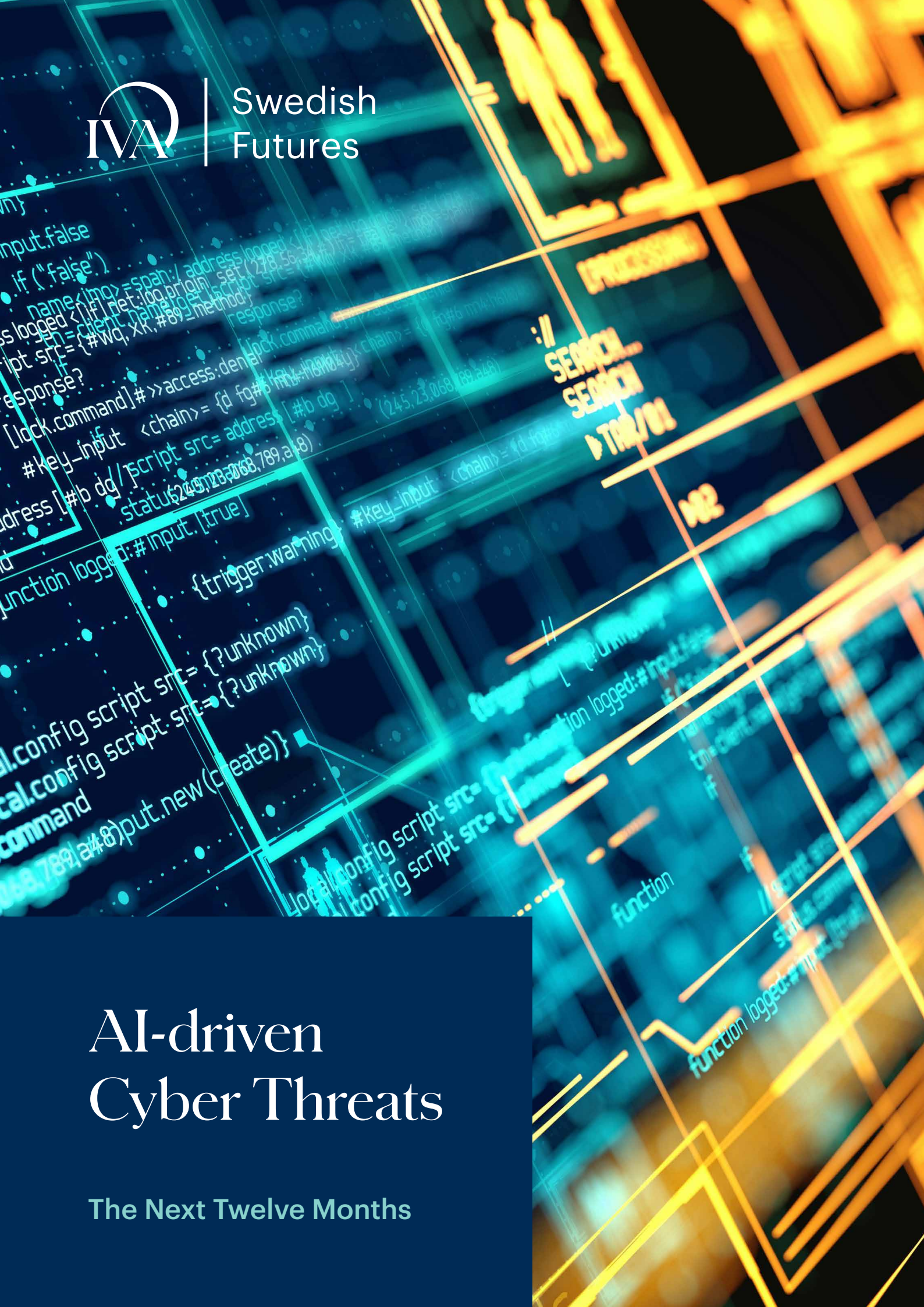


AI-driven Cyber Threats

The Next Twelve Months

Contents

Foreword 4

Executive Summary 6

A New Reality – Why the Issue Is Urgent 8

The Next Twelve Months 10

Priorities for Action 13

- Speed 14
- Basic cybersecurity work 14
- Governance 15
- Information sharing and collaboration 15
- Fighting AI with AI 15

Conclusion 17

Appendix 19

- Methodology 20
- References 21

Foreword

IVA is spearheading the *Swedish Futures* initiative to formulate a vision for Sweden as a leader in technology and innovation by 2035. Since autumn 2025, *Swedish Futures* has brought together stakeholders from academia, industry and the public sector to identify opportunities, challenges and strategic directions for competitiveness, sustainable development and security. Among other things, the initiative convenes working groups that quickly and systematically analyse challenges and opportunities in different technology areas and produce highly focused reports. These reports provide an overview of the status quo and outlook for the field under examination as well as present concrete proposals for action. They also serve as an important foundation for shaping an overarching vision for Sweden by 2035.

Artificial intelligence is rapidly changing the conditions for cybersecurity: capabilities that once required rare specialist expertise are becoming cheaper, faster and more widely available to attackers and defenders alike. The overarching question the group set out to answer is: What should we in Sweden do in the next twelve months to reduce the most likely and most serious cyber threats given what we see of accelerated use of AI?

This interim report is deliberately short and high-level. It conveys one view of the problem and a small number of priority directions that are relevant to society as a whole. The intention is to continue the work with a more in-depth follow-up report. The report has been prepared in close dialogue with the National Cyber Security Centre Sweden (NCSC-SE) and AI Sweden's project on national mitigation of AI-driven cyber threats.

As is the case with all IVA projects, all participants contribute in their personal capacity and not as representatives of the organisations for which they work. The report's analyses and recommendations are based on the experience and knowledge the contributors brought to the table, and the discussions these inputs engendered. The working group endorses the report, although this does not mean that all members necessarily endorse every formulation in detail.

The working group on AI-driven cyber threats convened from May to June 2026.

Working group

Pontus Johnson (Chair), KTH, IVA Fellow
Anne-Marie Eklund Löwinder, Amelsec, IVA Fellow
Erik Ekudden, Ericsson, IVA Fellow
Patrik Fältström, Netnod, IVA Fellow
Daniel Gillblad, Recorded Future, IVA Fellow
Fredrik Heintz, Linköping University, IVA Fellow
Simin Nadjm-Tehrani, Linköping University, IVA Fellow
Anders Sandberg, The Institute for Futures Studies, IVA Fellow
Staffan Truvé, Recorded Future, IVA Fellow

