AI Governance Alliance

WORLD ECONOMIC FORUM

In collaboration with the Global Cyber Security Capacity Centre, University of Oxford

Transformation of Industries in the Age of AI

Artificial Intelligence and Cybersecurity: Balancing Risks and Rewards

WHITE PAPER

JANUARY 2025



Images: Getty Images

Contents

Reading guide 3

Foreword 4

Executive summary 5

Introduction: The scope 6

1 The context of AI adoption – from experimentation to full business integration 8

2 Emerging cybersecurity practice for AI 10

- 2.1 Shift left 11
- 2.2 Shift left and expand right 11
- 2.3 Shift left, expand right and repeat 11
- 2.4 Taking an enterprise view 11

3 Actions for senior leadership 12

4 Steps towards effective management of AI cyber risk 14

- 4.1 Understanding how the organization's context influences the AI cyber risk 14
- 4.2 Understanding the rewards 15
- 4.3 Identifying the potential risks and vulnerabilities 15
- 4.4 Assessing potential negative impacts to the business 17
- 4.5 Identifying options for risk mitigation 19
- 4.6 Balancing residual risk against the potential rewards 21
- 4.7 Repeat throughout the AI life cycle 21

Conclusion 22

Contributors 23

Endnotes 27

Disclaimer
This document is published by the World Economic Forum as a contribution to a project, insight area or interaction. The findings, interpretations and conclusions expressed herein are a result of a collaborative process facilitated and endorsed by the World Economic Forum but whose results do not necessarily represent the views of the World Economic Forum, nor the entirety of its Members, Partners or other stakeholders.

© 2024 World Economic Forum. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, including photocopying and recording, or by any information storage and retrieval system.

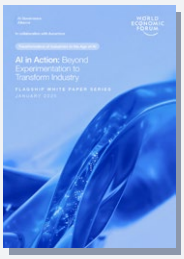
Reading guide

The World Economic Forum's AI Transformation of Industries initiative seeks to catalyse responsible industry transformation by exploring the strategic implications, opportunities and challenges of promoting artificial intelligence (AI)-driven innovation across business and operating models.

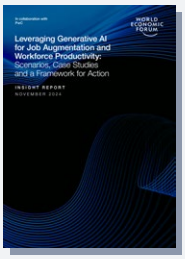
This white paper series explores the transformative role of AI across industries. It provides insights through both broad analyses and in-depth explorations of industry-specific and regional deep dives. The series includes:

Cross industry

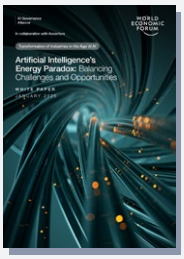
Impact on industrial ecosystems



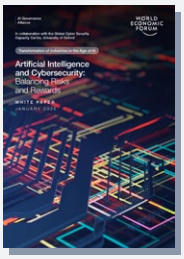
AI in Action: Beyond Experimentation to Transform Industry



Leveraging Generative AI for Job Augmentation and Workforce Productivity



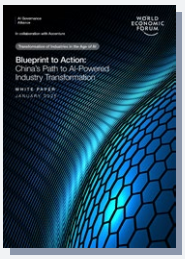
Artificial Intelligence's Energy Paradox: Balancing Challenges and Opportunities



Artificial Intelligence and Cybersecurity: Balancing Risks and Rewards

Regional specific

Impact on regions



Blueprint to Action: China's Path to AI-Powered Industry Transformation

Industry or function specific

Impact on industries, sectors and functions

Advanced manufacturing and supply chains



Frontier Technologies in Industrial Operations: The Rise of Artificial Intelligence Agents

Financial services



Artificial Intelligence in Financial Services

Media, entertainment and sport



Artificial Intelligence in Media, Entertainment and Sport

Healthcare



The Future of AI-Enabled Health: Leading the Way

Transport



Intelligent Transport, Greener Future: AI as a Catalyst to Decarbonize Global Logistics

Telecommunications

Upcoming industry report: Telecommunications

Consumer goods

Upcoming industry report: Consumer goods

Additional reports to be announced.

As AI continues to evolve at an unprecedented pace, each paper in this series captures a unique perspective on AI – including a detailed snapshot of the landscape at the time of writing. Recognizing that ongoing shifts and advancements are already in motion, the aim is to continuously deepen and update the understanding of AI's implications and applications through collaboration with the community of World Economic Forum partners and stakeholders engaged in AI strategy and implementation across organizations.

Together, these papers offer a comprehensive view of AI's current development and adoption, as well as a view of its future potential impact. Each paper can be read stand-alone or alongside the others, with common themes emerging across industries.

Foreword



Sadie Creese
Professor of Cybersecurity; Director and Technical Board Chair, Global Cyber Security Capacity Centre, University of Oxford



Jeremy Jurgens
Managing Director, World Economic Forum

Adoption of artificial intelligence (AI) is accelerating across the economy as organizations seek to harness its potential rewards. To support this, the AI Governance Alliance, launched by the World Economic Forum in June 2023, was established to provide guidance on the responsible design, development and deployment of AI systems.

Historically, insufficient attention has been given to the potential cybersecurity risks of AI adoption and use. This report highlights the steps that need to be taken to ensure that cybersecurity is fully embedded within the AI adoption life cycle.

Amid a business landscape that is increasingly focused on responsible innovation, this report offers a clear executive perspective on managing AI-related cyber risks. It empowers leaders to invest and innovate in AI with confidence, and exploit emerging opportunities for growth. To unlock full potential, it is essential to develop a comprehensive understanding of these cyber risks and related mitigation measures.

Throughout the report, we explore a central question: **How can organizations reap the benefits of AI adoption while mitigating the associated cybersecurity risks?**

This report provides a set of actions and guiding questions for business leaders, helping them to ensure that AI initiatives align with overall business goals and stay within the scope of organizations' risk tolerance.

It additionally offers a step-by-step approach to guide senior risk owners across businesses on the effective management of AI cyber risks. This approach includes: assessing the potential vulnerabilities and risks that AI adoption might create for an organization, evaluating the potential negative impacts to the business, identifying the controls required and balancing the residual risk against anticipated benefits.

Though focused on AI, the approach can be adapted for secure adoption of other emerging technologies.

This report draws on insights from a World Economic Forum initiative, developed in collaboration with the Global Cyber Security Capacity Centre (GCSCC) at the University of Oxford. Through collaborative workshops and interviews with cybersecurity and AI leaders from business, government, academia and civil society, participants explored key drivers of AI-related cyber risks and identified specific capability gaps that need to be addressed to secure AI adoption effectively.

Executive summary

A secure approach to AI adoption can allow organizations to innovate confidently.

AI technologies offer significant opportunities, and their application is becoming increasingly prevalent across the economy. As AI system compromise can have serious business impacts, organizations should adjust their approach to AI if they are to securely benefit from its adoption. Several foundational features capture best practices for securing and ensuring the resilience of AI systems:

1. Organizations need to apply a risk-based approach to AI adoption.

2. A wide range of stakeholders need to be involved in managing the risks end-to-end within the organization. A cross-disciplinary AI risk function is required, involving teams such as legal, cyber, compliance, technology, risk, human resources (HR), ethics and relevant front-line business units according to specific needs and contexts.

3. An inventory of AI applications can help organizations to assess how and where AI is being used within the organization, including whether it is part of the mission-critical supply chain, helping reduce "shadow AI" and risks related to the supply chain.

4. Organizations need to ensure adequate discipline in the transition from experimentation to operational use, especially in mission-critical applications.

5. Organizations should ensure that there is adequate investment in the essential cybersecurity controls needed to protect AI systems and ensure that they are prepared to respond to and recover from disruptions.

6. It is necessary to combine both pre-deployment security (i.e. the "security by design" principle – also called "shift left") and post-deployment measures to monitor and ensure resilience and recovery of the systems in use (referred to in this report as "expand right"). As the technology evolves, this approach needs to be repeated throughout the life cycle. This overall approach is described in the report as "shift left, expand right and repeat".

7. Technical controls around the AI systems themselves need to be complemented by people- and process-based controls on the interface between the technology and business operations.

8. Care needs to be paid to information governance – specifically, what data will be exposed to the AI and what controls are needed to ensure that organizational data policies are met.

It is crucial for top leaders to define key parameters for decision-making on AI adoption and associated cybersecurity concerns. This set of questions can guide them in assessing their strategies:

1. Has the appropriate risk tolerance for AI been established and is it understood by all risk owners?

2. Are risks weighed against rewards when new AI projects are considered?

3. Is there an effective process in place to govern and keep track of the deployment of AI projects?

4. Is there clear understanding of organization-specific vulnerabilities and cyber risks related to the use or adoption of AI technologies?

5. Is there clarity on which stakeholders need to be involved in assessing and mitigating the cyber risks of AI adoption?

6. Are there assurance processes in place to ensure that AI deployments are consistent with the organization's broader organizational policies and legal and regulatory obligations?

By prioritizing cybersecurity and mitigating risks, organizations can safeguard their investments in AI and support responsible innovation. A secure approach to AI adoption not only strengthens resilience but also reinforces the value and reliability of these powerful technologies.

Introduction: The scope

Cyber risks related to AI adoption have to be considered by business leaders and senior risk owners alike.

This report is part of a series exploring the transformative role of artificial intelligence (AI) across industrial ecosystems, along with cross-industry, industry-specific and regional perspectives. It is specifically focused on how organizations can reap the benefits of AI adoption while mitigating the associated cybersecurity risks.

