In collaboration with KPMG

WORLD ECONOMIC FORUM

Empowering Defenders: AI for Cybersecurity

WHITE PAPER

MAY 2026

Images: Getty Images

Contents

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.

© 2026 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.

Foreword

Akshay Joshi
Head of the Centre for Cybersecurity; Member of the Executive Committee, World Economic Forum

Laurent Gobbi
Partner, Global Head of Cyber & Tech Risk, KPMG

With the advent of the Intelligent Age, artificial intelligence (AI) and cybersecurity have become inextricably linked. In 2025, the World Economic Forum published *Artificial Intelligence and Cybersecurity: Balancing Risks and Rewards*, which highlighted the inherent cybersecurity risks associated with AI adoption and provided organizations with a roadmap to mitigate them. In this paper, the focus shifts to the transformative potential of AI in strengthening cybersecurity defences.

This white paper examines the application of AI technologies across the cybersecurity life cycle, drawing on real-world examples from World Economic Forum partner organizations. It outlines the strategic considerations and practical steps that executives and chief information security officers (CISOs) must undertake to enable the effective adoption and deployment of AI in cybersecurity.

Success will depend on key foundational elements, including strong executive support, high-quality data, a skilled workforce and integrated infrastructure. As automation becomes increasingly prevalent, preserving human judgement and expertise remains essential to mitigate potential systemic fragility.

Developed by the World Economic Forum's Cyber Frontiers initiative in collaboration with KPMG, this paper brings together contributions from leading experts in cybersecurity and AI. Through a series of workshops, these leaders shared insights into how AI can be harnessed effectively to enhance cyber resilience. As AI becomes further embedded within cybersecurity operations, strategic alignment will be critical to staying ahead of increasingly sophisticated cyber threats and to supporting sustainable growth in the digital economy.

Executive summary

While AI can enhance cyber defences, its effectiveness relies on having a clear strategy grounded in human oversight and judgement.

Artificial intelligence (AI) is reshaping the cybersecurity landscape. Attackers are increasingly using AI to increase the speed, scale and sophistication of threats. To address these evolving risks, cybersecurity must keep pace with the growing speed and sophistication of modern attacks. AI-driven tools are becoming central across the cybersecurity life cycle, from detecting and preventing incidents to response and recovery, enabling organizations to secure their digital assets and sensitive information more effectively.

This paper focuses on the use of AI by defenders, thereby providing actionable guidance to organizations on how to deploy AI to enhance defensive capabilities.

Real-world applications of AI in cybersecurity, illustrated through case studies submitted by World Economic Forum partners, demonstrate, for example, the use of AI to improve vulnerability detection, curate threat intelligence data and enhance defences against phishing and other attack vectors. This reflects a shift to operational adoption of AI solutions delivering measurable improvements in security performance.

Executive and cyber leaders embarking on the AI adoption journey for cyber defence should therefore:

- Align the adoption of AI in cybersecurity with organizational strategic priorities
- Establish organizational readiness across processes, data, infrastructure, skills and governance before deploying AI in cybersecurity
- Validate AI solutions through structured pilots prior to full deployment
- Scale and monitor the performance of AI in cybersecurity and optimize as needed

With the right approach in place, organizations can harness AI technologies to strengthen defences today, while building the agility to keep pace with evolving cyberthreats.

Introduction

The deployment of AI in cybersecurity is accelerating, although the pace and depth of adoption vary considerably.

AI is no longer a peripheral innovation; it is redefining cybersecurity by enabling advanced defence capabilities, creating new requirements to secure AI-driven systems and amplifying the tools available to adversaries. In fact, its strategic importance is reflected in the findings of the *Global Cybersecurity Outlook 2026*, which identifies AI as the most significant driver of change in cybersecurity, according to 94% of survey respondents.

While 77% of organizations report using AI in cybersecurity, deployment is closely tied to organizational size and resources.[1] Larger companies, supported by greater technical maturity and investment capacity, report higher adoption rates, whereas smaller entities, governments and non-governmental organizations (NGOs) tend to lag behind due to financial constraints, skills availability and data maturity.

When deployed effectively, AI delivers measurable operational and financial benefits: 88% of security teams report time saving and greater opportunity for proactive defence.[2] Organizations using AI extensively in security shortened breach times by approximately 80 days and reduced average breach costs by $1.9 million.[3] More broadly, AI helps address the structural challenges faced by cybersecurity, including the rising volume and sophistication of attacks, persistent talent shortages and increasing system complexity.

Yet progress on the defensive side is unfolding in parallel with a rapidly evolving threat environment, where adversaries are increasingly operating at machine speed, using AI to conduct reconnaissance of targets and vulnerabilities, generate malware, exploit code, evade detection and launch attacks at scale. What once required weeks of effort can now be executed in minutes, lowering technical barriers and dramatically expanding both the volume and impact of cyberattacks.[4]

Against this backdrop, AI has become a core enabler of modern cybersecurity. Its value lies not just in automation but in augmenting human cognition and expertise, accelerating detection and decision-making and strengthening organizational preparedness across technical, operational and governance dimensions.

Why is AI for cybersecurity a strategic necessity?

The rising complexity, volume and speed of cyberthreats is outpacing traditional defences, requiring new approaches to protect interconnected systems. AI has emerged as both a technical and strategic capability to address these challenges, enabling organizations to strengthen defensive operations and meet regulatory demands. Several factors underscore why AI has become essential for modern cyber defence:

Defensive asymmetry: Attackers need only find one entry point while defenders must defend everything. With AI, attackers can identify vulnerabilities faster. Defenders, however, can use AI to analyse and prioritize risks using internal proprietary data for contextual precision that attackers lack, thereby regaining strategic advantage.[5]

Complexity and scale: Modern digital ecosystems are highly interconnected, creating large attack surfaces and hidden vulnerabilities. Manual analysis is no longer sufficient. AI can operationalize and correlate data in real time, identifying configuration weaknesses and vulnerabilities at a scale and speed beyond human capability.

Operational burden: Security teams are often overwhelmed by alert overload and repetitive tasks, with 76% of professionals reporting exhaustion in 2025.[6] AI can automate routine processes and prioritize and filter alerts, freeing teams to focus on high-value, proactive cybersecurity activities.

Resource constraints: Expanding digital footprints often outpace cybersecurity budgets and staffing. In 2025, 53% of teams reported underfunding and 55% reported understaffing.[7] AI can help augment human expertise, scale security operations efficiently and help address talent gaps.