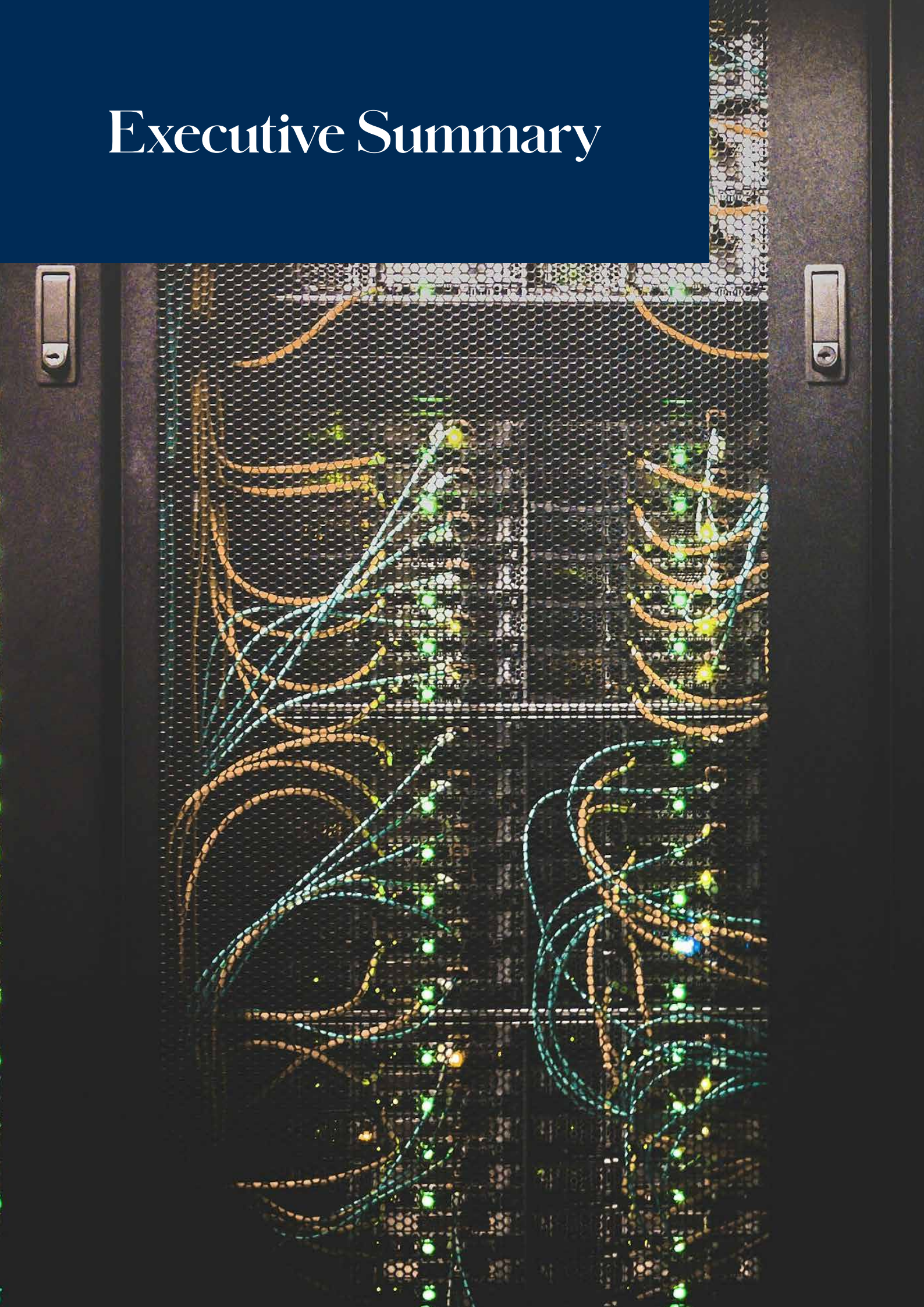
Support for the working group

Per Hjertén, IVA, Project Manager
Musard Balliu, KTH, Author

The report has been prepared in close dialogue with

John Billow, National Cyber Security Centre Sweden (NCSC-SE)
Otto Elmgart, National Cyber Security Centre Sweden (NCSC-SE)
Robert Valsjö, National Cyber Security Centre Sweden (NCSC-SE)
Robert A. Bridges, AI Sweden
Mauricio Muñoz, AI Sweden
Mats Nordlund, AI Sweden
Tommy Schönberg, AI Sweden

Executive Summary



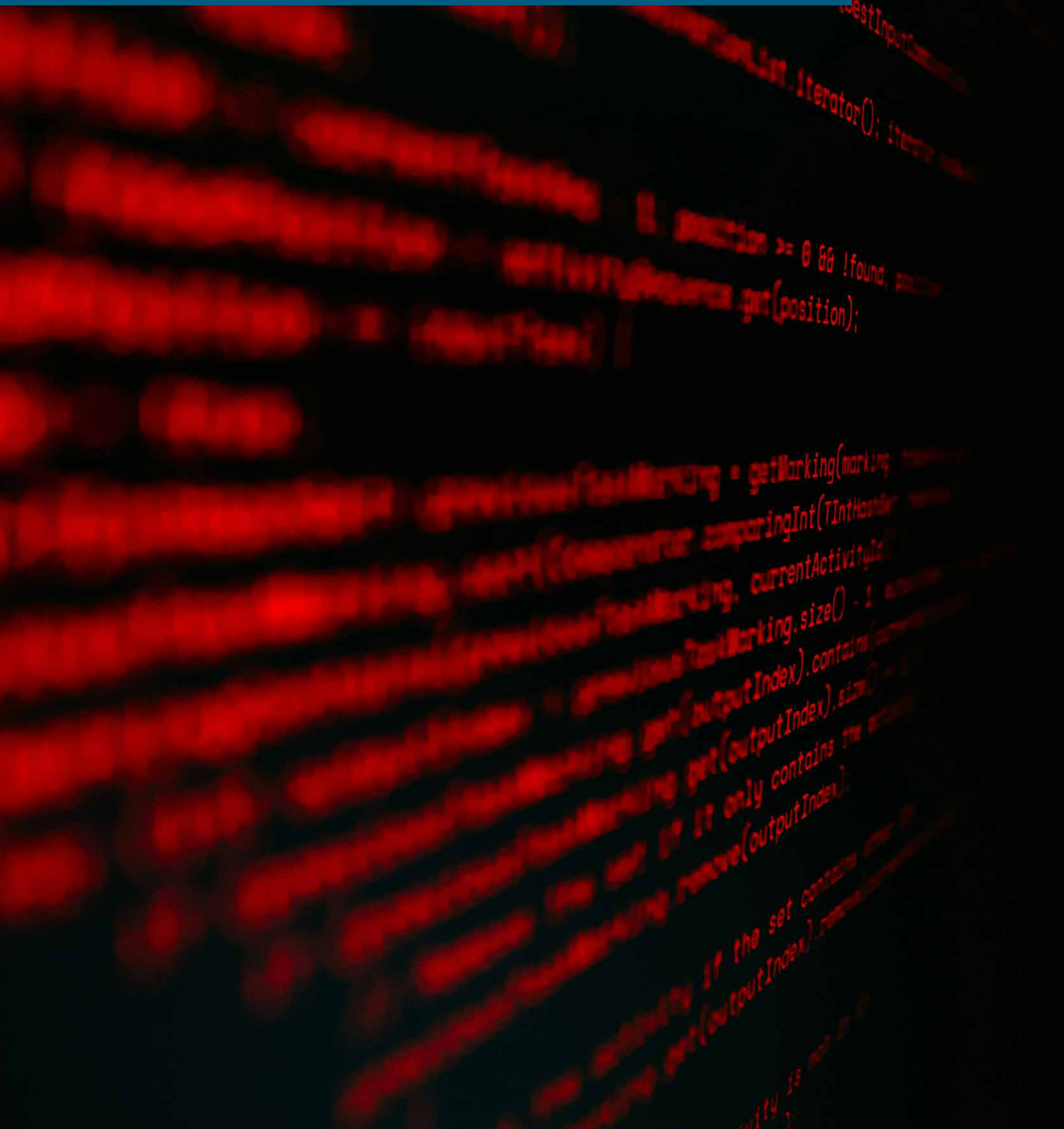
Artificial Intelligence (AI) is changing the tempo of cybersecurity. The central challenge is to shorten the time between understanding a risk and acting on it. This interim report concludes that AI increases the speed and scale at which many cyber weaknesses can be found and exploited. It is not that AI replaces cyber threats, but that it increases their reach, pace and practical impact, while also requiring organisations to secure the AI systems they put into use. Advanced AI models can support coding, troubleshooting, analysis, and technical reasoning. When given access to tools and data, they become practical aids for long chains of malicious cyber activity.