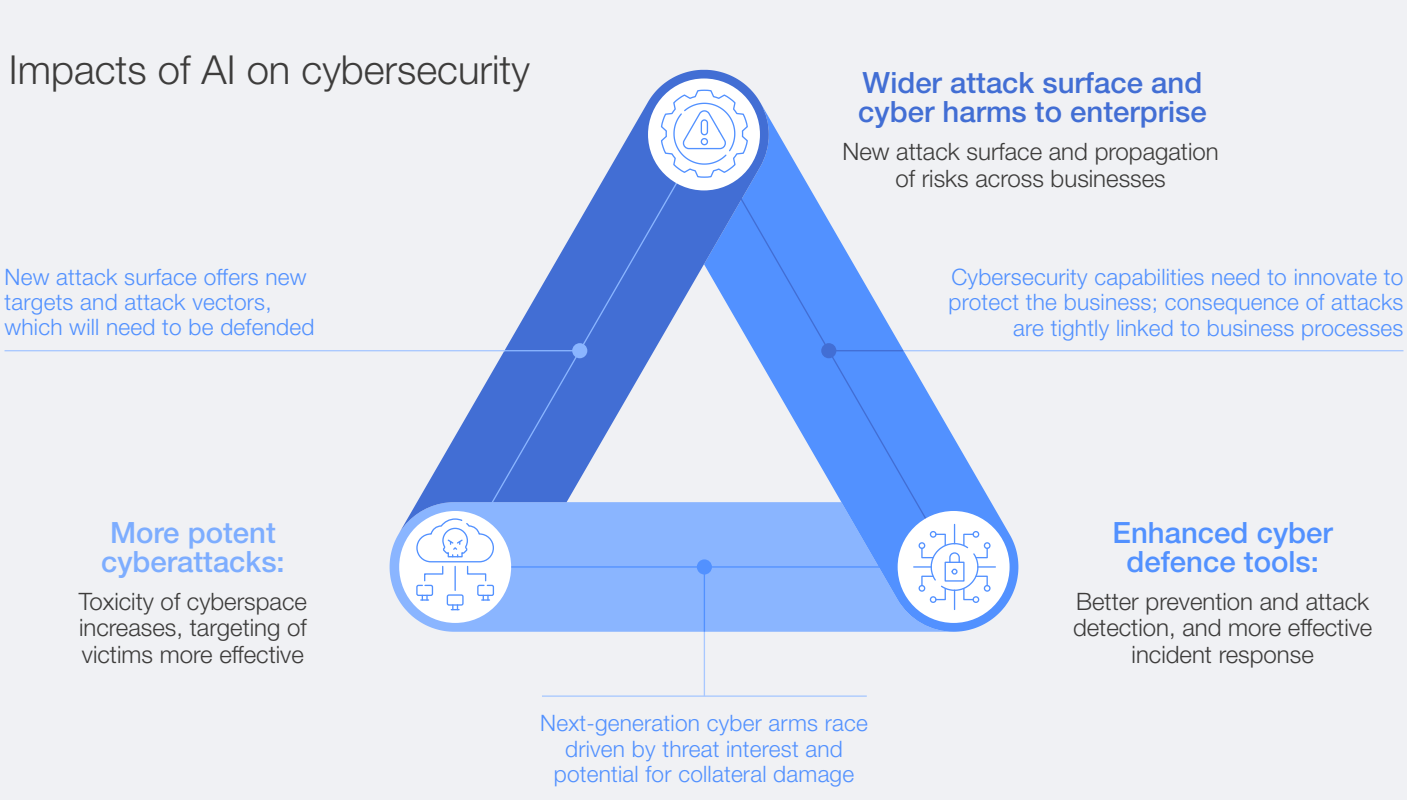
The business benefits of adopting AI can be considerable, but the cyber risks of embedding these technologies into an organization are not always considered from the outset. By adopting AI, businesses may find themselves vulnerable to new threats that they do not yet know how to defend themselves against.

The impact of AI on cybersecurity can be considered to fall into three broad categories:

- **The use of AI by threat actors**: Threat actors are using AI to enhance their capabilities and make their tactics, techniques and procedures more potent, and attacks more effective.
- **The use of AI by defenders**: In parallel, cyber defenders are harnessing AI to enhance cybersecurity capabilities, facilitating wider prevention, more accurate threat detection, autonomous remediation and more rapid and effective incident response.
- **Cybersecurity for AI**: The use of AI is creating an expanded attack surface that might be exploited by threat actors. Existing methods need to be extended to address new vulnerabilities that are inherent in AI, but that may not be as relevant for "classical" IT systems.

This report focuses on the third of these – namely, the need to adopt AI systems with due consideration for the emergent cyber risks. It contains guidance for business leaders and senior risk owners on managing the cyber risks associated with the implementation of AI technologies while innovating in their use of AI.

FIGURE 1 **The triangle of AI impacts on cybersecurity**



BOX 1 | The use of AI by threat actors

Cybercriminals can harness AI capabilities to amplify the scale, sophistication and speed of their malicious activities, presenting unprecedented challenges in cybersecurity defence.

- **Impersonation, social engineering and spear phishing**: The criminal use of AI has not only bolstered the scope and efficiency of cybercrime (including identity theft, fraud, data privacy violations and intellectual property breaches), but has also lowered the barriers to entry for criminal networks that previously lacked the technical skills.[1] A research study found that large language model (LLM)-automated phishing can lead to an over-95% reduction in costs, while maintaining or even exceeding previous success rates.[2]

- **Reconnaissance**: AI has enhanced reconnaissance efforts for cybercriminals by automating and refining the information-gathering process. Attackers can efficiently analyse vast amounts of data from various sources, such as by scraping social media, public records and network traffic to identify potential targets and vulnerabilities. Though not a novel use case, AI tools can process and correlate this data with greater speed and accuracy, making target selection and external surface scanning more efficient and effective.[3] For example, AI can detect and map out organizational structures, pinpoint weaknesses in security configurations and predict likely security behaviours and responses.

- **Discovering and exploiting zero-days**: AI allows cybercriminals to accelerate the process of discovering unpatched vulnerabilities such as zero-days – unknown vulnerabilities that do not have any patch or fix available – more efficiently and at scale. AI-enabled reconnaissance tools not only streamline the identification of zero-day vulnerabilities but also make it easier to create custom malware capable of exploiting these weaknesses before patches can be deployed. Researchers have also found that multiple GPT-4 models working in tandem are capable of autonomously exploiting zero-day vulnerabilities.[4]

- **Compromising AI systems**: This involves cybercriminals exploiting weaknesses in AI training datasets via data poisoning attacks,[5] model architectures and operational frameworks. Data poisoning can degrade a model's performance and reliability, leading to erroneous outputs[6] with far-reaching, sector-specific consequences. In the financial sector, for example, a successful data poisoning attack could manipulate algorithms used for credit scoring or fraud detection. Such outcomes not only undermine the integrity of systems, but also expose institutions to significant financial losses and reputational damage.

> **In the next decade, companies will be defined by their AI strategy: innovators will succeed, while resistors will vanish. Today's chief information security officers (CISOs) play a critical role in this journey, and must move from blocking the use of AI, to enabling it. But with the technology still in its infancy, the lack of understanding around AI has the potential to shift the balance of power to threat actors. The only viable defence is fighting AI with AI – developing personalized, adaptive security approaches that can protect an organization at speed and at scale.**
>
> Matthew Prince, CEO and Co-Founder, Cloudflare

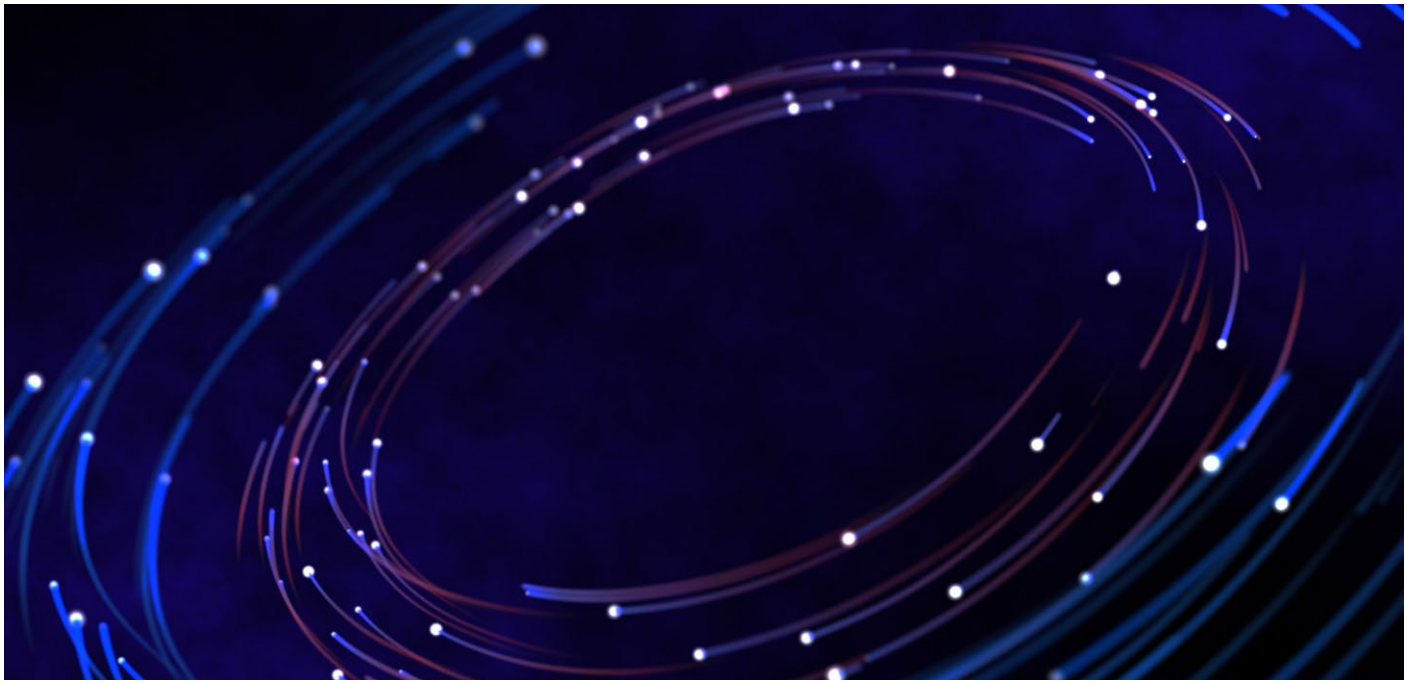
1 The context of AI adoption – from experimentation to full business integration

Understanding business context is essential for identifying the security needs of AI.