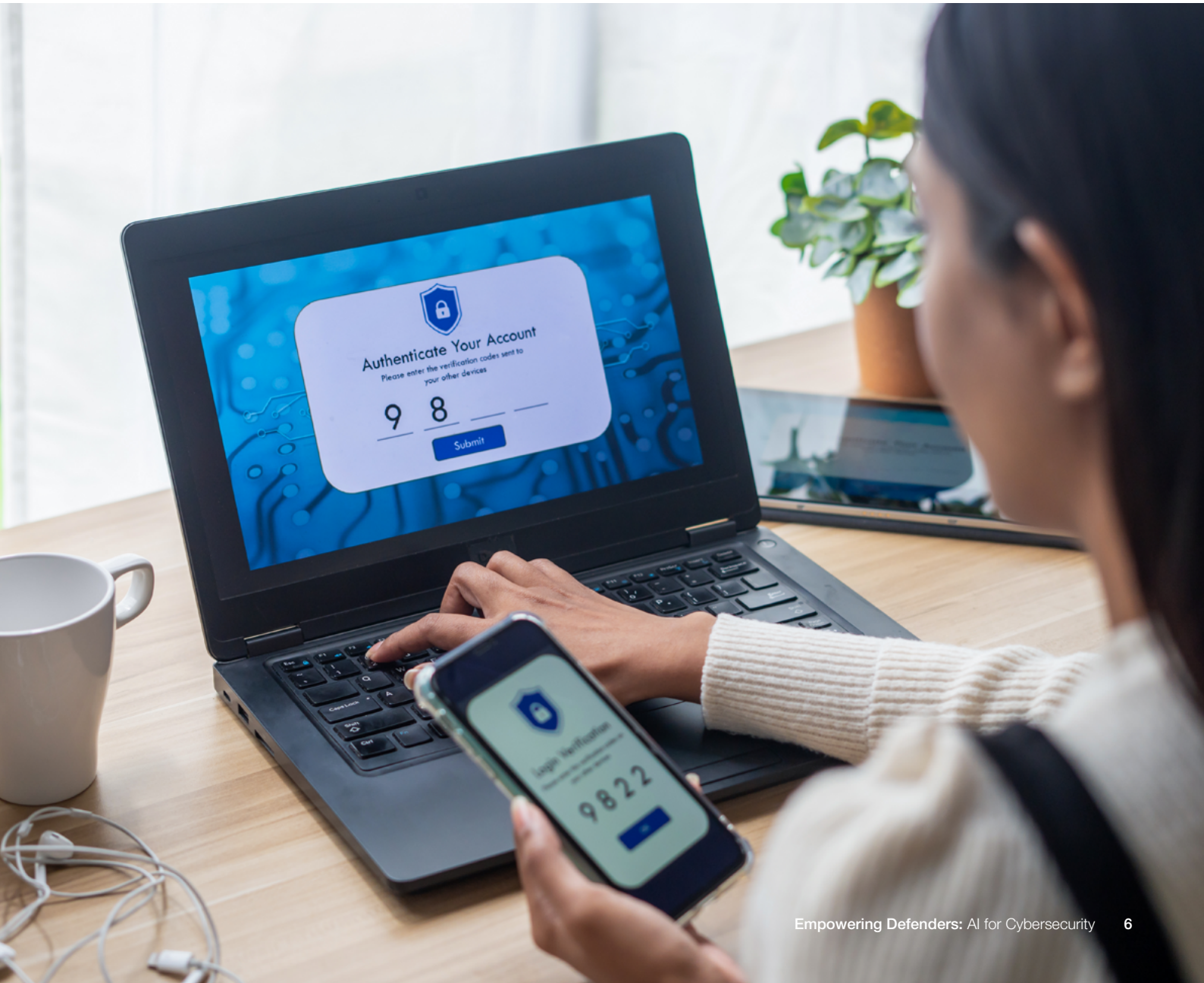
Regulatory and compliance pressures: Cybersecurity regulations – whether industry-specific, such as the Digital Operational Resilience Act (DORA), or cross-industry, such as the NIS2 Directive – mandate the rapid detection, reporting and mitigation of security incidents. As requirements expand in scope and complexity, AI can act as a force multiplier, accelerating incident detection and response, automating monitoring and documentation and maintaining regulatory readiness across the organization without increasing operational burden.

Ultimately, organizations that approach AI in cybersecurity strategically, aligning governance and operations, will be best positioned to counter adversaries and mitigate misuse.

Guarding against over-reliance on AI in cybersecurity

Heavy reliance on AI can undermine cyber resilience. Excessive trust in automated decisions creates a false sense of security and over time erodes the expertise needed to intervene when systems fail.

To prevent over-reliance on AI, security teams should combine AI with human judgement, simulate AI failures and design fail-safes that keep security operations functional during AI outages.



1 Present-day uses of AI in cybersecurity

AI in cybersecurity spans multiple functions, supporting activities across the entire security life cycle – from identifying risks and protecting systems to detecting threats, responding to incidents and enabling recovery.

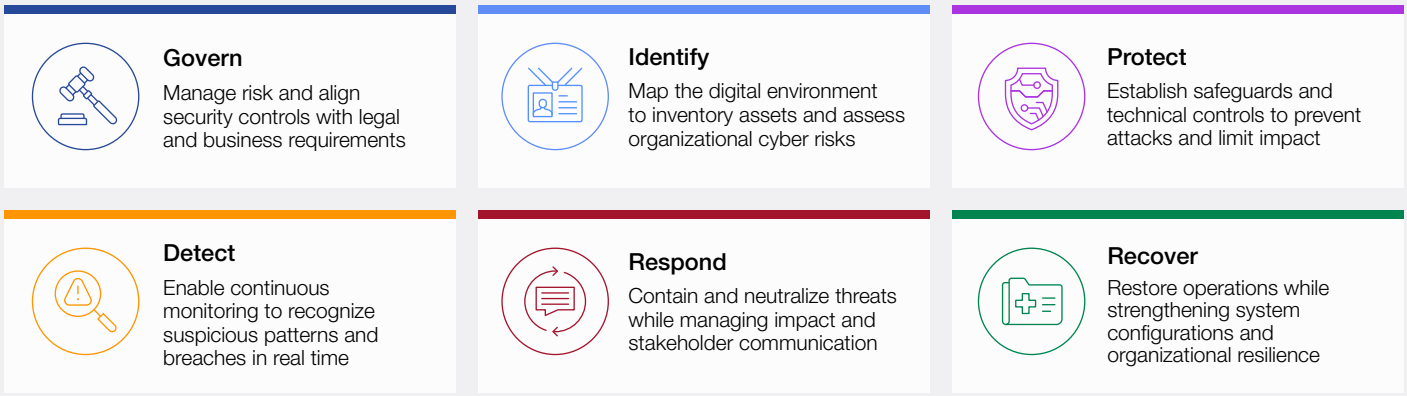
To understand the current state of AI in cybersecurity, this section draws on case studies submitted by World Economic Forum partners from various sectors and geographies, highlighting practical applications and real-world experience.