This changes the risk picture for us in Sweden. Many organisations lack a clear picture of their digital environment and do not have sufficiently robust routines for managing access control, keeping systems updated, reducing unnecessary exposure, and detecting incidents in time. AI increases the risk related to these weaknesses because AI helps attackers find and exploit them at a higher pace than before. Cost and knowledge barriers for attackers are falling, while powerful models and reusable tools are spreading to a wider range of actors.

The working group assesses that the next twelve months should be understood as a race between offensive and defensive adoption. Attackers can use AI to work faster, scale attempts, and adapt their methods. Defenders can use the same technology for code review, vulnerability prioritisation, monitoring, analysis, incident response, and training. The key question is whether we in Sweden can make AI raise the productivity of defenders faster than it raises the capability of attackers. For boards, executive teams and public-sector leaders, this points to three immediate priorities: know which services and dependencies must continue to work, shorten the time from risk awareness to action, and introduce AI into cybersecurity practices under clear governance.

The response must be coherent. AI increases the speed of the attack cycle. That forces a corresponding increase in the speed of defence. Speed is a central part of the story, but it is not the whole story. The ability to act at speed depends on sound governance, clear priorities, and capacity to proactively anticipate risks before they become incidents. AI changes the pace of cybersecurity, broadens the attack surface, and places new demands on how AI systems are governed and secured. The basic cybersecurity work that every organisation needs to have in place must therefore be carried out with much greater consistency, discipline, and agility. Governance must become more risk-based and able to operate at a tempo never seen before to support timely decisions in a fast-moving environment. This requires clear responsibility for critical services, AI-supported decisions, supplier dependencies, and the authority to act during incidents. Collaboration and information sharing must happen much faster and become more operational, and more trusted. Finally, just doing the same work faster will not be enough. AI must be integrated into cybersecurity operations. This should be done in a controlled and responsible way: tested before use, governed in practice, and applied where it strengthens defence without creating new unmanaged risks.

A New Reality – Why the Issue Is Urgent



Digitalisation has already made critical services dependent on software, data, networks, and complex supply chains. In Sweden, this dependency is visible in everyday digital infrastructure such as BankID and payment services, in municipalities, healthcare and connected industrial environments, and in reliance on international cloud, software and AI providers. IVA's earlier report "Cybersecurity for increased competitiveness" emphasised that Sweden's high level of digitalisation creates both opportunity and vulnerability, and further that cybersecurity is linked to competitiveness and resilience (IVA, 2022). This remains the basis for the future. AI adds a new layer to this dependency. It makes existing cyber risks more likely to turn into operational threats by increasing the speed, scale, and reach with which weaknesses can be found and exploited. Two complementary needs should therefore be addressed together from the start: protecting against AI-powered cyber threats and securing the AI systems that organisations use – whether for cybersecurity or for other purposes.

This development is not driven by one single breakthrough. It is the combined effect of several reinforcing changes. AI models are over time made more capable. The tools using the models have improved. Knowledge about how to use models is spreading. The result is that more actors can perform more tasks, more quickly, and with less expert support than before. This is true both for the attackers and defenders.

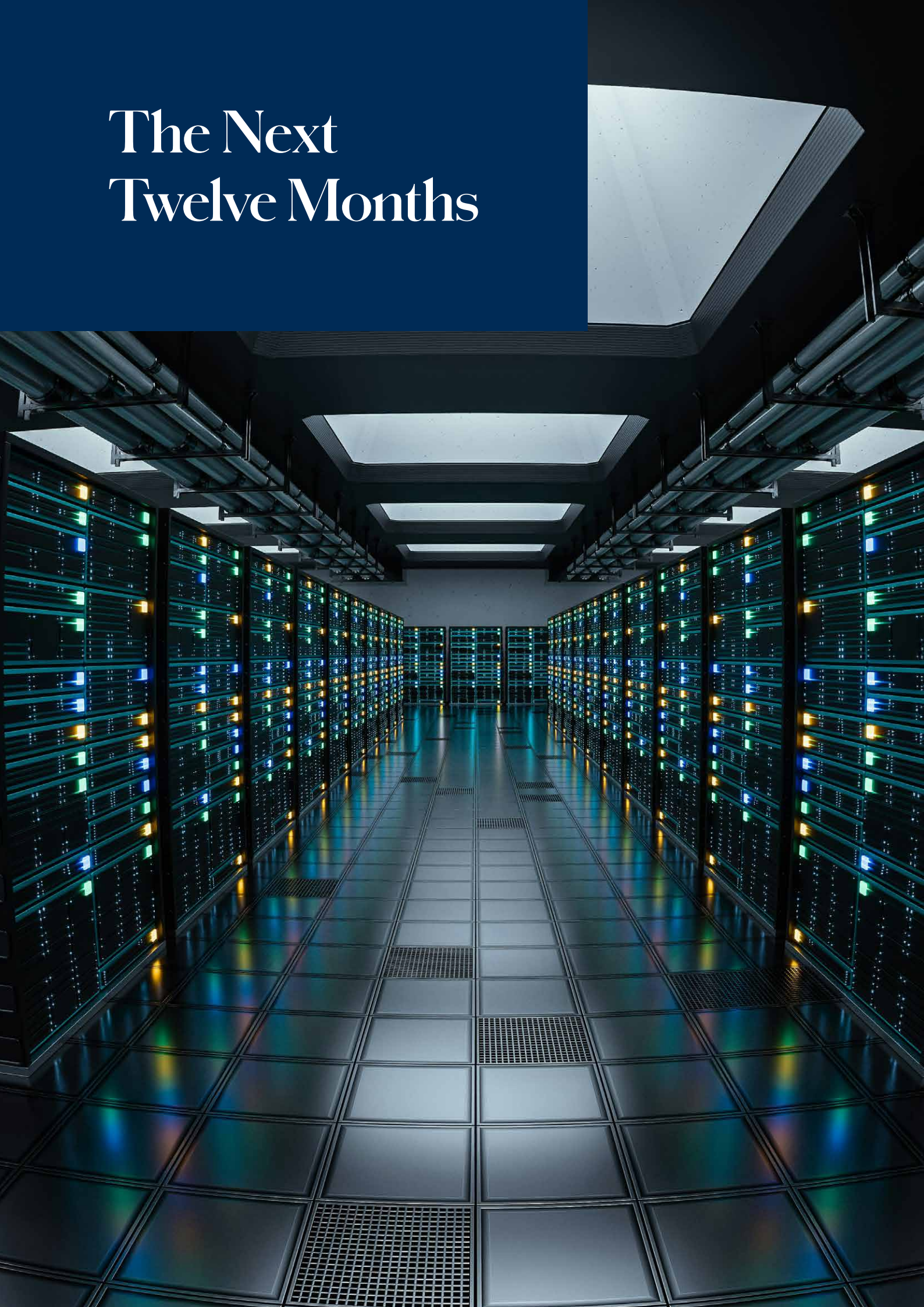
AI should be treated primarily as a productivity technology, not as an independent actor. This does not mean that automated or agent-like AI systems should be underestimated. It means that responsibility remains with the people and organisations that design, deploy, and depend on these AI systems. Many of the most important risks are still rooted in weak security baselines, unclear responsibilities, and poor understanding of supply chain dependencies.