Cybersecurity requirements for AI technologies should be considered in tandem with business requirements. How a business is using AI should determine security needs, what to protect and when. There are numerous influencing factors that drive cybersecurity requirements, including: the criticality of the business processes and control systems using AI and the degree of dependency these processes have on the AI system outputs; the sensitivity of the data and devices that AI is involved in processing and controlling; and the risk culture of the organization and its approach to digital innovation.

Businesses are innovating with AI in a range of ways, and are at various stages in the adoption cycle:

- **Experimentation and piloting**: Much of current AI deployment by businesses is explorative or experimental. According to research from the AI Governance Alliance, organizations are commonly using "smaller, use-case-based approaches that emphasize ideation and experimentation".[7] There is, however, a risk of experiments becoming embedded within live business operations without the rigorous risk assessment, system testing and user training required.
- **Unconscious use of AI through product features (off-the-shelf software)**: For some organizations, the adoption process involves a more gradual – and at times passive – approach. Under this approach, AI is introduced in enterprise processes through new features or the enhancement of tools and platforms already available in an organization's ecosystem – e.g. enterprise resource planning (ERP), HR and IT management platforms. This process presents the risk of introducing shadow AI. A lack of formal roll-out programmes may decrease transparency, which can in turn weaken management processes and leadership oversight. Businesses require visibility and close coordination with vendors to assess AI feature capabilities and effectively evaluate potential risks. Furthermore, lax software management in organizations can amplify this type of risk due to the introduction of AI through unsanctioned or unmonitored tools (e.g. open source tools used by developers, browsers or software plugins).



- **Roll-out and integration into live operations**: Some organizations have already identified the business opportunities presented by AI and are moving to full deployment. However, they may not be conducting proper cyber risk assessments or implementing appropriate controls. Organizations need to ensure that there's adequate discipline around the transition from experimentation to operational use, especially in mission-critical applications. The cybersecurity market's ability to support specialized tools for protecting the confidentiality, integrity and availability of related systems and services may also not be mature enough to enable these organizations to implement AI systems securely.

- **Disparate projects across the organization**: In most large businesses, there are multiple projects exploring the use of AI across different functions and channels. These are not necessarily following a coordinated process, so assessment of risk to the business may not be sufficiently aligned. This applies to both full roll-out and gradual creep scenarios.

- **Hosted by third-party versus on-premises**: Often, businesses are using third-party AI services hosted in the cloud. Such operations do not absolve the business from managing cybersecurity of the AI assets, but they do change the mitigation controls available and create a need to negotiate appropriate protections from the suppliers.

- **Internal AI tools development**: Many organizations started offering AI features in their public digital services. Some of these are based on existing commercial or open- source tools. Others are developed internally. In either case, security requirements need to be properly established at the development stage.

Organizations may also be entering the decision-making process on risk at different stages:

- AI technologies may already have been embedded into the business processes or core assets. In this case, risk owners need to map what has been implemented and assess how to manage security retroactively.

- In other cases, the process might start with a risk-reward-based decision about whether to embed AI into operations or products. Under this approach, the AI system is only moved into the live environment when the rewards are determined to outweigh or justify the risks. This risk-reward-based decision necessitates a proactive approach to security, which can be integrated during the design phase.

> **AI holds enormous potential to advance the way people live and work, but we must ensure that we apply these powerful tools ethically and sustainably. Rapid advances in AI create opportunities but also introduce significant cybersecurity and governance challenges. As AI systems become more integrated into our lives, we must build secure AI platforms that protect against adversarial attacks and safeguard data integrity by following secure-by-design principles. Additionally, we need to introduce the appropriate level of governance in both development and usage to ensure trustworthy AI.**
>
> Antonio Neri, President and Chief Executive Officer, Hewlett Packard Enterprise

2

Emerging cybersecurity practice for AI

Securing AI systems demands early mitigation, ongoing operational security, enterprise-level risk management, and frequent reassessment of vulnerabilities.

While the understanding of attackers' and defenders' use of AI is well established, the recognition of the AI system as an asset to be protected is relatively new. Literature is emerging on the cybersecurity risks associated with AI systems. A range of initiatives are seeking to outline and categorize the cybersecurity threats and risks emerging from the use of AI, including from MITRE[8] and the UK National Cyber Security Centre (NCSC).[9] Emerging guidance and policies are highlighting requirements needed to address these risks, including (but not limited to):

- The Dubai AI Security Policy[10]
- The Cyber Security Agency (CSA) of Singapore's *Guidelines and Companion Guide on Securing AI Systems*[11]
- The UK Department for Science, Innovation and Technology's (DSIT's) developing AI Cyber Security Code of Practice[12]
- The National Institute of Standards and Technology's (NIST's) taxonomy of attacks and mitigations[13]
- The Open Worldwide Application Security Project's (OWASP) AI Exchange[14]

Simultaneously, evidence of real-world AI cybersecurity vulnerabilities, threats and incidents is being collected, and numerous repositories and databases are being created.[15]



2.1 Shift left

The question of how to secure AI is closely related to a wider body of work related to AI safety. This work is a significant aspect of the AI Governance Alliance's (AIGA's) agenda. This approach promotes the need to "shift left", i.e. implement safety guardrails early in the AI system life cycle (namely, at the building and pre-deployment stages) to mitigate related risks.[16] As an example of safe and secure-by-design AI technologies, it mandates the use of processes that address inherent vulnerabilities in the AI systems and services being used and procured by organizations.

2.2 Shift left and expand right

Not all risks can be mitigated at the building and pre-deployment stages. It is not possible to eliminate all system vulnerabilities, and there will always be threat actors who will succeed in circumventing the mitigating measures in place. To complement the security-by-design practices that help organizations develop AI technologies securely and ethically, businesses need to implement cybersecurity practices that will protect AI systems once they are in use.

This requires:

- An understanding of the wider risks faced by businesses using and depending on AI
- An understanding of the risks associated with the criticality of the data being processed
- Effective operational cybersecurity capabilities to protect against these risks and detect attacks
- Effective response and recovery processes to deal with incidents when they occur

In short, organizations will need to both "shift left and expand right".

2.3 Shift left, expand right and repeat

Alongside shifting left and expanding right, any approach for mitigating the cybersecurity risks associated with AI adoption needs to consider how the technology will evolve and how business use will change over time. This should be facilitated via repeated re-evaluation of risks and controls, alongside frequent rehearsal and regular testing of the organization's preparedness (e.g. war gaming, tabletop exercises, disaster recovery drills). This presents another opportunity to further integrate cyber risk assessment and intelligence capabilities into the resilience cycle and adjust testing strategies based on evolving AI risk profiles and threat actor developments observed across the industry. This means that leaders need to expand right, i.e. embed cyber resilience*,* and **repeat**.

2.4 Taking an enterprise view

AI systems do not exist in isolation. Organizations need to consider how the business processes and data flows built around AI systems can be designed in a way that reduces the business impact that a cybersecurity failure might cause. Where assurance on the security of underlying AI or on the effectiveness of defences is limited, it's crucial to consider how any compromise might be overcome. This could include implementing additional controls outside the system itself, or reviewing what data should or should not be exposed to the AI.

To enable such an end-to-end view, risks and controls need to be integrated into wider governance structures and enterprise risk management processes.

3

Actions for senior leadership

Leaders' decision-making on AI adoption should be guided by security considerations

Leaders are responsible for ensuring that adoption of AI technologies aligns with their organization's goals and objectives, and that the risks that arise fall within the scope of their organization's risk tolerance.

Cutting through the hype to understand risk and reward

Before making any decision to deploy AI into core operations, businesses need to ensure that the benefit is commensurate with costs and risks. To be sure of this, businesses need to take the potential risks of AI system failures (either accidental or due to malicious attacks) into account. Because of the speed of AI evolution, the risk-reward balancing decision may need to be reviewed on a frequent basis.

Promoting AI security-by-design and by-default

Because AI is rapidly evolving and security standards are relatively immature, business leaders should be aware that some products are likely to be less secure than others, and should therefore be approached with more caution. Leaders should demand robust third-party risk management and use the organization's purchasing power to promote AI security-by-design and by-default.

Embedding AI cyber risks into cross-organizational risk management

Managing AI-related cyber risks effectively requires a multidisciplinary approach. Technology and security teams alone cannot prevent incidents from occurring. Front-line business teams need to assess the potential business impacts, and specialists – e.g. in HR and/or legal teams – need to assess the potential liabilities that might arise. They have a significant role to play in establishing contingent mitigation. Such multidisciplinary arrangements may already be embedded within the organization's enterprise risk management. If not, they will need to be created bespoke to AI challenges.

Managing the decision-making process in a large organization can be complex. Some organizations may have a central AI policy, with divisional or local leadership responsible for decision-making within that policy. Smaller organizations may be able to operate a flatter governance structure, with decisions being made by the boardroom. In both cases, it is important to be very clear about where accountability for cyber risks sits.

Ensuring adequate investment in essential cybersecurity operations

Leaders need to ensure adequate investment in the cybersecurity controls and tools that are needed to protect AI systems, and ensure that the business is prepared to respond to and recover from disruptions. Chief information security officers need to be empowered to challenge both technology teams and business teams seeking to embed the technology within their operations. Security teams should be equipped with the necessary resources to adapt their capabilities and address new threats arising from AI use within the organization. Innovation investments for AI should be coupled with security investments to ensure that security is embedded throughout the AI system life cycle. This approach will help organizations define a reusable approach for mitigating complex technology risks, leaving them better prepared for future disruptions.