While the case studies are mapped against the six functions of the NIST Cybersecurity Framework 2.0 (see Figure 1) – govern, identify, protect, detect, respond and recover[8] – it is important to note that the framework is not applied prescriptively but rather serves to structure and categorize the case studies, ensuring that they are presented within a coherent and standardized context.

Many case studies span multiple functions, reflecting the interconnected nature of cybersecurity activities. While not exhaustive, the concentration of case studies in the identify and detect functions is consistent with the *Global Cybersecurity Outlook 2026*, which finds that 52% of organizations use AI for phishing detection, 46% for intrusion and anomaly detection and 40% for user behaviour analytics.

Additional insights come from workshops conducted under the World Economic Forum's Cyber Frontiers initiative, convening 105 representatives from 84 organizations across 15 industries.

FIGURE 1 **The six functions of the NIST Cybersecurity Framework[9]**



Source: World Economic Forum

1.1 AI-driven cyber governance

The "govern" function outlines basic considerations for managing cybersecurity risk, ensuring that policies and controls are followed and aligning security controls with legal and business requirements. AI can support this function in the following areas:

- **Regulations and compliance:** Through assessment of system configurations, audit trails and control implementations to ensure adherence to regulatory frameworks; AI can also help harmonize regulatory requirements across jurisdictions, reducing complexity and promoting consistent compliance.
- **Policy validation:** By checking whether internal security policies, such as password standards or multifactor authentication, are implemented correctly on premises and in cloud and hybrid environments, helping identify gaps and violations.
- **Audit and accreditation:** By generating audit-ready documentation, providing traceable evidence and enabling automated reporting to improve transparency and reduce manual effort: an example can be found in case study 1.
- **Cybersecurity supply chain risk management:** By validating supplier documentation, monitoring risk signals and correlating them with procurement data to detect compliance gaps and potential disruptions, optimizing the third-party risk management process.
- **Cyber-risk management:** By reviewing overall security posture and control effectiveness, identifying emerging risks and generating reports to support informed decision-making aligned with organizational risk appetite and strategic objectives.
- **Business impact analysis:** By evaluating the potential effects of disruptions on critical business functions, analysing operational data, dependencies and financial metrics; this supports asset prioritization, resource allocation and definition of recovery objectives.

CASE STUDY 1

Scaling security review for product portfolio with AI agents

Submitting organization: Rubrik

Challenge

Traditional manual threat modelling and security design reviews were time-intensive, difficult to scale and limited in coverage and consistency. Rubrik sought to replace these processes with an intelligent, automated system for structured risk analysis, validation and certification with full auditability.

Solution

Rubrik developed an AI-driven security review platform that replaces manual threat modelling with a multi-agent workflow.

- Agent 1: Security reviewer analyst

  Ingests design documents and architectural diagrams, performing structured security assessments using Rubrik's risk scoring matrix, STRIDE[10] threat modelling, OWASP Application Security Verification Standard (ASVS) 4.0[11] and Rubrik-specific context (back-up integrity, multi-tenancy, ransomware protection). It generates prioritized findings with common weakness enumeration (CWE) classifications, MITRE ATT&CK[12] mappings and remediation guidance.

- Agent 2: Code validation inspector

  Fetches code from source control and cross-references findings against actual code changes. Using static and differential code analysis, it detects implemented remediations, identifies partial implementations and flags unaddressed findings, providing code snippets as evidence.

- Agent 3: General availability (GA) certifier

  Aggregates security review findings and validation status, applies risk scoring and generates certification reports with residual risk assessments, compliance attestations and complete evidence chains from design through implementation, enabling data-driven GA sign-off decisions with a full audit trail.

Impact

This AI-driven approach replaces traditional manual threat modelling with an intelligent, automated system that delivers three times the coverage, reduces review time by 50% and generates more accurate security findings.

1.2 AI-enabled risk identification

The "identify" function focuses on mapping the organization's digital environment, inventorying assets and assessing cyber risks. AI is increasingly being applied for:

- **Threat intelligence:** By converting raw threat feeds into actionable insights tailored to the organization's context, improving situational awareness, prioritizing threats and supporting informed decision-making in security operations; case studies showcasing this include 2, 3, 5 and 6.
- **Security assessments and testing:** By simulating attacker behaviour to continuously test systems, uncover weaknesses and support large-scale adversary simulations, including automated penetration testing and red teaming; case study 4 illustrates AI applied to website security testing.

CASE STUDY 2

Enhancing threat intelligence through AI-driven correlation and attribution

Submitting organization: KPMG

Challenge
KPMG's threat intelligence team had a well-established data collection process, yet analysts spent long hours piecing together threat indicators, validating context and tracing attacks back to specific threat actors. Investigations progressed, but the effort required often meant extended analyst workloads during active engagement cycles.

Solution
A custom AI model was trained on the existing threat repository and was introduced as a chatbot within the intelligence platform. Analysts now query the system in plain language and receive relevant context, suggested links between attacks, sandboxing[13] results and guidance on tracing attacks back to specific threat actors. The model supports forensic investigations, incident response, red team scenario design using mapped tactics, techniques and procedures (TTPs) and rapid intelligence reporting. When required, a controlled fallback enables external AI-assisted research with cited sources.

Impact
The solution delivered a 25% increase in operational efficiency by reducing manual effort. The team transitioned to an autonomous, analyst-led structure, shifting senior leadership from close oversight to strategic quality assurance. By processing raw signals into structured intelligence without disrupting existing workflows, the system enables analysts to track a higher volume of threat actors in parallel with greater clarity.

CASE STUDY 3

Microsoft's Digital Crimes Unit (DCU) uses AI to curate and communicate threat intelligence data

Submitting organization: Microsoft

Challenge
Cybercrime fighters struggle to uncover hidden threat intelligence quickly within large volumes of data such as legal discovery responses and documents from hosting providers. Valuable threat indicators are often buried and difficult to find, which slows down forensic investigations and makes it harder to communicate risks to defenders and customers.

Solution
Microsoft's DCU developed an AI-powered tool to address these challenges. Haystack is an AI application that automatically tags likely threat intelligence indicators and uses an Azure OpenAI-based chatbot to curate intelligence rapidly. This enables analysts to identify relevant information within seconds or minutes rather than hours.

Impact
The solution significantly accelerated the identification and curation of threat intelligence, reducing forensic investigation times from hours to minutes. Moreover, it enabled clearer communication of complex threat data to both internal defenders at the DCU and customers, turning intelligence into actionable insights.