An organisation that cannot identify its exposed systems, that does not know which suppliers support critical functions or that lacks working incident-response routines, will not become secure by focusing on the choice of the AI tool. Fundamental cybersecurity basics such as segmentation, monitoring, patching, access control, and accountability, remain more important than the brand of a model.

The UK National Cyber Security Centre assesses that AI is already making parts of cyber intrusion more effective and efficient and that, up to 2027, it is likely to increase the volume and impact of intrusions mainly by enhancing existing methods rather than creating wholly new ones (UK NCSC, 2025). This is consistent with the working group's view. Vulnerabilities can be found more quickly, many attacks require less expert knowledge, and repetitive tasks such as reconnaissance and social engineering can be scaled up at low cost.

At the same time, AI is not perfect. It will generate failed attempts, noisy probing, and inaccurate outputs. Defenders will need to decide how to respond to a higher volume of both meaningful and low-quality activity. The practical pressure is still clear. If vulnerability detection and exploitation move faster, the time between exposure, discovery and attack becomes shorter. Organisations with strong control over assets, identities, patching, dependencies, and security processes will remain far more resilient than organisations without such control. But even strong organisations will need to shorten decision cycles, test new defence methods (for example by use of AI) and treat AI systems themselves as part of the attack surface.

The Next Twelve Months



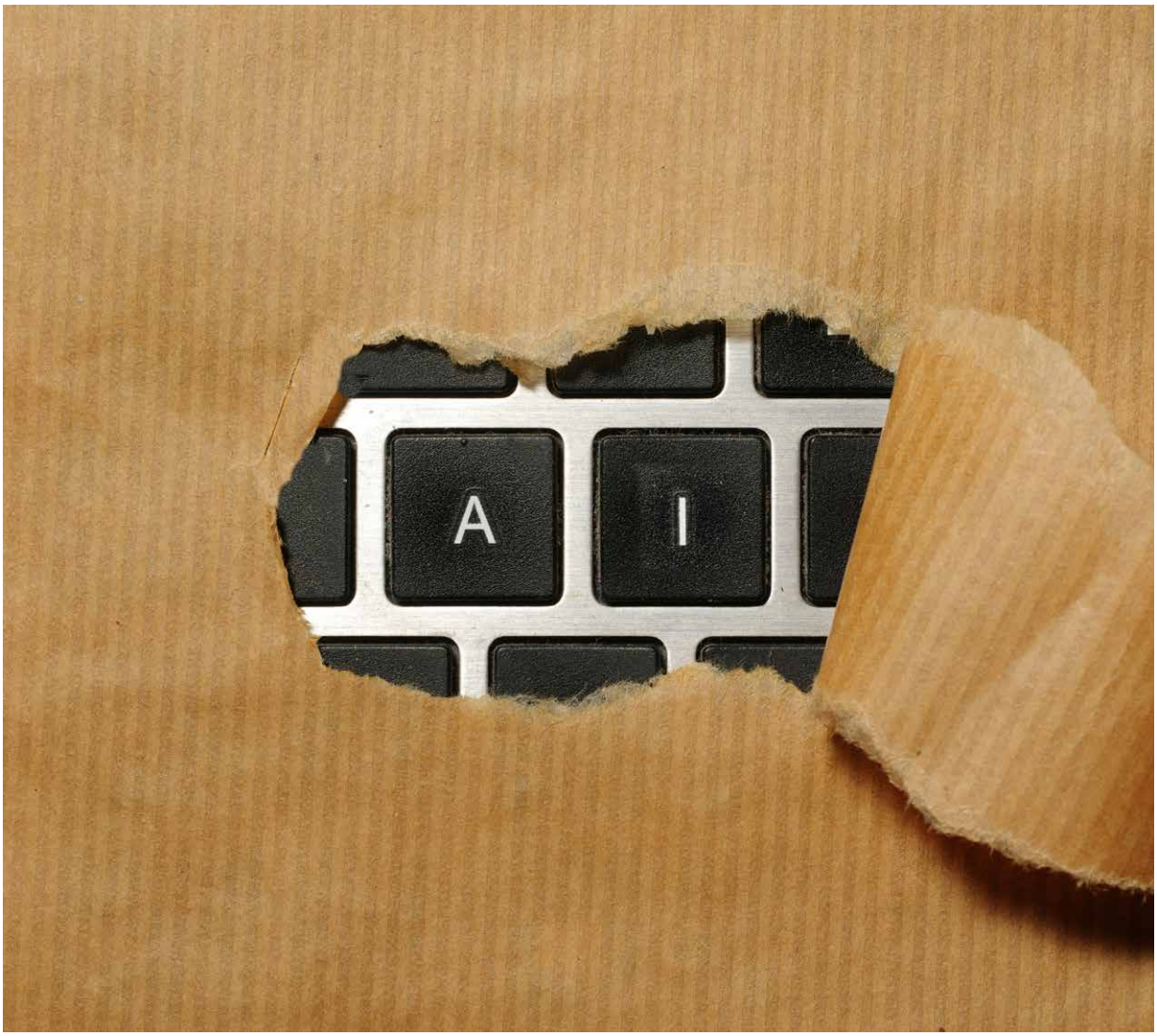
The next year is likely to bring further increase in AI-powered cyber capability. Stronger models, better tools, and more capable AI agents will make it easier to analyse systems, learn unfamiliar tools, troubleshoot problems, and automate repetitive work. The UK NCSC and the UK AI Security Institute argue that defenders should assume that at least some attackers already have access to capable AI tools, while also stressing that defenders can use the same capabilities to gain an advantage (UK NCSC and UK AISI, 2026). The most advanced capabilities may initially be concentrated among states and other well-resourced actors, but the trend still matters for all organisations as tools become easier to use and adapt.

Vulnerabilities and weak security practices will therefore become more consequential. Weak passwords, misconfigured systems, old software, and poor logging are not new problems. What changes is the speed and cost with which they can be found and exploited. AI can make the long tail of ordinary weaknesses more visible and usable to attackers. AI-assisted software development may also add risk if functional code is produced and deployed faster than organisations can review, test, and maintain it securely (AI Sweden, 2026). It may also make more tailored attacks affordable, especially where the target is valuable. Financially motivated actors will often continue to choose easier targets, but actors seeking disruption or strategic effect may also direct AI-supported efforts against more hardened environments.