Engaging with national and sector-specific strategies and standards

Business leaders should be aware of the rapidly changing regulatory environment (particularly that relating to the markets they operate in). It will be necessary to consider how the specific local and regional AI contexts – including strategies and standards – impact business operations and risks. Additionally, relevant controls will need to be put in place to ensure businesses are meeting their obligations. For many, this will mean not only a watching brief on legal and regulatory compliance matters, but also on emerging threats and technological risks.

Questions for business leaders to consider

It is crucial for business leaders to define and communicate key parameters within which decision-making on AI adoption and its associated cybersecurity can be conducted. This set of questions is designed to guide them in assessing their current strategies, identifying potential vulnerabilities and cultivating a culture of security within their organizations.

1. Has the right risk tolerance for AI technologies been established and is it understood by all risk owners?

The organization might choose to be an early mover, recognizing the potential risks, or might take a more conservative approach. In both cases, there is a need to oversee the management of cybersecurity risks before, during and after the deployment of AI systems. The oversight and leadership scrutiny should generate evidence that AI risks are well understood, that stretch scenarios have been considered and that choices are in line with the wider risk tolerance of the business.

2. Is there a proper balancing of the risks against the rewards when new AI projects are considered?

It's crucial to assess how the potential upsides of AI projects align with the strategic direction of the business, when balanced against the novel risks these technologies might introduce. The potential rewards should be well qualified, and consideration should be given to the potential risks in any decision to use in operations.

3. Is there an effective process in place to govern and keep track of the deployment of AI projects within the organization?

This is particularly challenging in complex organizations in which experimentation and deployment may be occurring in multiple departments and subsidiaries. A clear process should be defined for making decisions on AI projects (including when to move them from experimentation to operational use). It is also important to monitor live AI systems to make sure users are not inadvertently exposing the organization to additional risk.

4. Is there a clear understanding of the organization-specific vulnerabilities and cyber risks related to the use or adoption of AI technologies?

There are novel vulnerabilities associated with AI technologies such as data-poisoning, inference engine sabotage and prompt jailbreaking. These could lead to operational disruption and data loss, or could exacerbate issues such as a lack of explainability and reliability, or potential for bias. A comprehensive risk assessment is required to identify the vulnerabilities of the AI systems and potential impact of compromise on the business. Timely access to relevant threat intelligence and advice will support greater situational awareness of the organization's risk exposure.

5. Is there clarity on which stakeholders within the organization need to be involved in assessing and mitigating the cyber risks from AI adoption?

There must be involvement from relevant front-line business teams, from legal, risk, audit and compliance, and from communications and technology. The various ways in which the AI is embedded into the operational and decision-making processes of the business need to account for the possibility of security failure, and mitigating controls put in place around deployment and operation need to limit the potential impact of adverse cyber events. The relevant accountable stakeholders should be identified. Clear responsibilities need to be set for AI-related cyber risks, and associated duties need to be clarified should a cyber incident occur.

6. Are there assurance processes in place to ensure that AI deployments are consistent with the organization's broader organizational policies and legal and regulatory obligations (for example relating to data protection or health and safety)?

Proposals for new AI deployments need to be tested to ensure compliance with wider organizational policies. Formal sign-off by relevant accountable stakeholders within the organization may be required. This review process will need to be revisited as the technology and its business use evolve.

4

Steps towards effective management of AI cyber risk

Evaluating the cyber risks resulting from AI adoption is essential for all organisations intending to innovate.

This chapter presents a set of steps for implementing oversight and control of cyber risks related to AI adoption and use. It is designed to be used by senior risk owners within an organization. The steps aim to guide the assessment of cybersecurity risks resulting from the adoption of AI technologies, and the implementation of the necessary mitigations.

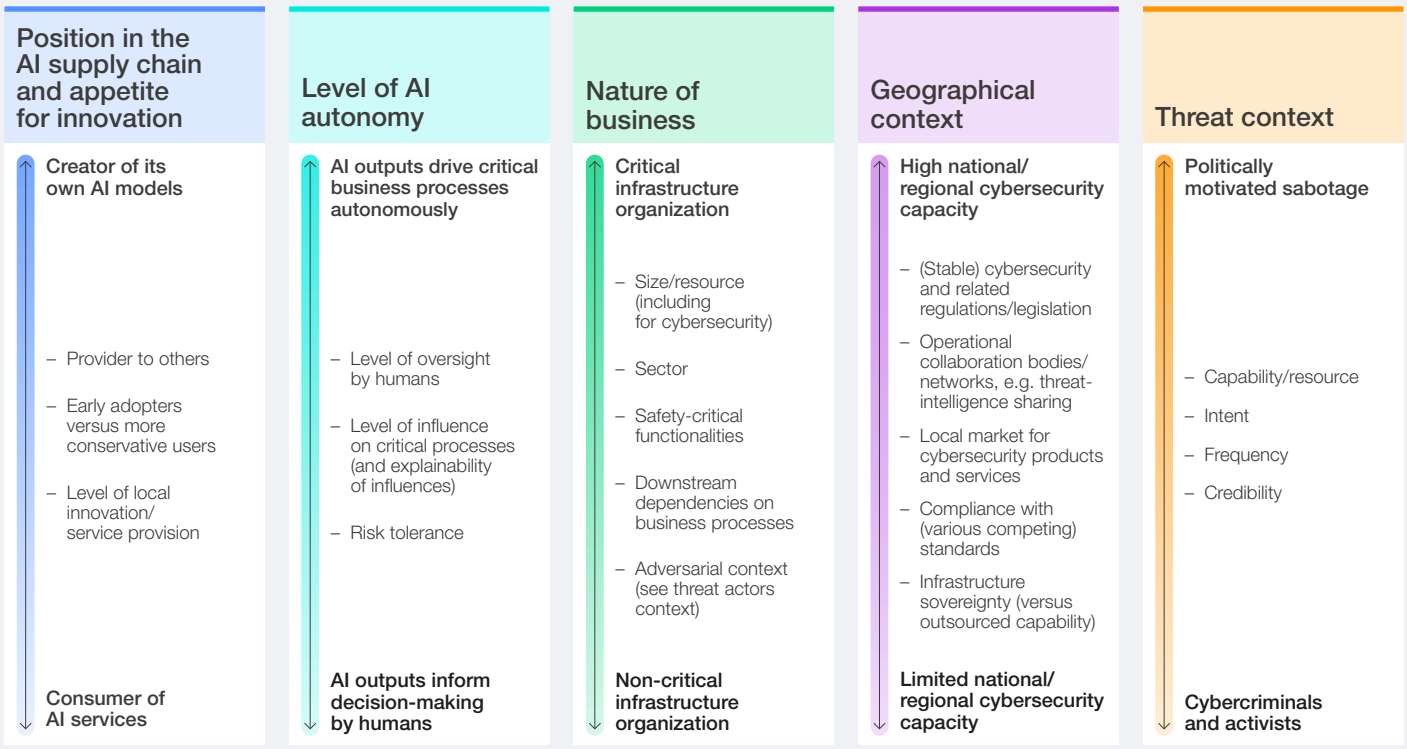
The decision-making process will, in many cases, be iterative. Senior risk owners should revisit risk-reward evaluations after analysing the potential impact scenarios. The process starts with an assessment of the AI risk context of the organization, and ends with the deployment of leading practices throughout the AI life cycle.

Step 1

Understanding how the organization's context influences the AI cyber risk

There are several contextual factors that may influence the risk exposure of organizations adopting AI:

FIGURE 2 **Characteristics influencing the cyber risks faced by organizations adopting AI**



Position in the supply chain and appetite for innovation: Organizations leading in AI innovation (either as sellers or consumers with market-leading capabilities) are likely to face risks from using newer technologies that may contain undiscovered vulnerabilities. More conservative users that procure more mature AI technologies may face fewer risks, as more will be known about vulnerabilities and effective control practices.

Nature of business: Which sectors the business operates in can affect their risk exposure. For example, critical national infrastructure organizations may be more likely to face high threat levels from attackers motivated by high harm potential or value, and to be subject to cybersecurity regulation. The size of the business could influence its resources for implementing AI risk mitigation, while the level of dependence from other businesses downstream affects the extent to which impacts of compromise might propagate.

Geographical context: Where the organization is conducting business will have a strong influence on their cybersecurity posture and residual cyber risk. The level of cybersecurity capacity of the country may influence the level of cybersecurity regulation that the organization is subject to. This might also affect the organization's access to a skilled professional workforce – though this might be less of an issue for large multinational organizations – and the availability of trusted sovereign cybersecurity infrastructures or threat/intelligence sharing channels.

Level of AI autonomy: Where autonomous AI drives business processes without human oversight, this may create greater risk. Lower risk is faced when there is little autonomy or strong human oversight to limit risk propagation.

Threat context: The type of threat actor faced by an organization determines the level of risk. More capable, resourced and motivated threat actors will create greater risk for potential victims.

It is necessary for organizations to consider how these risk contexts apply to them. This then informs later steps, during which the potential risks and impacts will be identified.

Step 2

Understanding the rewards

There may be a lack of clarity around the true benefits of AI technologies, as use cases are still in development, making accurate risk-reward analysis challenging. However, understanding the business drivers for the implementation of AI technologies will help to promote understanding of the expected rewards that are being sought. Research by the AI Governance Alliance has informed categorization of the opportunities that generative AI is perceived to be creating for businesses:[17]

- Enhancing enterprise productivity
- Creating new products or services
- Redefining industries and societies (e.g. making sectors such as healthcare more efficient and responsive to market changes – e.g. accelerating drug discovery).

It is essential to build understanding of the proposed integration of AI in the business. This should incorporate which systems, processes, information and data is involved, as well as which stakeholders and why.

Step 3

Identifying the potential risks and vulnerabilities

Key questions can help organizations to develop an understanding of the new risk exposure that the use of AI might bring:

1. What parts of the business might be dependent on AI and could be impacted should the AI systems be compromised?
2. What key business value, e.g. revenue, reputation, process efficiency, need to be protected?
3. Might the deployment of AI put crown jewels – assets of greatest value to the organization – and broader critical assets and processes at risk?
4. What new assets and processes related to the AI system itself need to be protected?