CASE STUDY 4

Agent Oliver, Accenture cybersecurity AI innovation

Submitting organization: Accenture

Challenge
Accenture's information security team faced what could be described as a fundamental scaling challenge: the enterprise attack surface included hundreds of thousands of internet-facing sites. Manually reviewing each site for common security issues required 15 minutes per site, meaning that scaling this process across the entire attack surface would take years for a resource-constrained team.

Solution
Agent Oliver is an advanced AI capability that has transformed how Accenture manages its external attack surface. It combines two distinct toolsets into a single capability:

- The Automated Site Crawler tool automatically scans the enterprise's internet-facing sites, visits each site and flags compliance issues (e.g. missing multifactor authentication).
- The Automated Web Application Tester tool tests web applications using four coordinated execution agents: crawlers map the application, configuration agents check settings, injection agents probe for vulnerabilities and a reporter consolidates the findings into a single report with recommendations for mitigating identified risks.

Impact
Agent Oliver was deployed across more than 100,000 sites. The time required to analyse each site dropped from approximately 15 minutes to under one minute, resulting in a 93% reduction in manual effort. This enabled security engineers to focus their time on more complex tasks.

CASE STUDY 5

Optimizing security operations centre (SOC) activities with AI integration

Submitting organization: Repsol

Challenge
Repsol identified the need to strengthen its cybersecurity posture across a complex enterprise environment. It faced challenges managing SOC activities, migrating security, information and event management (SIEM) platforms and processing large volumes of threat intelligence from multiple sources. Manual processes limited the speed and scalability needed to address evolving threats.

Solution
Repsol positioned AI as a strategic pillar, deploying an enterprise-grade agentic AI platform and expanding the use of large language models (LLMs). Within cybersecurity, AI was integrated into SOC operations and governance. Key initiatives included:

- Accelerating SIEM migration using AI-driven agents, enabling faster transitions
- Enhancing threat intelligence by using predesigned and optimized AI prompts, automating the selection, extraction and refinement of data on vulnerabilities and threat actors from multiple sources

Impact
SIEM migration was accelerated, reducing the lead time for detection rule creation and migration by 15%. Threat intelligence became more automated and actionable, improving 11% of related operational activities. These advances strengthened Repsol's defensive posture, improved efficiency and supported a more proactive approach to cybersecurity.

CASE STUDY 6

Automating and standardizing threat intelligence with agentic AI

Submitting organization: Check Point Software Technologies

Challenge
Security organizations need standardized threat intelligence workflows for consistent outputs. Manual research can be slow, inconsistent and difficult to scale. Threat indicators may be incomplete, contextual enrichment can vary in quality and turning analysis into actionable guidance for protection, response and recovery can take weeks. This creates duplicated effort and hinders intelligence sharing.

Solution
Check Point Research – the threat intelligence and research division within Check Point – developed Universe, an AI-powered research cycle that converts expert investigative knowledge into repeatable automation. Universe uses AI-driven multi-agent workflows and curated prompts that reflect Check Point Research's analytical standards. It reverse-engineers samples at scale, extracts and validates indicators of compromise (IOCs), enriches them with internal and public data and supports telemetry hunting[14] by generating and executing hunting queries to confirm prevalence and expand the intelligence picture. It generates standardized outputs, including IOC context, relationships and confidence levels alongside detection rules, threat patterns, concise analyst reports and remediation recommendations. Governance is strengthened through consistent structure, a clear evidence trail and a closed-loop system that feeds validated outcomes back into future investigations.

Impact
Internal measurements show that length of investigation dropped from roughly three weeks of manual effort to approximately one hour. This strengthens protection and response capabilities by turning intelligence into action more quickly, enabling earlier containment, reducing analyst-to-analyst variability and feeding validated intelligence into product improvements.

AI can also support other aspects of the identify function, namely:

- **Vulnerability management:** By ranking vulnerabilities based on threat intelligence, asset criticality and likelihood of exploitation, suggesting prioritized remediation actions to address the most critical security gaps efficiently; AI can further correlate vulnerabilities across the enterprise to predict attack paths, identifying how individually low-rated vulnerabilities could be chained together into a significant breach.
- **Asset management:** To discover, track and categorize hardware and software assets, predicting maintenance needs, life-cycle changes and potential compliance issues.
- **Open-source intelligence (OSINT):** By collecting and organizing publicly available intelligence, scanning underground forums, marketplaces and leak sites to identify exposed credentials, sensitive data and indicators of compromise relevant to the organization; this enables teams to anticipate external threats proactively.

1.3 AI-augmented cyber protection

The "protect" function establishes safeguards, including technical controls and employee practices to prevent attacks or limit the impact of security events. To advance this function, AI technologies are used in:

- **Secure software development:** To support developers in identifying insecure coding patterns, configuration flaws and vulnerabilities early, suggesting and implementing fixes to embed security into software from the start; case study 7 demonstrates this application.
- **Configuration management:** Including monitoring of system configurations across on-premises and cloud environments to ensure alignment with security baselines and policies; misconfigurations and unauthorized changes are detected, enabling rapid remediation and keeping systems secure and consistent with minimal manual effort – a practical example can be found in case study 8.
- **Evaluating technology infrastructure resilience:** By analysing architectural patterns and system interactions to identify weaknesses and assess configurations, helping build resilient systems before deployment; case study 8 illustrates this capability.
- **Domain protection:** By monitoring domains for spoofing, hijacking and other fraudulent activities such as impersonation, helping prevent brand misuse and automate takedown processes to reduce response time and reputational damage; an example of this can be found in case study 9.

CASE STUDY 7

Operationalizing AI agents for large-scale vulnerability detection and remediation

Submitting organization: Google

Challenge

Google needed to secure large, complex software codebases, where manual review could not keep pace with the speed and scale required to detect and address vulnerabilities. Rapid identification and remediation of unknown security flaws was essential.

Solution

Google deployed AI agents to enhance software security by increasing the speed, consistency and scale of vulnerability detection and remediation.

- Big Sleep is an AI-based security agent that actively searches for and identifies unknown security vulnerabilities in software, enabling security teams to act before flaws are exploited. Big Sleep has been deployed across Google's products as well as on widely used open-source projects.
- CodeMender, developed by Google DeepMind, is an AI-based agent that automatically improves code security by generating patches for identified vulnerabilities. To ensure reliability and safety, all CodeMender-generated patches are reviewed by human researchers before being submitted upstream.

Impact

The deployment of AI agents has reduced detection latency and accelerated remediation timelines across large and complex codebases. This approach has lessened reliance on scarce security engineering resources and strengthened security not only within Google but also across the wider digital ecosystem. Since its launch, CodeMender has patched more than 100 critical security issues, including in complex codebases such as the V8 JavaScript engine.