AI systems and AI-supported workflows will themselves become more attractive targets. The issue is not only whether a model can produce harmful technical advice. It is also what an AI system is allowed to do inside an organisation, what data it can access, and what actions it can trigger. Tricking AI systems is often not difficult. An attacker may try to make an AI assistant reveal information, follow hidden instructions, make an inappropriate decision or misuse a business process. We should assume that organisations will face various forms of manipulation that are difficult to predict in advance. NIST's work on adversarial machine learning underlines that AI security must be managed across the life cycle of AI systems, not added as an afterthought (NIST, 2025).

While general access to AI continues to broaden, a further concern over the next twelve months is reliable access to the specific models, services and tools that organisations may come to depend on. As AI becomes more useful for cybersecurity operations, access to capable AI systems will itself become part of the question of resilience. This is a practical operational issue: organisations should avoid building critical defensive capabilities around services that they cannot control, replace or continue without. The same applies regardless of whether the service or supplier is domestic, European, or global. The key question is whether the organisation understands the dependencies and has realistic alternatives if the availability of those capabilities changes.

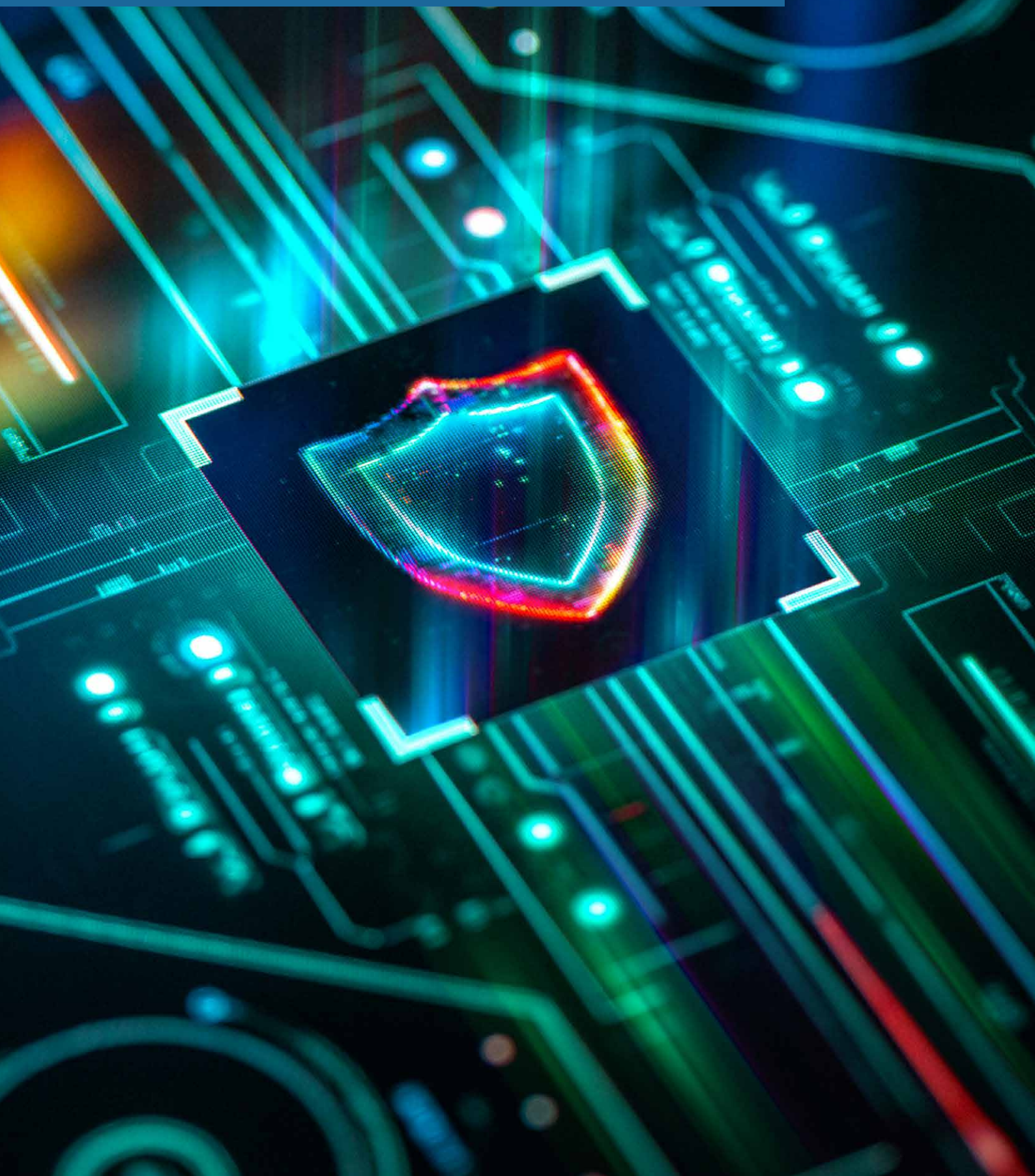
Recent public examples point in the same direction. Anthropic has reported a case in which agentic AI capabilities were used to support a cyber espionage operation against global targets (Anthropic, 2025). Another attack against an AI-supported customer-service workflow at Meta also illustrates that the risk is not only advanced code exploitation, but the manipulation of AI-enabled processes that have authority to act (MIT Technology Review, 2026). Anthropic's suspension of access to Fable 5 and Mythos 5 in May 2026 shows that access to advanced AI can change abruptly for legal, geopolitical or security reasons, making model access a resilience issue for AI-supported defence (Anthropic, 2026). These cases should not be treated as exact forecasts for Sweden, but they show why AI security, basic cybersecurity and control over critical dependencies increasingly belong together.



Several agencies in other countries are reaching similar conclusions. New Zealand's NCSC describes frontier AI as a dual-use development that can make known vulnerabilities, legacy systems and weak cyber hygiene more dangerous (NCSC New Zealand, 2026). Canada's National Cyber Threat Assessment and the Australian Signals Directorate's annual threat report both stress that cyber threats are increasingly tied to geopolitics, critical infrastructure and the ability of public and private actors to work together (Canadian Centre for Cyber Security, 2024; Australian Signals Directorate, 2024).

The potential impact is not one-sided. Defenders will also gain better tools. AI can support code review, vulnerability analysis, monitoring, incident response, triage, and education. CISA's AI roadmap frames AI both as a technology to secure and as a technology that can strengthen cybersecurity efforts when adopted responsibly (CISA, 2023). A realistic objective for us in Sweden should be to make defensive adoption faster, safer and more systematic than the aggressors' offensive adoption.

Priorities for Action



The central storyline is simple but not simplistic. AI changes the pace of cybersecurity, but it also changes how software is produced, how attacks can be automated, and how AI-powered processes can be misused. This forces a corresponding increase in the speed of defence. The most important challenge is not a single product, platform, or threat actor. The challenge is that many organisations still operate with processes and decision structures designed for a slower-paced environment, where we do not have any enemies intentionally trying to destroy our systems. We therefore need a response built around five connected pillars: speed, basic cybersecurity work, governance, information sharing and collaboration, and fighting AI with AI.