New technology brings the potential for new vulnerabilities. These typically fall into the following categories:

- Inherent software vulnerabilities
- Vulnerabilities introduced by humans' configuration and use of the technologies, particularly since this may require new and untrained practice
- Vulnerabilities in interfaces with other digital systems, e.g. weak links between software, hardware, operating system

Organizations need to develop an understanding of what vulnerabilities might be introduced as they adopt AI technologies, and of which security properties might be weakened should threat actors successfully exploit them.

Consider Figure 3, which details the potential areas of vulnerability of the AI system:

- The core AI infrastructure and supporting infrastructure that needs to be taken into consideration
- How this could expand attack surface and how this infrastructure might be compromised
- The security properties that must therefore be considered at risk

BOX 2

New tech, same need for security

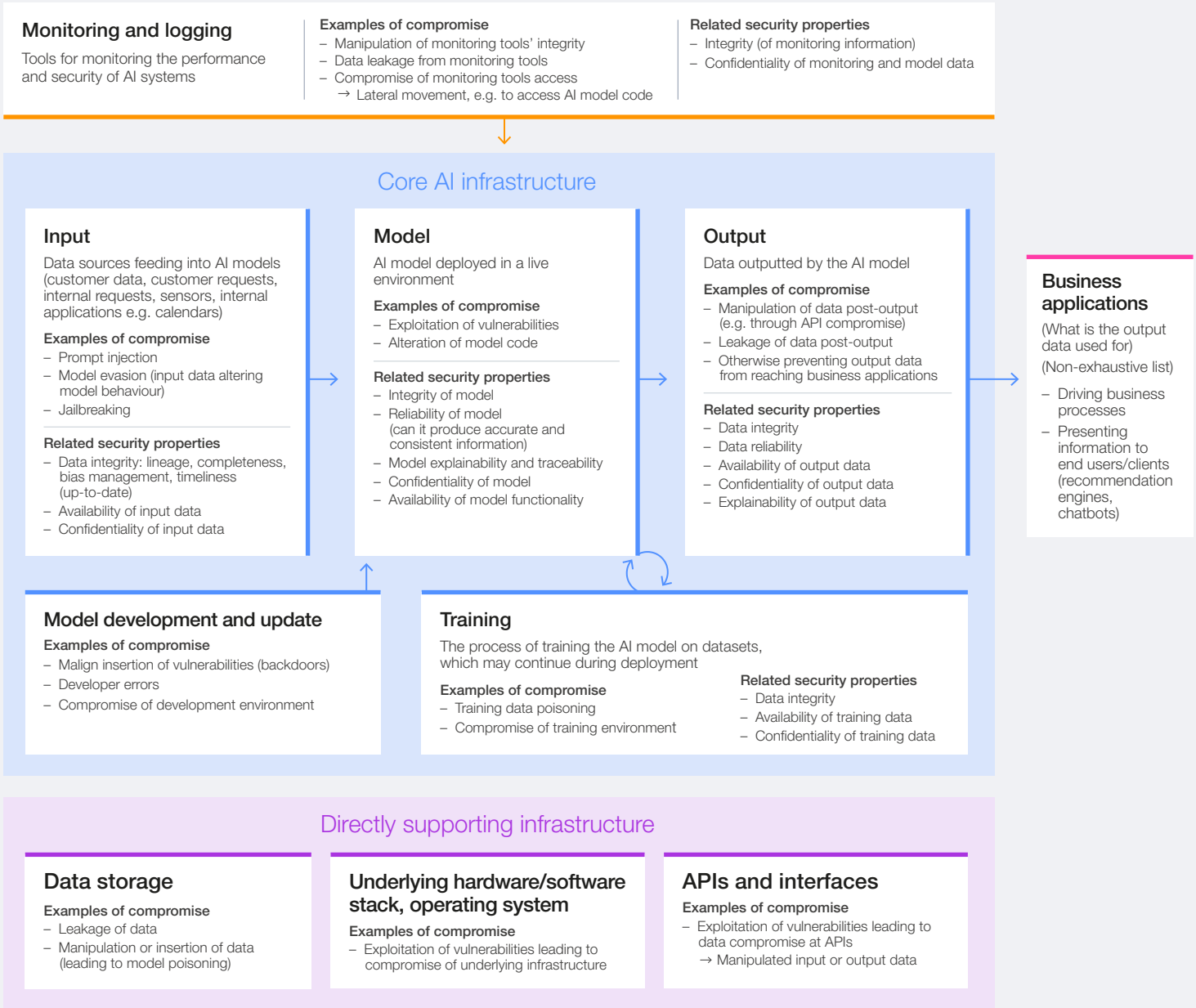
The traditional CIA triad remains critical: the compromise of AI systems and supporting infrastructure has the potential to impact on the Confidentiality, Integrity and Availability of data and assets. Other important security properties include:

- Explainability: refers to the concept that human users can comprehend the outputs generated by the AI model.
- Traceability: a property of the AI that signifies whether it allows users to track its processes – including understanding the data used and how it was processed by the models.

A lack of explainability or traceability may affect the organization's ability to investigate and mitigate against the impacts of an AI-system compromise.

FIGURE 3

AI system attack surface and security properties



Step 4

Assessing potential negative impacts to the business

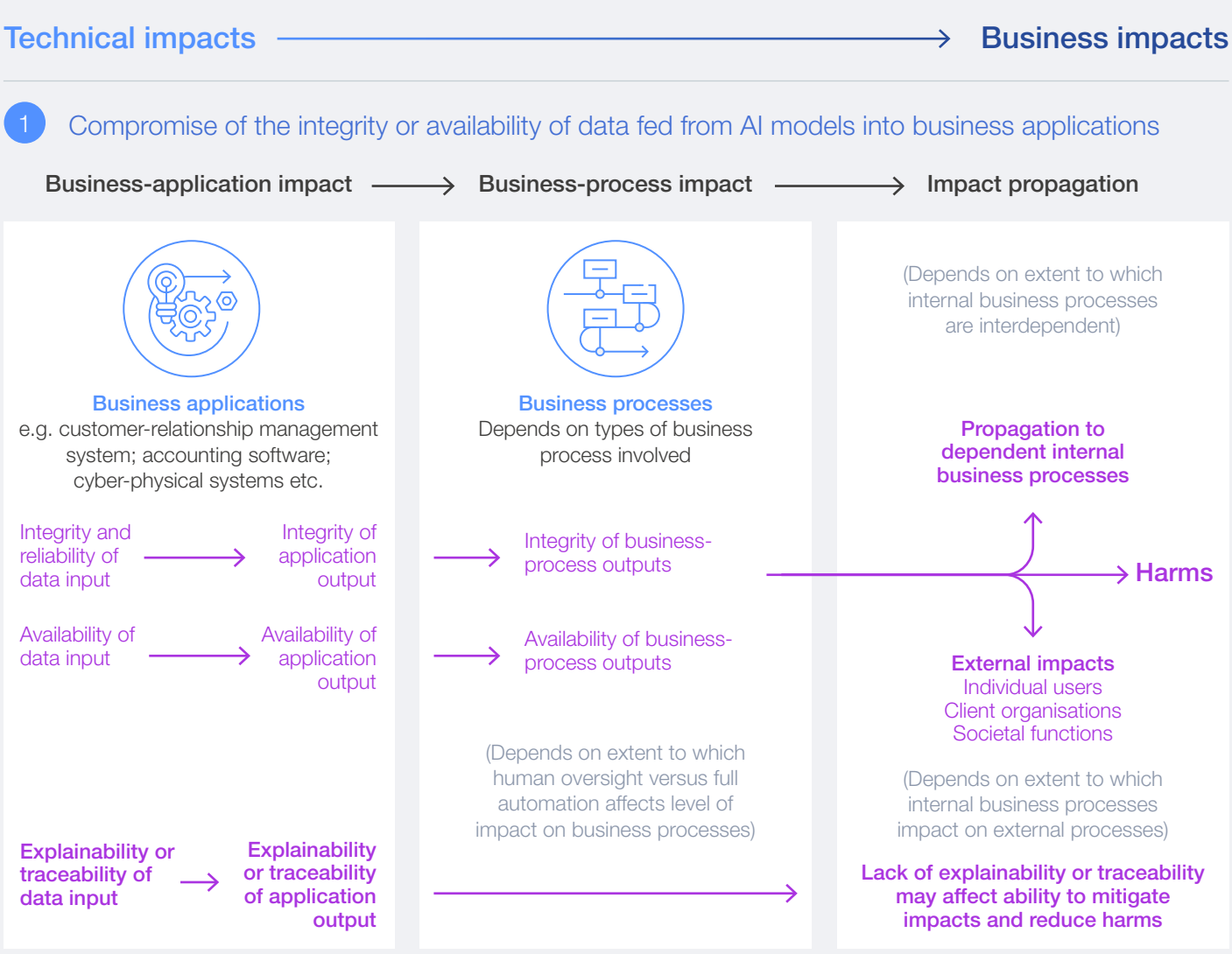
The negative impacts caused by the compromise of AI technologies may go beyond those associated with traditional cyber risks.

Key novel risks of AI-enabled business

1. Limited fairness due to inherent bias in products
2. Limited explainability of AI model, leading to reduced potential for human scrutiny
3. Unreliable outputs that decrease confidence and impede the ability to check the system reliability
4. New exploitable attack surface with limited controls
5. Privacy risks relating to personal data exposure via pattern-of-life generation
6. Exposure of confidential data through (possibly accidental) inclusion in AI training datasets

These risks can lead to negative impacts to the business, including reputational damage, loss of market position, loss of revenue, and legal and regulatory violations.