CASE STUDY 8

AI-driven security architecture and engineering for scalable, secure-by-design systems

Submitting organization: AXIS Capital

Challenge

The AXIS Capital security architecture and engineering team wanted to embed security intelligence directly into application and cloud design, minimizing developer friction and reducing time spent managing alerts and backlogs. Traditional manual reviews could not scale to meet the demands of complex cloud environments.

Solution

The team applied AI to analyse code and cloud configurations, using static application security testing and cloud security posture management capabilities. Uninterrupted scanning within continuous integration and continuous deployment pipelines enables proactive risk assessment, identifying and addressing vulnerabilities before code reaches production. AI prioritized the risks based on exploitability and business impact, recommending actionable remediation. AI-generated guidance and fixes were delivered within existing workflows, accelerating remediation and reducing developer friction. Teams also use AI to correlate data across cloud environments to identify architectural weaknesses and surface misconfigurations that are difficult to detect manually. Through ongoing cloud security posture management, AI detects misconfigurations in real-time.

Impact

This approach reduced time spent managing alerts and backlogs, freeing teams to focus on designing resilient systems and ensuring that security measures can be scaled without affecting speed of delivery and business development. It also accelerated response times, reduced remediation costs and strengthened overall cloud security while maintaining engineering productivity.

CASE STUDY 9

AI-driven resilience: Elevating urban cybersecurity with RZAM

Submitting organization: Dubai Electronic Security Center (DESC)

Challenge
As Dubai's digital economy expands, individuals and businesses face growing volumes of sophisticated web-based threats, such as phishing and malicious URLs, which traditional security measures often fail to detect in real time. The smart city ecosystem has widened the attack surface, increasing exposure to data theft and financial fraud during everyday browsing. Static threat databases are no longer sufficient to counter evolving cybercrime.

Solution
DESC developed RZAM, an AI-powered browser extension and mobile app providing a "fearless browsing" experience. Using machine learning trained on more than 1 million URLs, RZAM performs real-time webpage analysis to identify and block malicious content within milliseconds, acting as a seamless digital shield across platforms. It enables proactive protection without collecting or storing personal user data.

Impact
Since deployment, RZAM has strengthened Dubai's digital resilience, identifying malicious sites with more than 95% accuracy. By neutralizing threats before they compromise devices, it reduces phishing success and strengthens trust in digital services. RZAM supports Dubai's ambition to be the safest city in cyberspace while enabling a secure environment for innovation and investment.

Other possible applications of AI to advance the protect function include:

- **Identity management, authentication and access control:** By analysing user behaviour and access patterns to automate permissions, enforce least-privilege policies, detect and remediate entitlement accumulation, reduce insider risk and securely manage machine and agent identities.
- **Awareness and training:** Tailored to individual roles and behavioural patterns, increasing engagement and fostering a security-conscious culture throughout the organization.
- **Data security:** Through classification and labelling of data, monitoring data flows, access patterns and user behaviours to detect anomalies and enforce data protection policies.

1.4 AI-powered threat detection

The "detect" function ensures continuous monitoring and timely recognition of suspicious patterns and activities as well as potential security breaches in real time. AI enhances detection by supporting:

- **Threat hunting:** Analysing large volumes of data to uncover hidden patterns and anomalies, automatically generating queries to validate findings; AI-powered threat hunting enables cyber teams to focus on higher-complexity investigations – an example of AI-enabled hypothesis-based investigation can be found in case study 10.
- **Adverse event analysis:** Through the examination of logs and malware samples to classify and report incidents, accelerating early investigations and identifying evasive or novel threats; relevant case studies include 10, 11, 13 and 15.
- **Phishing and email threat detection:** By analysing message content, sender behaviour and communication patterns to detect and block phishing, spam and business email compromise attempts, reducing susceptibility to impersonation; case studies 12 and 16 illustrate this application.
- **Anomaly detection:** Including monitoring of user behaviour, endpoint activity, network traffic or website activity to detect unusual patterns, uncovering insider threats, compromised accounts and stealthy attacks; practical examples are shown in case studies 14 and 15.

CASE STUDY 10

Hypothesis-based AI analysis for autonomous cyberthreat detection

Submitting organization: Allianz

Challenge

Traditional cybersecurity operations face a critical scalability problem in autonomous threat detection. Collecting 10–20 GB/s per endpoint across 350,000 endpoints would generate 15 petabytes daily, making centralized data collection impossible. This creates dangerous gaps between alert generation and investigation, increasing the risk of missed threats and prolonged incident response times.

Solution

Allianz developed a hypothesis-based AI analysis system that reimagined threat investigation. Rather than collecting all endpoint data centrally, the AI generates hypotheses when alerts are triggered, identifies the data points needed to validate them and retrieves them on demand using forensic application programming interfaces (APIs). The system iteratively refines its analysis, mirroring the on-demand data-processing approach used in autonomous vehicle technology. This enables targeted dynamic data collection, avoiding infrastructure overload.

Impact

This approach enables comprehensive forensic analysis at scale, even for minor cases, while maintaining operational efficiency. By bridging the gap between conditional and high automation, the system reduces mean time to respond (MTTR) and strengthens cyber detection capabilities.

CASE STUDY 11

AI-enhanced malware analysis efficiency through Assemblyline

Submitting organization: Canadian Centre for Cyber Security

Challenge

The Canadian Centre for Cyber Security, like many other organizations, faces an escalating volume of increasingly complex malicious files, placing pressure on cybersecurity teams to conduct rapid and accurate analysis at scale. Efficiently processing and prioritizing large inflows of files while minimizing manual reporting and cognitive load is critical for maintaining effective cyber defence.

Solution

To address this, the Canadian Centre for Cyber Security developed Assemblyline, an open-source automated malware analysis and triage system used by defenders across government, industry and academia.[15] It integrates more than 50 analytic services and offers a scalable infrastructure capable of handling millions of files daily. Since 2024, the Cyber Centre has added AI-driven summarization enhancements to Assemblyline, making this capability available to the broader open-source user community. This feature streamlines manual reporting and helps translate technical findings into actionable intelligence.[16]

Impact

The deployment of Assemblyline accelerates incident response and increases throughput without proportional increases in analyst headcount. AI-driven summarization reduces analysts' cognitive load and demonstrates how defenders can responsibly leverage AI to enhance operational capacity in modern cyber defence environments.

CASE STUDY 12

AI-driven phishing detection through persuasion analysis

Submitting organization: Santander Group

Challenge

Phishing is one of the most prevalent cyberthreats, increasingly relying on social engineering rather than technical exploits. Modern phishing campaigns are often targeted, multilingual and fast-evolving, exploiting behavioural and psychological tactics including false urgency, authority impersonation and the imitation of trusted sources. Traditional detection approaches based on signatures, sender reputation or static rules struggle to identify these novel attacks, particularly when infrastructure and content are previously unseen.