Speed

Speed must become a strategic priority. That means urgency with discipline, understanding expectations as well as dependencies, and not speed for its own sake. Organisations need to proactively reduce the time between knowing and acting. This applies to vulnerability management, patching, incident response, procurement, decision-making, and communication. The aim is not reckless action, but a better balance between caution and delay. In an environment where attackers can test, adapt and scale fast, waiting for perfect information, or the perfect solution, may itself become a risk.

This means that organisations need to reassess how they weigh risks, speed, and action. Leaders and decision-makers in boardrooms should ask which decisions can be made earlier, which mitigations can be introduced temporarily (before patches become available), and which processes can be simplified without losing accountability. They should also ask whether responsibility is placed where timely action can be taken, and whether some decisions need to be delegated, pre-authorised or clarified before an incident occurs. The same discipline is needed when organisations deploy new AI tools: moving quickly without guardrails can create new vulnerabilities. The organisations that are best prepared will be those that can act quickly, learn from real events, and adjust their course continuously.

Basic cybersecurity work

Faster evolving threats make basic cybersecurity more important, not less. Multi-factor authentication, sound access control, asset identification, secure configuration, logging, monitoring, backup, incident response routines and patch management remain the foundation of resilience (NCSC-SE, n.d.). These are not new measures, but AI makes the consequences of failing to implement them much more severe.

The practical task is to make basic cybersecurity work more consistent and more agile, integrated as a part of daily work. Organisations need to understand their own digital environment: which services matter most, which systems and suppliers support them, who has access, how vulnerabilities are handled, and what must continue to work during disruption or patching, no matter what. Cyber hygiene must be treated as an operational capability, not as a one-off compliance exercise. Organisations should also work from an "assume breach" mindset: planning on the basis that attacks will occur, and focusing on how they will be detected, contained and limited in impact. Once these basics are in place, organisations should improve how they sustain them as AI becomes part of daily operations and security work.

A further practical need is basic education for users and decision-makers. People do not need to understand every AI tool in detail, but they do

need to know that AI agents may be given access to files, credentials, customer data, systems, and other sensitive resources. That understanding should become part of ordinary cybersecurity awareness as AI tools enter daily work.

Governance

Governance must adapt to a higher-paced threat environment. Boards, executive teams and public-sector leaders should treat AI-related cyber risk as part of ordinary organisational risk, not as a narrow technical issue. The use of AI does not remove responsibility. The organisation that provides a service remains responsible for the risks, dependencies, and outcomes connected to that service.

Governance also needs to start with a clear understanding of what the organisation is expected to deliver. Many organisations try to protect everything equally, which makes the task almost impossible. A more realistic approach is to first identify the organisation's mission, and then what services, information assets, systems, and suppliers are truly critical for that delivery. Leaders should know which dependencies could prevent delivery, which risks are outside the organisation's risk appetite, and where a single supplier or service has become a single point of failure. And this regardless of whether it is internal or external suppliers or services. The same logic applies to AI adoption: leaders should make sure that new tools do not bypass established responsibilities, security checks, or procurement requirements.

The most useful governance questions are practical. What functions are critical? Which suppliers, systems, and data do they depend on? How quickly can the organisation make decisions during an incident? Who has the authority to act? How can AI help, how is AI being used, and how are AI-powered processes secured? Governance should help organisations ask the right questions early enough for the answers to matter. These questions must be asked proactively and routinely, not later as a reaction to some incident.

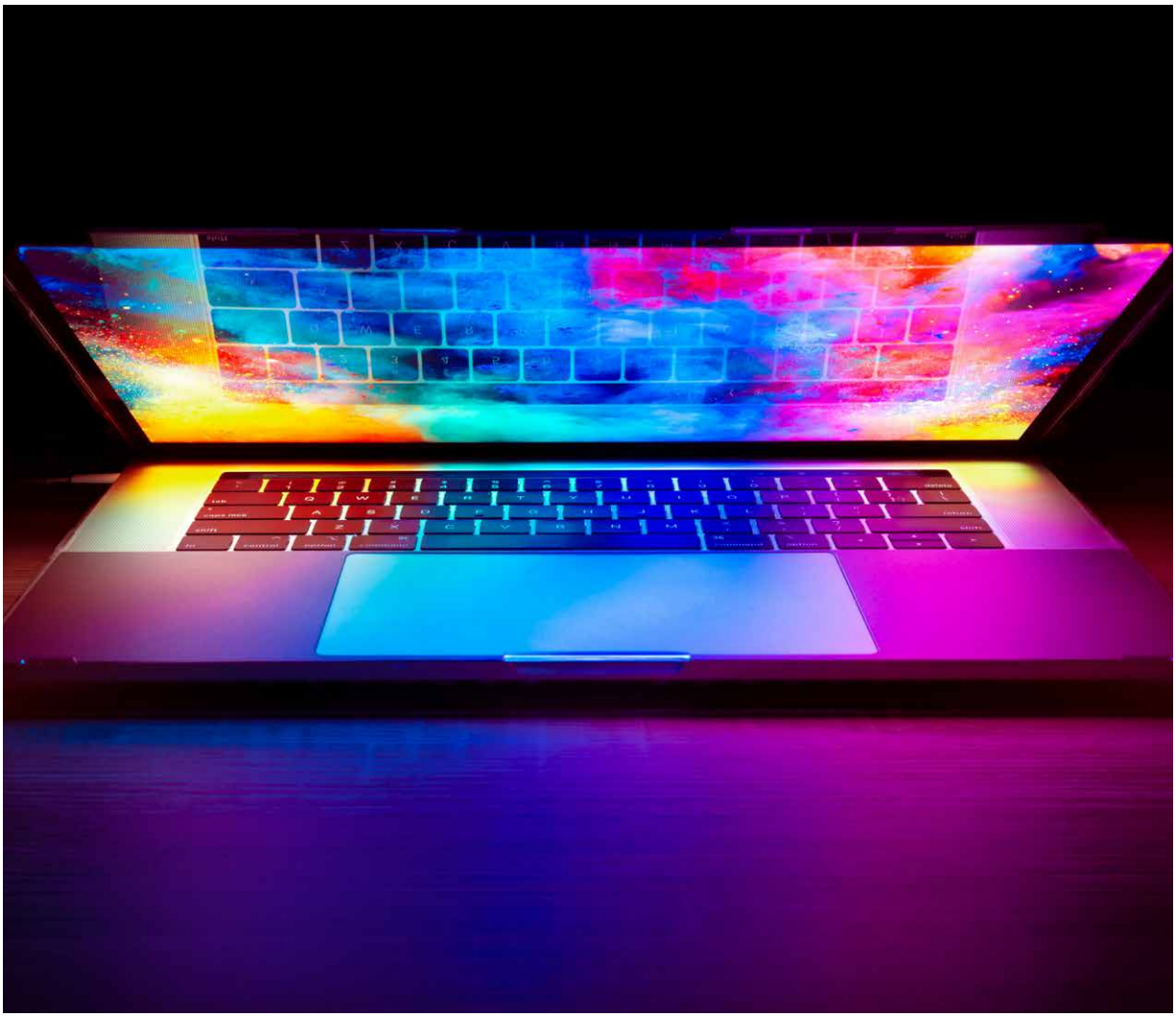
Information sharing and collaboration

Faster operations also require faster collaboration. Threat information, lessons learned and practical insights need to move more rapidly between public authorities, companies, researchers, operators of critical infrastructure, and trusted international partners. We need well-functioning and well-established collaboration mechanisms that can be used in normal operations, not only during major crises.