FIGURE 4 **Technical impacts of AI compromise can lead to business impacts**



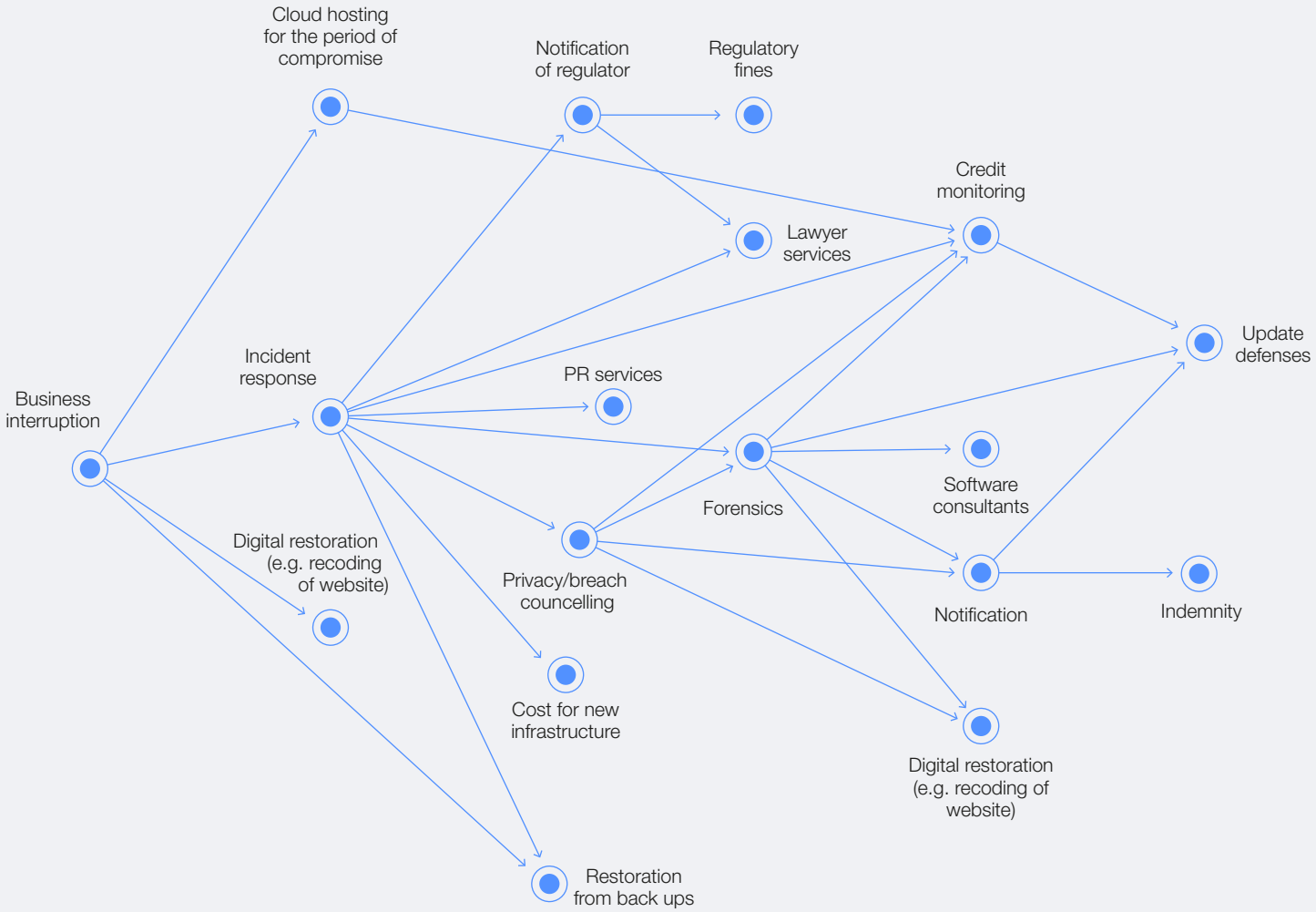
Harm-propagation trees

Attacks on AI systems can propagate further harms to businesses. They can also affect the wider ecosystem – for example, through impacts on downstream clients' processes or on societal processes that affect citizens. Analysing how an initial impact event might lead to further harms can strengthen resilience planning, as a more intricate set of events can be forecasted and planned for.

Harm-propagation trees are a tool for achieving this. They are a map of the negative consequences resulting from each event.[18] The process of creating a harm tree starts with identifying an initial impact event, and recording any impacts that could potentially result from it. Any further impacts that might result from these new impacts are then recorded in an iterative process.

Figure 5 shows an example of the harms a business might experience from an initial business interruption. The full scale of potential harms is broad, including the costs of incident response services such as legal and public relations (PR) services, forensics and breach counselling. It also includes other technical costs, such as for restoration and hosting during the period of compromise.

FIGURE 5 **Harm tree example. Initial impact: business interruption**



Source: Axon, L. et al. (2019). *2019 International Conference on Cyber Situational Awareness, Data Analytics And Assessment.*

Step 5

Identifying options for risk mitigation

Many existing cybersecurity control frameworks that are not AI-specific remain relevant for addressing cyber risks associated with AI adoption. What may differ is the way in which these controls need to be applied to protect the AI system, as well as any potential gaps they leave for specific risks.

Basic cyber hygiene is the foundation

It is critical to have a secure foundation of existing cybersecurity controls in place – i.e. basic cyber hygiene – to manage the cyber risks related to AI adoption. Some key practices include:

Avoiding vulnerabilities in the AI systems

Robust threat and vulnerability management practices help remediate critical exposures detected across systems, including AI technologies. It must also be complemented by secure configurations of the underlying hardware and software.

Limiting blast radius

Implementing controls for protecting the perimeters of systems – such as segmentation of networks and databases, and data-loss prevention – help limit the impact of an initial compromise of AI systems.

Accessing control

Ensuring that the AI systems and the infrastructure hosting AI algorithms and data are protected by access controls such as multi-factor authentication and strong privileged access management (PAM). These should be embedded as foundational security measures.

Third-party risk management

Strong procurement processes for assessing the security of AI models and training data are also critical to avoiding integrity issues and reducing cyber risk exposures.

Information sharing

Organizations should collaborate with peers – across businesses and governments – to ensure that threat- and incident-sharing mechanisms take AI-related cyber risks into account.

Education and awareness

Leaders need to develop an understanding of both the opportunities and risks associated with AI, and invest in training programmes to enhance AI awareness, create an organization-wide culture of responsible AI adoption and help employees recognize potential risks. Training should be tailored to the role of employees.

Mind the gaps: basic cyber hygiene is not enough

Some existing critical control capabilities will need to be tailored and updated in order to mitigate the cyber risks related to AI adoption, while other critical control capabilities will need to be developed from scratch to adequately mitigate the cyber risks of AI adoption. Examples of the former are set out in Table 1 and examples of the latter in Table 2.

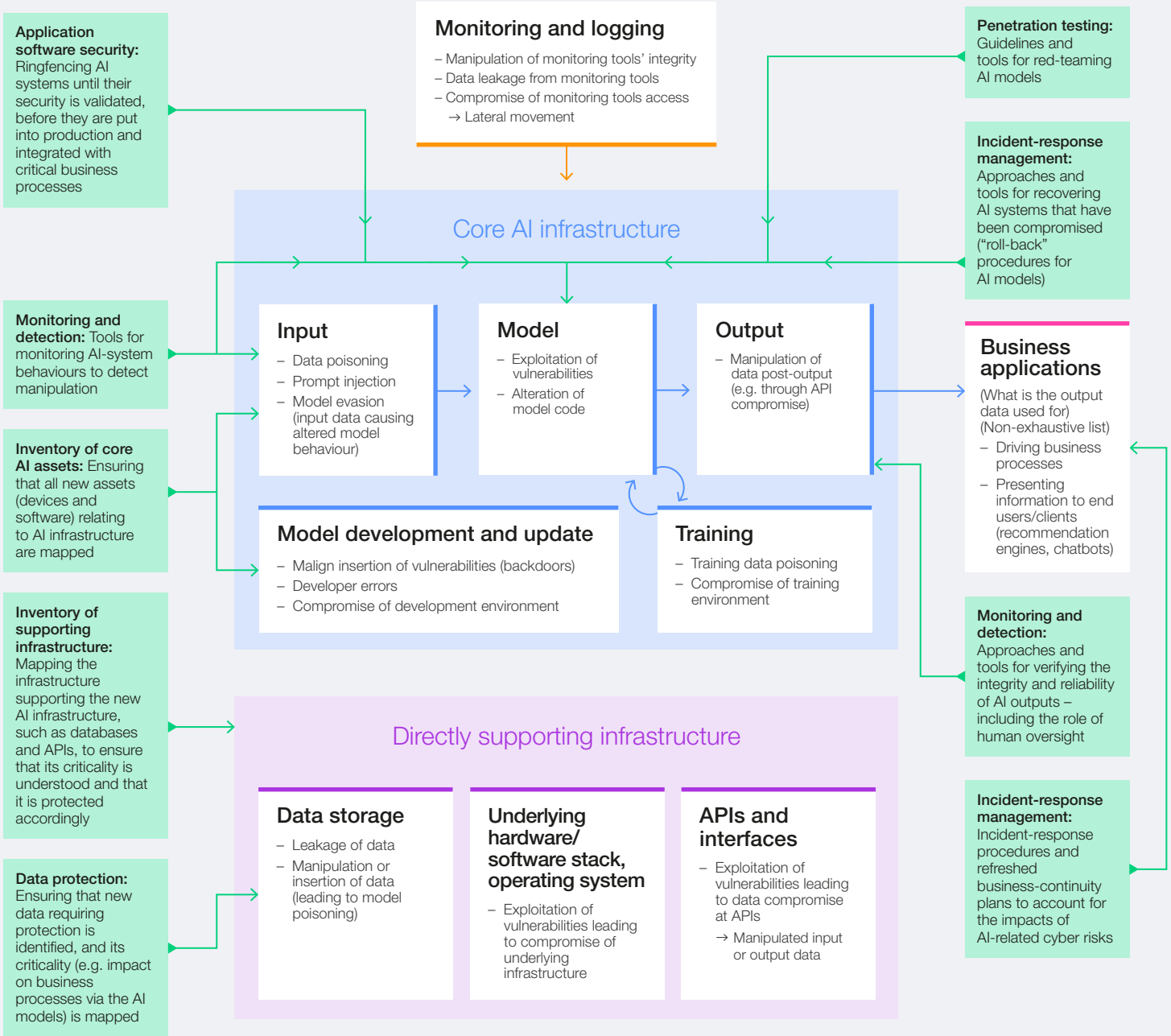
TABLE 1 **Example of existing control capabilities that need to be tailored**