Solution

An AI-driven phishing detection solution was deployed as an augmentation layer into email security operations at Santander Global SOC. The solution combines multiple analytical capabilities, including domain analysis, brand and visual inspection and advanced semantic analysis. A key differentiator is an internally trained, multilingual LLM designed to identify psychological persuasion traits embedded in phishing messages, such as scarcity, authority, reciprocity and commitment. By focusing on semantic intent and humancentric manipulation patterns, the system improves detection of previously unseen, highly targeted phishing campaigns that evade traditional controls.

Impact

Overall phishing detection effectiveness improved by at least 10%, while also accelerating threat identification and enabling earlier disruption of social engineering-driven attacks. This shifts phishing detection from indicators of compromise to a human-centric threat analysis paradigm, strengthening cyber resilience.

CASE STUDY 13

Threat detection and response with autonomous threat operations machine (ATOM)

Submitting organization: IBM

Challenge

Adversaries increasingly use automation and AI to move faster and operate at scale, placing sustained pressure on security operations. IBM needed to expand its 24x7 threat investigations without adding headcount or compromising governance or quality. At the same time, it aimed to reduce manual effort, investigation time and analyst alert fatigue. The challenge was to automate investigations at scale while maintaining trusted, explainable and auditable outcomes across global managed security services.

Solution

IBM implemented ATOM as the central investigation and triage engine within its managed security services. ATOM autonomously investigates, enriches and scores alerts at scale using agentic AI, with human analysts focusing on oversight and escalation. Today, ATOM handles about 95% of daily investigations, delivering consistent, explainable and auditable results. IBM operates as Client Zero, running its global managed security services on ATOM. ATOM enables automation, improving speed, accuracy and investigation quality. ATOM applies deep context, correlation and explainable AI to strengthen detection while ensuring human-in-the-loop governance.

Impact

ATOM automates more than 850 analyst hours per month, enabling IBM to scale global 24x7 threat detection and response without adding headcount. Investigation time decreased by 37%, with annotation completed more than 40% faster than if done manually. Detection quality improved, with 6,700+ high-risk activities identified in sampled production alerts affecting quality, governance or accountability.

CASE STUDY 14

AI-enhanced real-time intellectual property access monitoring

Submitting organization: Cybervergent

Challenge
Traditional static controls often fail to distinguish between legitimate team activity and malicious mass downloads, leaving organizations vulnerable to insider threats, corporate espionage and confidential information leaks. The rise of shadow AI has created new pathways for intellectual property theft, and the gradual accumulation of excessive access privileges increases the risk of attackers moving across systems within the network and of regulatory non-compliance.

Solution
Using agentic AI, Cybervergent implemented real-time monitoring of the speed and volume of data access across its intellectual property channels. This solution automates governance and secures IP by continuously analysing the speed and volume of data movement. Key features include:

- Source code/proprietary information exfiltration detection: Real-time monitoring distinguishes between normal developer activity and suspicious mass downloads, identifying potential insider threats.
- Shadow AI leak prevention: Tracking data egress to public AI platforms prevents proprietary code and information from entering external training sets, safeguarding sensitive assets from exposure.
- Managing access privileges: Identifying and responding to the use of expired or escalated permissions, which may signal account takeover or unauthorized movement across systems.
- Regulatory compliance: Automating controls ensures that data is not transferred to unvetted AI tools, thereby reducing the risk of regulatory fines and legal liabilities.

Impact
The solution has helped to almost eliminate source code and proprietary information exfiltration, reduced shadow AI risks and improved detection of privilege misuse, helping keep sensitive IP within secure internal environments.

CASE STUDY 15

AI as a force multiplier: How Aramco closes cybersecurity gaps through in-house AI innovation

Submitting organization: Aramco

Challenge
Aramco's cybersecurity landscape demands solutions that can keep pace with rapidly evolving threats. In several critical domains, no commercial technology existed to address the scale, complexity or specificity of the challenges facing the world's largest energy company.

Solution
To bridge these gaps, Aramco Cybersecurity embedded AI at the core of its defensive ecosystem. Some 40% of its cybersecurity solutions use AI not as an add-on but as a foundational capability to detect malicious activity, automate response and enhance its incident response with operational efficiency. An example is the development of more than 50 patented AI solutions designed to identify sophisticated attack patterns, analyse massive volumes of telemetry and autonomously trigger containment actions. These solutions were deployed through a modular architecture that integrates seamlessly with existing security platforms, enabling real-time threat detection and automated workflows that previously required extensive manual effort.

Impact
AI has empowered defenders with proactive capabilities, allowing Aramco to anticipate rather than merely react to emerging threats with a 99% accuracy rate.

CASE STUDY 16

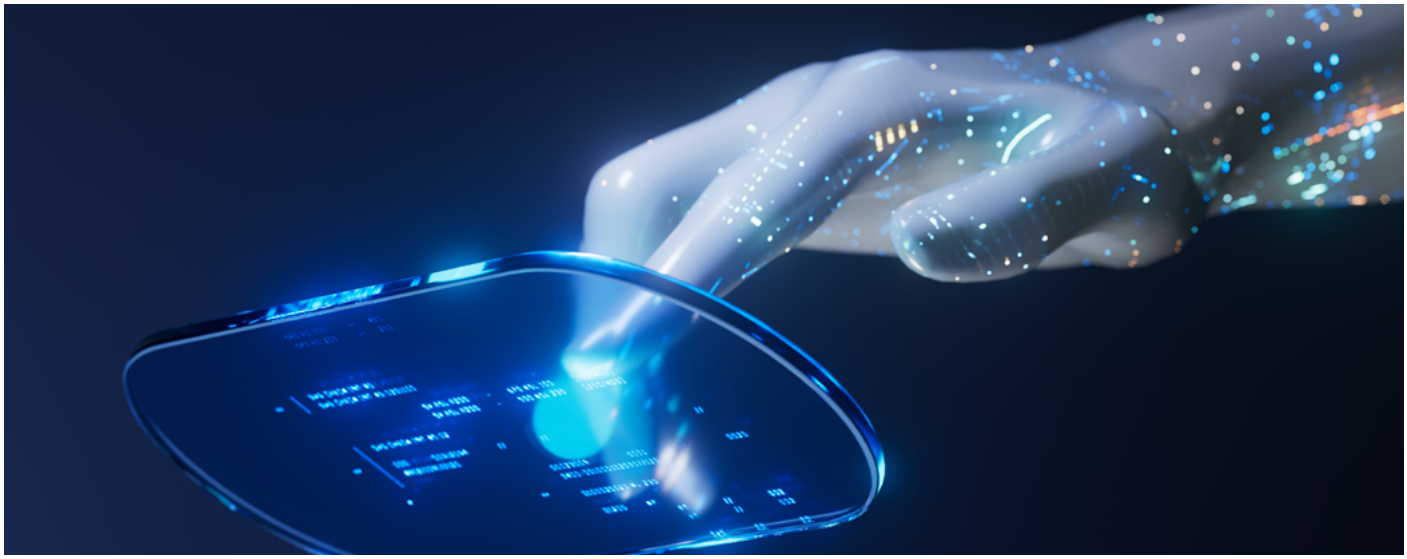
ING's AI-powered data leakage prevention

Submitting organization: ING

Challenge
ING, a global financial institution, has developed a machine learning solution built on top of its existing data leakage prevention (DLP) cybersecurity tooling to better process and prioritize alerts related to potential data leaks across multilingual email attachments, web uploads and metadata.

