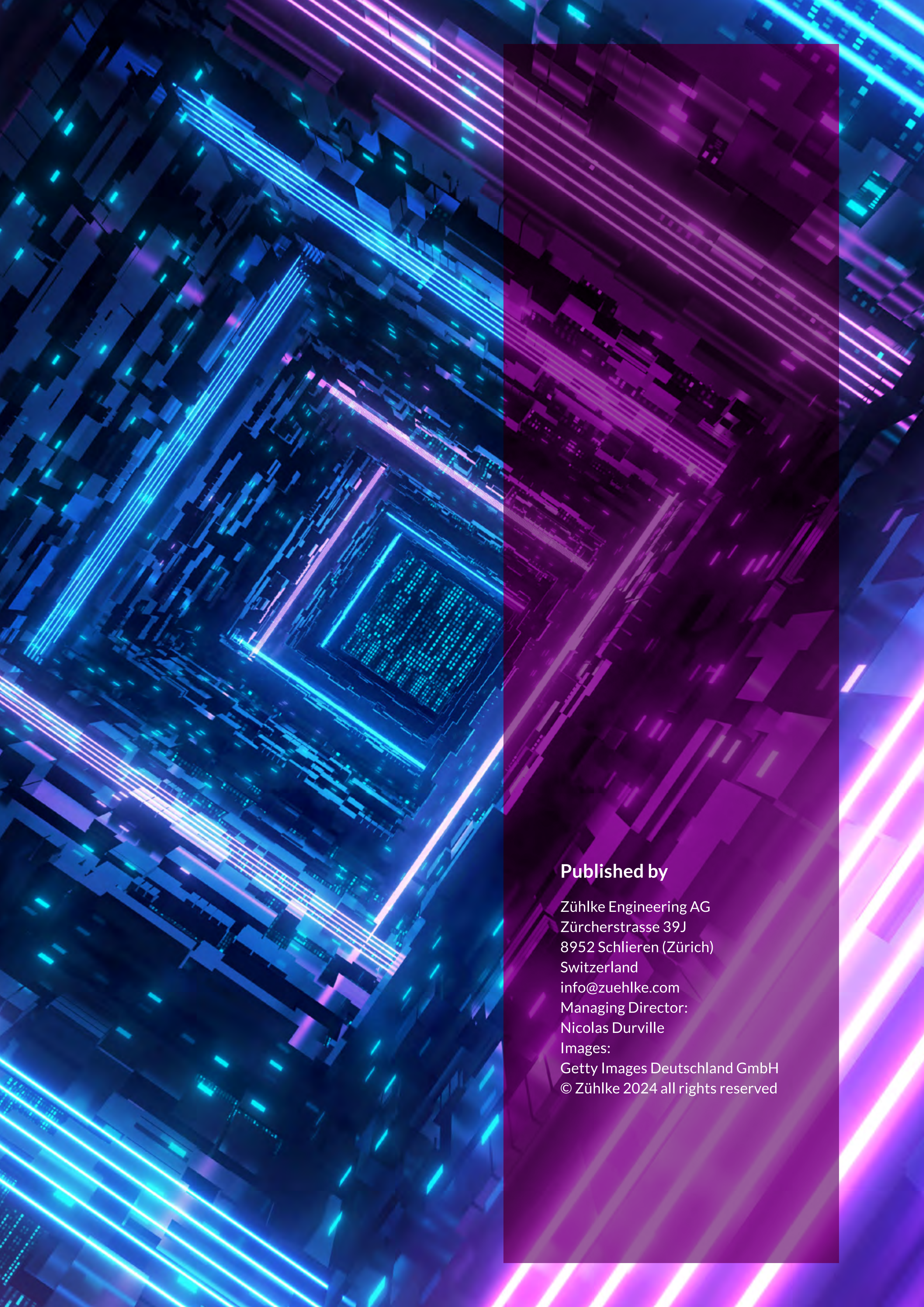
Dr Raphael M. Reischuk, Group Head Cybersecurity & Partner
LinkedIn

Acknowledgements

We would like to thank all our participants for their time and great contributions.

Zühlke – Empowering Ideas.

Zühlke is a global innovation and transformation partner. We turn opportunities into early and ongoing impact. For businesses, people, and the planet. As a holistic growth partner to our clients, we close the gaps between strategic ambition, tech implementation, and operational excellence, helping organisations thrive in a business environment that's tech enabled, data empowered, and hyper connected. Zühlke was founded in Switzerland and is a global team of 1,900 colleagues, with 17 locations in Europe and Asia. Our multidisciplinary teams deliver innovation and transformation initiatives in complex and regulated industries and ecosystems including health and financial services.



Published by

Zühlke Engineering AG
Zürcherstrasse 39J
8952 Schlieren (Zürich)
Switzerland
info@zuehlke.com
Managing Director:
Nicolas Durville
Images:
Getty Images Deutschland GmbH