Control	Description
Inventory of enterprise devices and software	Ensuring that all new assets (devices and software) relating to AI infrastructure (as well as the models) are inventoried
Business critical asset mapping	Mapping the infrastructure supporting the new AI system – including databases and application programming interfaces (APIs) – to ensure that its criticality is understood and that it is protected accordingly
Information governance	Ensuring that the application of AI to personal and other sensitive data does not undermine organizational information governance policies and data protection regulations
Pre-deployment integrity processes	Tailoring security-by-design processes (such as hardening, secure coding, etc.) specifically for AI data, inference models and technologies.
Business incident response strategy	Refreshing incident response procedures and business continuity plans to account for the impacts of AI-related cyber risks
Incident recovery tools and management	Updating tools and playbooks for recovering AI systems that have been compromised (e.g. “roll-back” procedures for AI models) Defining the criteria under which AI should be switched off, if possible
Exercising	Adapting the exercises with AI-related cybersecurity incidents to cover all major scenarios

TABLE 2 **Example of existing control capabilities that need to be developed**

Control	Description
Training data security	Data inputs need to be protected and managed to avoid deliberate poisoning and accidental damage to the AI system.
Prompt curation	Prompts need to be curated to mitigate risks of prompt injection and jailbreaking.
Output verification	The integrity and reliability of AI outputs need to be verified. Currently, this is mostly driven by humans.
Monitoring and detection	The behaviours of AI systems need to be monitored to detect manipulation in a timely manner.
Red teaming and adversarial testing	Guidelines and tools for red-teaming AI models, systems and processes using AI outputs are required. This is particularly critical for regulated sectors that already mandate such testing. AI systems could be harnessed to red-team AI models with greater efficacy.

FIGURE 6 **Application of risk controls to the attack surface**





Step 6

Balancing residual risk against the potential rewards

While implementing controls will reduce the cyber risk exposure of the organization, some residual risks are likely to remain. Decision-making on the adoption of AI should be informed by consideration of these risks in light of the potential rewards. Clarity on the qualified opportunity facilitates decision-making on residual risk exposure. Leadership needs to acknowledge the residual risk, and make a decision on whether to accept it or refuse it. In a case of refusal, additional controls will need to be put in place.

Step 7

Repeat throughout the AI life cycle

Threats are constantly evolving, so organizations will need to regularly review the steps outlined above to ensure they are properly positioned. These steps are meant to be an iterative process and not a one-time activity.

Conclusion

To fully benefit from the opportunities that AI technologies can bring, organizations need to ensure that the associated risks are proactively understood and managed. This is not a task that technology and security teams can perform in isolation. The process has to involve multiple stakeholder groups within the business, including top leadership and senior risk owners. Decision-making and investment choices need to be informed by proper evaluation of risks and rewards. The questions for business leaders and steps for senior risk owners outlined in this report highlight key considerations, and are designed to aid decision-making processes. They can be applied to help organizations ensure that the value from these technologies is realized and sustained.

AI and its associated risks are in constant evolution. As such, it is crucial that business leaders continuously update their understanding of the technology to keep up to date. Successful businesses will be well positioned to harness cybersecurity as a competitive advantage. In the context of AI adoption, this will enable organizations to innovate confidently and build trust in their services and brands.

Security leaders have an important role to play in aiding the secure adoption of AI technology across the wider economy. The community should collaborate on a global scale to develop and align AI security tools and standards that accommodate the diverse functionalities of different AI models. The community should also work together to exchange good practices in the secure deployment of AI systems, and in the protection of these systems (and their business interfaces) when in use. There is a need to enhance collaboration between the AI and cybersecurity communities, regulators and policy-makers through dialogues and joint initiatives. It will also be crucial to establish clear accountability mechanisms for securing the AI supply chain and provide effective incentives for security-by-design within AI products.

Lastly, it should be recognized that new tools and techniques are required to manage the novel security vulnerabilities driven by AI. While the market is maturing, remaining capability gaps should be addressed with some urgency.

Contributors

Lead authors

Louise Axon
Research Fellow, Global Cyber Security Capacity Centre, University of Oxford

Joanna Bouckaert
Community Lead, Centre for Cybersecurity, World Economic Forum

Sadie Creese
Professor, Cybersecurity, University of Oxford

Akshay Joshi
Head, Centre for Cybersecurity, World Economic Forum

Jamie Saunders
Oxford Martin Fellow, University of Oxford

Acknowledgements

Maria Basso
Digital Technologies Portfolio Manager, Centre for the Fourth Industrial Revolution Digital Technologies, World Economic Forum

Tal Goldstein
Head, Strategy and Policy, Centre for Cybersecurity, World Economic Forum

Jill Hoang
Initiatives Lead, AI and Digital Technologies, Centre for the Fourth Industrial Revolution Digital Technologies, World Economic Forum

Cathy Li
Head, AI, Data and Metaverse; Member, Executive Committee, World Economic Forum

Giulia Moschetta
Initiative Lead, Centre for Cybersecurity, World Economic Forum

We extend our thanks to all experts and leaders who contributed to the research:

Paige Adams
Group Chief Information Security Officer, Zurich Insurance Group

Bushra AlBlooshi
Senior Consultant, Research and Innovation, Dubai Electronic Security Center (DESC)

Hussain Aldawood
Cybersecurity Innovation & Partnerships Director, NEOM

Lampis Alevizos
Head, Cyber Defense Innovation, Volvo Group

Hessah Almajhad
Chief Cybersecurity Officer, Saudi Information Technology Company (SITE)

Doron Bar Shalom
Director, Strategic Product Innovation, Security, Microsoft

Alejandro Becerra
Digital Security Director, Telefonica HispAm

Mauricio Benavides
Chief Executive Officer, Metabase Q

Sarith Bhavan
Head, Cybersecurity and Technology Platform, Mubadala Investment Company

Janus Friis Bindslev
Chief Digital Risk Officer, PensionDanmark

Francesca Bosco
Chief of Staff, CyberPeace Institute

Jalal Bouhdada
Global Cybersecurity Director, DNV

Grant Bourzikas
Chief Security Officer, Cloudflare

Marijus Briedis
Chief Technology Officer, Nord Security

Niall Browne
Global Chief Security Officer, Palo Alto Networks

Ian Buffey
Chief Information Security Officer, AtkinsRéalis Group

Nicholas Butts
Director, Global Cybersecurity and AI/Emerging Tech Policy, Microsoft

Claudio Calvino
Senior Managing Director and Global Head, Data Science, FTI Consulting

David Caswell
Managing Director, Cyber-AI Leader for U.S. Government & Public Sector, Deloitte

Ronald Charron
Senior Cybersecurity Technology Advisor, Canadian Centre for Cyber Security

Piotr Ciepiela
Partner, Global/EMEIA Cyber Technologies Leader, EY

Claudionor Coelho
Chief AI Officer, Zscaler

Michael Daniel
President and Chief Executive Officer, Cyber Threat Alliance

Debashis Das
Principal, Office of the Chief Information Security Officer, Amazon Web Services

Maria del Rosario Romero
Head, IT Security, Pan American Energy

Tyler Derr
Chief Technology Officer, Broadridge Financial Solutions

Stefan Deutscher
Partner and Director, Cybersecurity and IT Infrastructure, Boston Consulting Group

Hazel Diez Castaño
Chief Information Security Officer, Banco Santander

Glenda Dsouza
Strategy Lead, Group Technology Office, Mahindra

Stephane Duguin
Chief Executive Officer, CyberPeace Institute

Gregory Eskins
Head, Global Cyber Insurance Center, Marsh McLennan

Sabrina Feng
Chief Risk Officer, Technology, Cyber and Resilience, London Stock Exchange Group

Sergio Fidalgo
Group Chief Security Officer and Group Chief Information Security Officer, BBVA

Bobby Ford
Senior Vice-President and Chief Security Officer, Hewlett Packard Enterprise

Shannan Fort
International Cyber Product Leader, Marsh McLennan

Simon Ganiere
Global Head, Cyber Intelligence Center, UBS

Javier Garcia Quintela
Chief Information Security Officer, Repsol

Akash Kumar Garg
Artificial Intelligence Technical Advice Production Lead, Australian Cyber Security Centre

Matan Getz
Chief Executive Officer, Aim Security

Jonathan Gill
Chief Executive Officer, Panaseer

Daniel Gisler
Chief Information Security Officer, Oerlikon Group

Pankaj Goyal
Chief Operating Officer, Safe Securities

Richard Hale
Senior Vice-President, Global Cyber Security Strategy, Sony

Randy Herold
Chief Information Security Officer, ManpowerGroup

Mark Hughes
Global Managing Partner, Cyber Security Services, IBM

Lars Idland
Vice-President, Security; Chief Information Security Officer, Equinor

Ann Irvine
Chief Data and Analytics Officer, Resilience

Amit Jain
Executive Vice-President and Head, Cybersecurity, HCLTech

Stefan Jäschke
Senior Vice-President and Head, Enterprise IT Security, Volvo Group