Solution
The implementation provides a reproducible workflow that can be extended to other organizations to process multilingual email attachments, web uploads and metadata. ING uses an AI model to categorize related email attachments and combines it with a classification model to prioritize and identify potential leaks. ING has also implemented a "browser uploads" production pipeline that provides SOC analysts with very-near-real-time insights via easy-to-read dashboards, enabling them to handle issues promptly as they seek to protect more than 60,000 employees globally.

Impact
To date, the solution has processed 5 million alerts, resulting in a 20% increase in analyst precision. ING's internal SOC survey indicates that analyst job satisfaction has significantly increased with the use of this AI-based workflow.



1.5 AI-orchestrated incident response

The "respond" function defines coordinated actions to contain and neutralize threats, manage their impact and communicate effectively with stakeholders. AI supports this function through:

- **Incident analysis:** By collection, correlation and analysis of large datasets to identify the root cause of the incident through pattern recognition, timeline reconstruction and detection of indicators of compromise; in malware incidents, AI helps reverse-engineer code to reveal behaviour and impact, strengthening containment, attribution and remediation efforts, as illustrated in case studies 17 and 19.
- **Incident mitigation:** By recommending or automating countermeasures such as blocking traffic, revoking credentials or isolating compromised systems to limit the impact of security breaches; automating time-sensitive steps reduces response delays and minimizes business disruptions, as illustrated in case studies 17 and 20.
- **Incident management:** Via prioritization and classification of incidents, escalating as needed to ensure timely handling of critical events while reducing human workload; a practical example of this AI application is provided in case study 18.
- **Incident reporting and communication:** By generating concise and audience-specific summaries of security incidents, translating complex data into actionable insights for technical teams and executive leadership, as demonstrated in case study 18.

CASE STUDY 17

AI-enabled cyber defence: Accelerating SOC performance with embedded intelligence

Submitting organization: Petroliam Nasional Berhad (PETRONAS)

Challenge
PETRONAS sought to elevate analyst effectiveness, eliminate repeatable manual work and accelerate response across its enterprise SOC.

Solution
Advanced AI capabilities were integrated directly into analyst workflows, providing real-time incident summaries, guiding next steps, translating natural language to queries and automating context gathering. This standardizes investigation quality and enables junior analysts to contribute earlier in their career.

Building on this, AI agents are being deployed that autonomously triage high-volume incidents, correlate alerts across telemetry and link anomalies with threat intelligence. These agents surface actionable patterns with structured recommendations, reducing manual correlation work. Integrated natural language intelligence look-ups provide context on threat actors, campaigns, malware families and IOCs, strengthening prioritization and proactive detection.

The initiative was shaped through a six-month pilot in which use cases were prioritized based on analyst time spent, incident recurrence, task consistency and risk relevance, ensuring high-impact investment. Adoption considerations – including analyst mindset, workflow adjustments and change management needs – were addressed through continuous feedback loops. Deployment of AI agents required no new hardware, and it was scaled incrementally based on real-world performance metrics and analyst input, enabling safe AI adoption, operational continuity and value validation at each stage.

Impact
Within three months, a 30–40% reduction in incident response and resolution times was achieved, freeing capacity for advanced threat hunting and investigations. New analyst ramp-up time improved by 50%, strengthening SOC readiness and resilience in the face of an increasingly complex threat landscape.

CASE STUDY 18

AI hyper-automation for SOC and case management

Submitting organization: Standard Chartered

Challenge
At Standard Chartered, SOC and response teams face increasing pressure from growing alert volumes, complex investigations and rising expectations in terms of speed, consistency and auditability.

Solution
To address these challenges, Standard Chartered implemented an AI hyper-automation strategy embedded into SOC and case management workflows, with a strong focus on analyst enablement and governed automation. The solution is an intelligent platform providing real-time situational awareness of users, applications and threats across the bank. The bank-wide solution applies machine learning and LLMs to dynamically score risk, prioritize alerts and cases and enrich detections with contextual intelligence. AI-driven triage automatically classifies events and averts duplication before case creation, while generative AI supports analysts by producing concise case summaries and drafting communications. An in-console AI co-pilot provides real-time guidance, similar-case recommendations and next-best actions, operating under a strict human-in-the-loop model.

Impact
Deployed incrementally with guardrails, full observability and kill-switch controls, the approach delivered measurable efficiency gains, including a 25–35% reduction in manual triage effort and 20–30% improvement in time-to-triage. Low-risk, repetitive cases are auto-resolved within defined thresholds, enabling teams to focus on high-impact investigations. The result is a responsive, scalable operating model that strengthens security outcomes without compromising human accountability or control.

CASE STUDY 19

AI-assisted deep malware analysis for scalable response

Submitting organization: Dream Group

Challenge
As Dream's security operations encountered increasingly sophisticated malware linked to multistage attack campaigns, deep technical analysis became critical for response and eradication. Although detection flagged suspicious activity, producing remediation guidance often required multiday reverse-engineering efforts by specialists across multiple programming languages. This caused delays and limited how many complex incidents could be handled simultaneously. Scaling high-quality malware response without expanding specialist headcount became a priority.

Solution
An internal AI-assisted malware analysis capability was developed and integrated into detection and response workflows. Designed and refined by Dream's cybersecurity AI researchers embedded in operational investigations, the system analyses malicious code to examine how it runs, remains active on a system, escalates access, moves across systems and communicates with attackers. It extracts TTPs and generates structured, explainable outputs supporting containment, environment-wide hunting for related variants and remediation and eradication guidance. Outputs are transparent, evidence-backed and reviewed by human analysts before action is taken.