Information sharing is not only a technical matter. It depends on trust, clear roles, and relationships built before incidents occur. Lightweight crisis management practices should become part of ordinary preparedness. The goal should be that relevant actors can find one another quickly, understand what information can be shared, and turn warning signs into practical action.

Fighting AI with AI

Finally, simply doing the same work faster will not be enough. Sweden must fight AI with AI and use AI where it gives defenders a clear and governed advantage. Organisations should use AI to strengthen code review, vulnerability management, monitoring, threat analysis, incident triage, decision support, and training. The objective should be to make AI improve defensive productivity more than it improves offensive capability. This should be done with an explicit

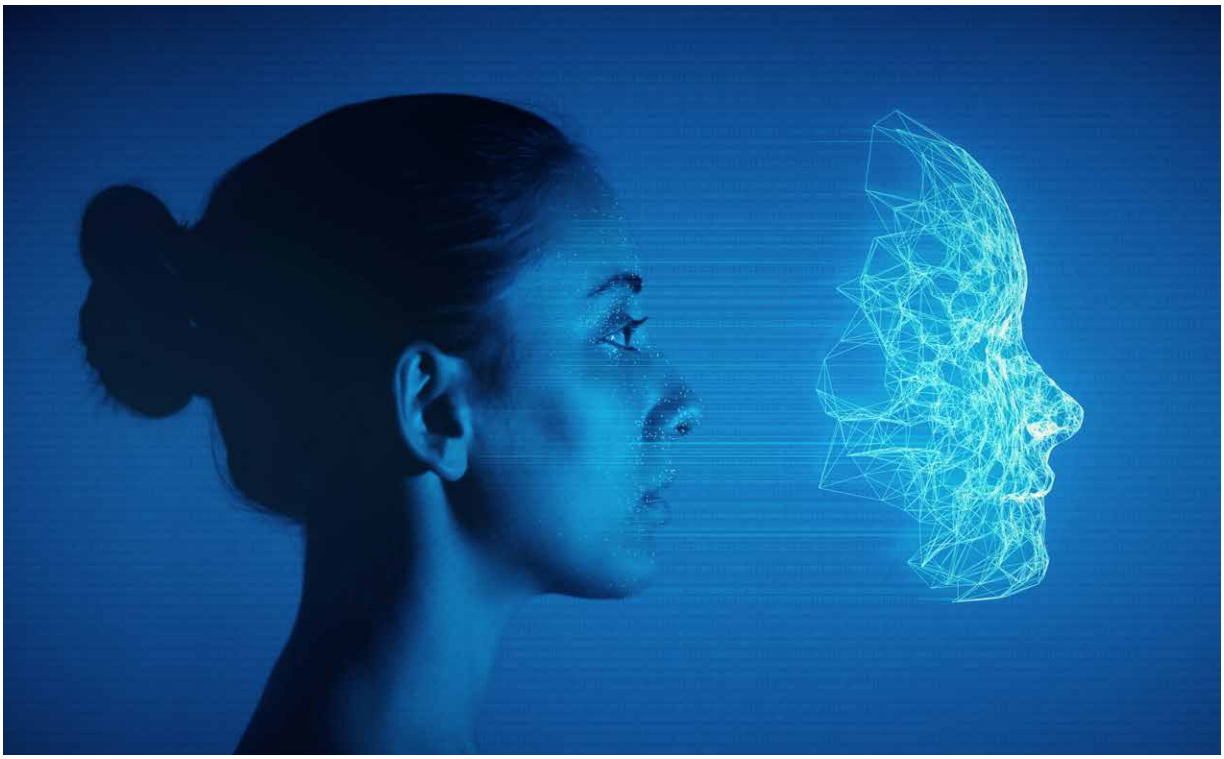


understanding that every new AI-powered process may also create new dependencies, data flows, and attack surfaces.

AI should be used as a productivity tool under clear human and organisational control, with responsibility remaining with the people and organisations that decide how it is deployed and used. AI-powered defence should be tested, governed, and integrated into existing security work. The organisations that learn to use AI safely and effectively will be better placed to manage the faster and more complex cyber environment that is now emerging.

Taken together, the five pillars form one response. Faster attacks require faster defence. Faster defence requires strong basic cybersecurity work, governance that enables action, collaboration that moves at operational speed, and the systematic use of AI in defence. This is what we should prioritise over the next twelve months in Sweden. And we should not wait for perfect conditions or for others to move first. Each of us should do what we can, individually and together with others, at each moment in time within the context in which we operate.

Conclusion



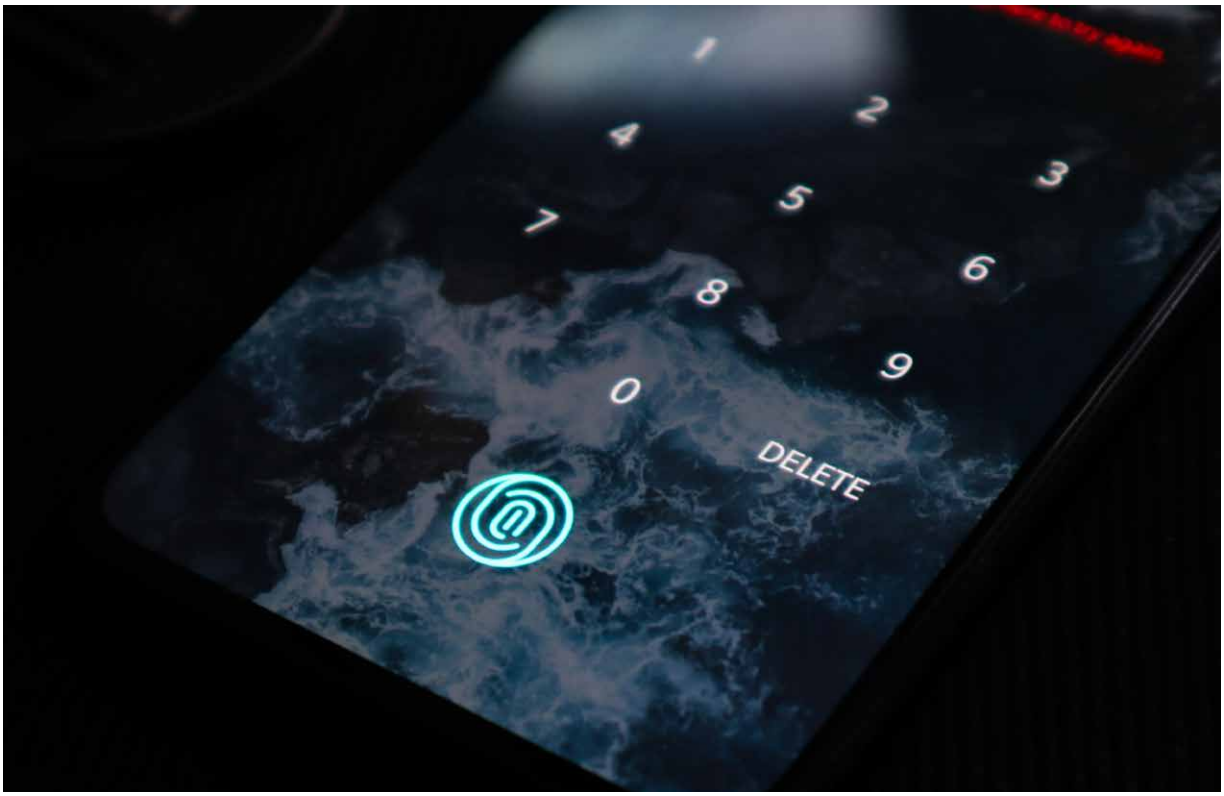
AI-driven cyber threats should be treated as a whole-of-society challenge. Sweden is not alone in facing them, and no single actor can solve them. The point of asking what we in Sweden should do is therefore a question of shared responsibility between organisations that design, operate, regulate, supply, and depend on critical digital infrastructure.