Ali El Kaafarani
Chief Executive Officer, PQShield

Mohit Kapoor
Group Chief Technology Officer, Mahindra

Steven Kelly
Chief Trust Officer, Institute for Security and Technology

Daniel Kendzior
Global Data and Artificial Intelligence (AI) Security Practice Lead, Accenture

Shaun Khalfan
Senior Vice-President and Chief Information Security Officer, PayPal

Hoda Al Khzaimi
Director, Centre for Cybersecurity, New York University Abu Dhabi

Sigmund Kristiansen
Chief Cyber Security Officer, Aker BP

Georgios Kryparos
Chief Information Security Officer, Einride

Ayelet Kutner
Chief Technology Officer, At-Bay

Christine Lai
AI Security Lead, Cybersecurity and Infrastructure Security Agency

Aamir Lakhani
Global Strategist and Architect, Fortinet

Philomena Lavery
Global Chief Information Security Officer, AVEVA

Jason Lee
Chief Information Security Officer, Splunk a Cisco Company

Simon Leech
Director, Cybersecurity Center of Excellence, Hewlett Packard Enterprise

Chris Lyth
Chief Information Security Officer, Arup Group

David Mabry
Vice-President and Chief Information Security Officer, Gulfstream Aerospace

Derek Manky
Chief Security Strategist and Global Vice-President, Threat Intelligence, Fortinet

Clemens Meiser
Division AI and Security, German Federal Office for Information Security

Eduardo Melendez
Chief Information Security Officer, Grupo Salinas

Michael Meli
Group Chief Information Security Officer and Managing Director, Julius Baer

Eiichiro Mitani
Executive Officer, Chief Information Officer, Mitsubishi Electric

Paulo Moniz
Head, CyberSecurity and Information Technology Risk, Energias de Portugal (EDP)

Sean Morton
Senior Vice President, Strategy & Services, Trellix

Barbara O'Neill
Global Chief Information Security Officer, EY

Mark Orsi
Chief Executive Officer, Global Resilience Federation

Christine Palmer
AI Expert, US Department of Homeland Security

Periklis Papadopoulos
AI Security Strategist, Accenture

Tom Parker
Chief Executive Officer, Hubble Technology

Pankaj Paul
Director, Strategy and Innovation, Burjeel Holdings

Sriram Ramachandran
Chief Risk Officer, Mahindra

Amanda Reath
Director, Cyber Programme Management, Canadian Centre for Cyber Security

Philip Reiner
Chief Executive Officer, Institute for Security and Technology

Cyril Reol
Group Chief Information Officer, Mercuria

Craig Rice
Chief Executive Officer, Cyber Defence Alliance

Harold Rivas
Chief Information Security Officer, Trellix

Jason Ruger
Chief Information Security Officer, Lenovo

Sreekumar S
Global Head, Practise for Cybersecurity, HCLTech

Amir Abdul Samad
Head, Cybersecurity, PETRONAS

Miguel Sanchez San Venancio
Global Chief Security and Intelligence Officer, Telefónica

Stephen Scharf
Managing Director and Chief Information Security Officer, Blackrock

Ralf Schneider
Senior Fellow and Head, Cybersecurity and NextGenIT Think Tank, Allianz

Tomer Schwartz
Co-Founder and Chief Technology Officer, Dazz

Leo Simonovich
Vice-President; Global Head, Industrial Cyber and Digital Security, Siemens Energy

Charley Snyder
Head, Security Policy, Google

Colin Soutar
Managing Director, Cyber Risk, Deloitte

Emanuele Spagnoli
Chief Information Security Officer, Mundys

Mark Stamford
Founder and Chief Executive Officer, OccamSec

Mark Swift
Chief Information Security Officer, Trafigura Group

Neha Taneja
Chief Information Security Officer, Hero Group

Jennifer Tang
Associate, Cybersecurity and Emerging Technologies, Institute for Security and Technology (IST)

Omar Al-Thukair
Vice President and Chief Digital Officer, Saudi Aramco

Ian Tien
Chief Executive Officer, Mattermost

Phil Tonkin
Field Chief Technology Officer, Dragos

Abdullah Al Turaifi
Head, AI Cybersecurity, Saudi Aramco

Swantje Westpfahl
Director, Institute for Security and Safety

Fabian Willi
Head, Cyber Key Accounts, Swiss Re

Rainer Zahner
Head, Cybersecurity Governance & Cyber Risk Management, Siemens

Jelena Zelenovic Matone
Chief Information Security Officer, European Investment Bank

Raphael Zimmer
Head, Division, AI and Security, German Federal Office for Information Security

Cyber Security Agency of Singapore

Israel National Cyber Directorate

We additionally thank the World Economic Forum's Partnership against Cybercrime (PAC) community members for their insights on AI and cybercrime.

Production

Phoebe Barker
Designer, Studio Miko

Louis Chaplin
Editor, Studio Miko

Laurence Denmark
Creative Director, Studio Miko

Endnotes

1. Federal Bureau of Investigation (FBI) San Francisco. (2024). *FBI Warns of Increasing Threat of Cyber Criminals Utilizing Artificial Intelligence.* https://www.fbi.gov/contact-us/field-offices/sanfrancisco/news/fbi-warns-of-increasing-threat-of-cyber-criminals-utilizing-artificial-intelligence; United Nations Office on Drugs and Crime. (2024). *Transnational Organized Crime and the Convergence of Cyber-Enabled Fraud, Underground Banking and Technological Innovation in Southeast Asia: A Shifting Threat Landscape.* https://www.unodc.org/roseap/uploads/documents/Publications/2024/TOC_Convergence_Report_2024.pdf.
2. Heiding, F., Schneier, B. & Vishwanath, A. (2024). AI Will Increase the Quantity – and Quality – of Phishing Scams. *Harvard Business Review*. https://hbr.org/2024/05/ai-will-increase-the-quantity-and-quality-of-phishing-scams.
3. Cantos, M., Riddell, S. & Revelli, A. (2023). *Threat Actors are Interested in Generative AI, but Use Remains Limited.* Google Cloud. https://cloud.google.com/blog/topics/threat-intelligence/threat-actors-generative-ai-limited.
4. Fang, R. et al. (2024). Teams of LLM Agents can Exploit Zero-Day Vulnerabilities. *Arxiv.* https://arxiv.org/abs/2406.01637.
5. Oprea, A., Fordyce, A. & Andersen, H. (2024). *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations.* National Institute of Standards and Technology. https://www.nist.gov/publications/adversarial-machine-learning-taxonomy-and-terminology-attacks-and-mitigations.
6. Cantos, M. (2019). Breaking the Bank: Weakness in Financial AI Applications. *Google Cloud Blog.* https://cloud.google.com/blog/topics/threat-intelligence/breaking-bank-weakness-financial-ai-applications/.
7. World Economic Forum. (2024). *Unlocking Value from Generative AI: Guidance for Responsible Transformation.* https://www3.weforum.org/docs/WEF_Unlocking_Value_from_Generative_AI_2024.pdf.
8. MITRE Adversarial Threat Landscape for Artificial-Intelligence Systems (ATLAS). (n.d.). *Home*. https://atlas.mitre.org/.
9. UK National Cyber Security Centre. (n.d.). *The near-term impact of AI on the cyber threat*. https://www.ncsc.gov.uk/report/impact-of-ai-on-cyber-threat.
10. Government of Dubai. (n.d.). *Dubai Electronic Security Center launches the Dubai AI Security Policy.* https://www.desc.gov.ae/dubai-electronic-security-center-launches-the-dubai-ai-security-policy/.
11. Cyber Security Agency of Singapore. (2024). *Guidelines and Companion Guide on Securing AI Systems.* https://www.csa.gov.sg/Tips-Resource/publications/2024/guidelines-on-securing-ai.
12. UK Department for Science, Innovation & Technology. (2024). *Call for views on the Cyber Security of AI.* https://www.gov.uk/government/calls-for-evidence/call-for-views-on-the-cyber-security-of-ai/call-for-views-on-the-cyber-security-of-ai.
13. Oprea, A., Fordyce, A. & Andersen, H. (2024). *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations.* National Institute of Standards and Technology. https://www.nist.gov/publications/adversarial-machine-learning-taxonomy-and-terminology-attacks-and-mitigations.
14. Open Worldwide Application Security Process. (2024). *AI Exchange*. https://owaspai.org/.
15. AI Risk and Vulnerability Alliance. (2024). *AI Vulnerability Database.* https://avidml.org/; Organisation for Economic Co-operation and Development (OECD) Policy Observatory. (2024). *OECD AI Incidents Monitor*. https://oecd.ai/en/incidents-methodology; Open AI. (2024). *Disrupting malicious uses of AI by state-affiliated threat actors.* https://openai.com/index/disrupting-malicious-uses-of-ai-by-state-affiliated-threat-actors/.
16. World Economic Forum. (2024). *Presidio AI Framework: Towards Safe Generative AI Models.* https://www3.weforum.org/docs/WEF_Presidio_AI%20Framework_2024.pdf.
17. World Economic Forum. (2024). *Unlocking Value from Generative AI: Guidance for Responsible Transformation.* https://www3.weforum.org/docs/WEF_Unlocking_Value_from_Generative_AI_2024.pdf.
18. Axon, L. et al. (2019). *Analysing cyber-insurance claims to design harm-propagation trees.* https://ora.ox.ac.uk/objects/uuid:496b5fb7-9da3-4305-a0b1-e4cbf0c41bfb/files/m75e3108c23c67618e62a642ac8c3f8f8.

WORLD ECONOMIC FORUM

COMMITTED TO IMPROVING THE STATE OF THE WORLD

The World Economic Forum, committed to improving the state of the world, is the International Organization for Public-Private Cooperation.

The Forum engages the foremost political, business and other leaders of society to shape global, regional and industry agendas.

World Economic Forum
91–93 route de la Capite
CH-1223 Cologny/Geneva
Switzerland

Tel.: +41 (0) 22 869 1212
Fax: +41 (0) 22 786 2744
contact@weforum.org
www.weforum.org