The necessary trade-offs should be handled openly. Faster action reduces exposure, but it must be supported by clear judgement and accountability. Wider use of AI can strengthen defence while also creating new dependencies, new information flows, and new authority boundaries, which themselves can be attacked by an enemy. Greater information sharing can improve collective resilience, provided that sensitive information is protected, and trust is maintained.

The practical test for the coming year is whether organisations can shorten the time between knowing and acting. We should focus on measures that can be implemented quickly, tested in real operations, and improved over time. That means prioritising the services and dependencies that matter most, accepting that not everything can be perfected at once, and learning fast enough to stay ahead of the threat. We cannot wait.

This is an interim report. It is deliberately short and high-level, and it will be followed by a more detailed report later in the year. The next twelve months should be used to raise the quality and tempo of defence: stronger cyber hygiene, faster processes, more operational information sharing, governance that asks the right questions, and practical adoption of AI in cybersecurity work. The main conclusion is practical: the key advantage will belong to organisations that can turn knowledge into action faster, with clear responsibility, trusted collaboration, and disciplined use of AI.

Appendix



Methodology

This interim report follows a rapid but traceable methodology. It does not aim to develop an entirely new theory of AI-driven cyber threats. Instead, it starts from already published international reports and tests how much agreement exists around their most relevant claims.

The process began with a review of international literature from government agencies, AI security institutes, standards bodies, cybersecurity organisations, and frontier AI labs. Statements were distilled from those sources and grouped into four areas: problem description, current situation, forecast, and recommendations. The purpose was to identify what the literature broadly appears to agree on, where uncertainty remains and which conclusions are relevant for us in Sweden in a twelve-month horizon.

The distilled statements were assessed through an expert survey using the response categories Agree, Disagree and Abstain, with free-text comments for each segment. The responses were summarised and classified on a consensus basis. High-consensus findings could be used as the group's own position. Lower-consensus findings were treated more cautiously and, where relevant, attributed to the underlying literature.

The working group was used to test whether the problem description was reasonable, whether important aspects were missing, whether the forecast was well calibrated, and whether the proposed directions were feasible within the coming twelve months. The final report presents a concise synthesis of the areas where the group found sufficient agreement for an interim report.

References

AI Sweden. (2026). *AI and Cyber Resilience*. https://www.ai.se/en/ai-labs/secure-ai/ai-and-cyber-resilience/

Anthropic. (2025). *Disrupting AI-enabled espionage operations*. https://www.anthropic.com/news/disrupting-AI-espionage

Anthropic. (2026). *Statement on the US government directive to suspend access to Fable 5 and Mythos 5*. Anthropic. https://www.anthropic.com/news/fable-mythos-access

Australian Signals Directorate. (2024). *Annual Cyber Threat Report 2023–2024*. https://www.cyber.gov.au/about-us/view-all-content/reports-and-statistics/annual-cyber-threat-report-2023-2024

Canadian Centre for Cyber Security. (2024). *National Cyber Threat Assessment 2025–2026*. https://www.cyber.gc.ca/en/guidance/national-cyber-threat-assessment-2025-2026

Cybersecurity and Infrastructure Security Agency. (2023). *Roadmap for Artificial Intelligence*. https://www.cisa.gov/resources-tools/resources/roadmap-ai

IVA. (2022). *Cybersäkerhet för ökad konkurrenskraft*. Kungl. Ingenjörsvetenskapsakademien. https://www.iva.se/publicerat/rapport-cybersakerhet-for-okad-konkurrenskraft/

MIT Technology Review. (2026). *The Meta hack shows there is more to AI security than Mythos*. https://www.technologyreview.com/2026/06/05/1138437/the-meta-hack-shows-theres-more-to-ai-security-than-mythos/

National Institute of Standards and Technology. (2025). *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations*. NIST AI 100-2e2025. https://csrc.nist.gov/pubs/ai/100/2/e2025/final

National Cyber Security Centre Sweden (NCSC-SE). (n.d.). *10 rekommenderade säkerhetsåtgärder*. https://www.ncsc.se/sv/cybersakerhet/10-rekommenderade-sakerhetsatgarder/

National Cyber Security Centre New Zealand. (2026). *Cyber readiness in the Frontier AI era*. https://www.ncsc.govt.nz/protect-your-organisation/cyber-readiness-in-the-frontier-ai-era/

UK AI Security Institute. (2026). *Our evaluation of Claude Mythos Preview's cyber capabilities*. https://www.aisi.gov.uk/blog/our-evaluation-of-claude-mythos-previews-cyber-capabilities

UK National Cyber Security Centre. (2025). *Impact of AI on cyber threat from now to 2027*. https://www.ncsc.gov.uk/report/impact-ai-cyber-threat-now-2027

UK National Cyber Security Centre and UK AI Security Institute. (2026). *Why cyber defenders need to be ready for frontier AI*. https://www.ncsc.gov.uk/blogs/why-cyber-defenders-need-to-be-ready-for-frontier-ai

The Royal Swedish Academy of Engineering Sciences (IVA) is an independent academy whose mission is to promote the engineering and economic sciences and the advancement of business and industry. In cooperation with the business community and academia, IVA initiates and proposes measures to improve Sweden's industrial expertise and competitiveness. For more information about IVA and the Academy's projects, see the website www.iva.se.

Published by: The Royal Swedish Academy of Engineering Sciences (IVA), 2026
Box 5073, SE-102 42 Stockholm, Sweden
Tel: +46 (0)8 791 29 00

IVA-R 540
ISSN: 1100-5645
ISBN: 978-91-89181-86-1

Author: Musard Balliu, KTH
Project management: Per Hjertén, IVA
Graphic design: Pelle Isaksson, IVA

This report is available to download as a pdf file at www.iva.se

IVA's project **Swedish Futures** seeks to establish an overarching and coherent vision for Sweden as a leading nation in technology and innovation by the year 2035 – with a focus on competitiveness, sustainability, and security.

AI SWEDEN

IVA
Royal Swedish Academy of Engineering Sciences

NATIONELLT CYBERSÄKERHETSCENTER
En del av FRA