Impact
Since deployment, remediation guidance time has been reduced by up to 95%, compressing multiday reverse-engineering efforts into minutes. This reduces manual workload, enabling faster response and greater parallel investigation capacity without increasing specialist headcount. By standardizing advanced malware response workflows and reducing reliance on language-specific experts, Dream strengthened its ability to detect, mitigate, hunt and eradicate malware variants at scale.

CASE STUDY 20

From classification to containment: Streamlining cyber defence with AI-driven software classifier and magic scripts

Submitting organization: Batuta

Challenge
Batuta's service teams, IT managers and solution engineers spend significant time manually writing scripts to detect and remediate endpoint vulnerabilities and deploy technologies. Script development requires both deep technical expertise and situational awareness, creating a bottleneck. To scale, the business must boost productivity by five times without increasing headcount or compromising security.

Solution
Magic scripts convert natural language instructions (e.g. "block port 445") into safe, executable code, enabling rapid containment and remediation. The system is context-aware, adapting scripts to the active module. For example, compliance modules generate read-only scripts, while remediation modules produce scripts that enforce controls. Scripts are tailored to the target operating system (OS) and environment, accounting for platform constraints. Built-in safety checks prevent harmful actions (e.g. disk wiping or the disabling of security tools), with user approval required for sensitive operations.

Impact
Magic scripts accelerated script creation by up to 12 times and reduced manual effort by 99%. Tasks that took an hour now take under a minute. As a result, teams are up to eight times more productive, enabling greater focus on higher-value work such as incident investigation, proactive security assessments and scaled service delivery.

AI can further enhance this function in several key areas, namely:

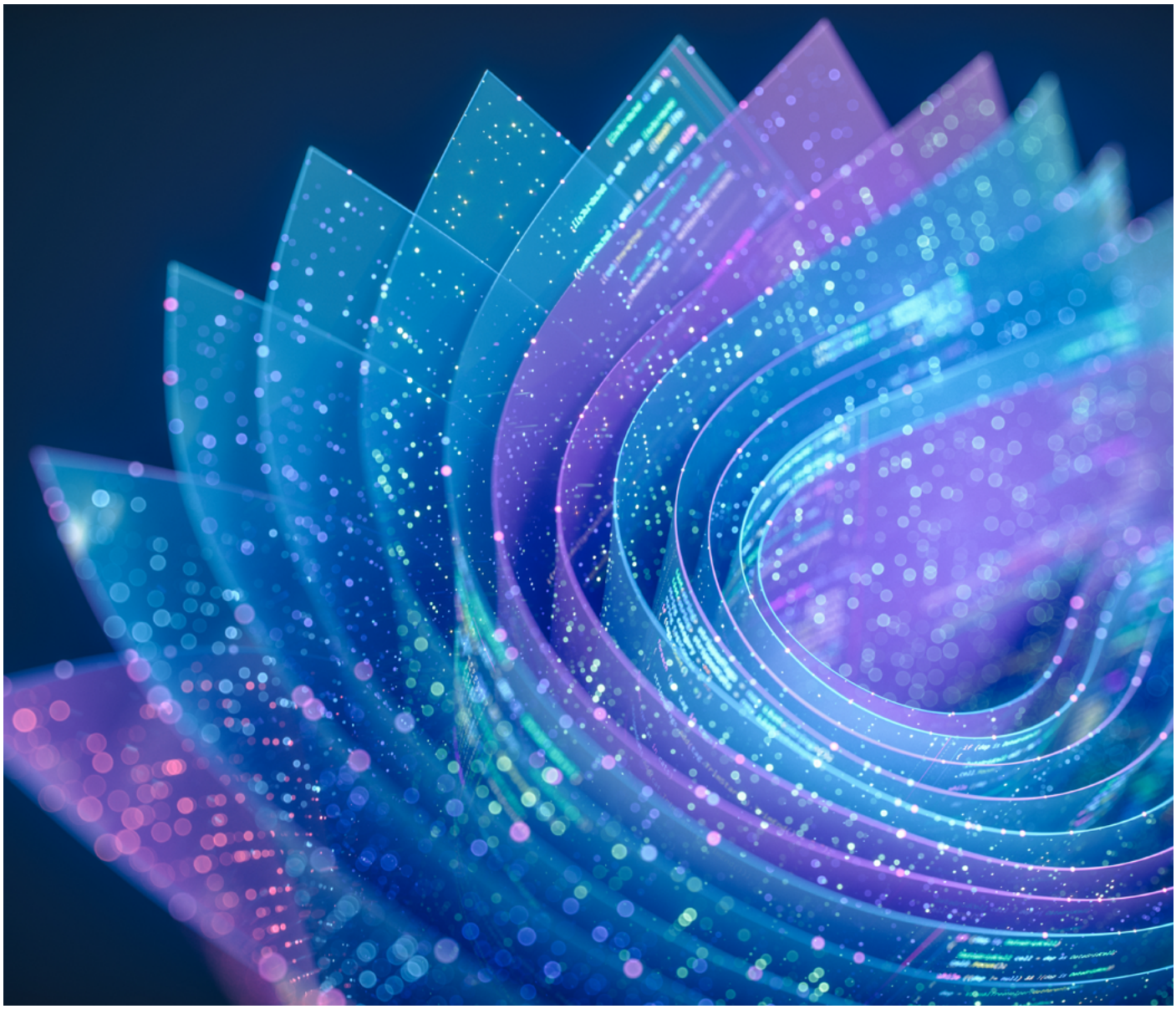
- **Response plan creation and management:** By supporting the development and ongoing refinement of incident response playbooks, through aligning procedures across teams and partners.
- **Response plan testing and validation:** Through assessment of quality and readiness of an organization's incident response procedures by simulating realistic attack scenarios; in conducting purple team exercises and sandbox-based testing, AI identifies gaps and areas for improvement in playbook design and execution.

1.6 AI-supported incident recovery

The “recover” function focuses on restoring normal operations following a cybersecurity incident, while strengthening system configurations, processes and organizational resilience to better withstand future attacks. To date, the adoption of AI within the “recover” function remains relatively limited in practice, with most applications largely conceptual or at early stages of exploration. That said, AI has the potential to support this function in:

– **Recovery plan creation and management:** Through the development and refinement of recovery plans based on analysis of system dependencies, risk scenarios and historical incident data; automated plan updates ensure readiness, minimize downtime and enhance organizational resilience.

– **Recovery plan testing and validation:** By reviewing their quality and using AI to simulate ecosystem-wide failures and assess response effectiveness; this helps to identify gaps, enable continuous stress testing and generate actionable improvements.



The case studies presented throughout this section highlight how AI supports defence capabilities. Collectively, they demonstrate improvements in threat intelligence, vulnerability and threat detection, incident response and resilience against attack vectors such as phishing, underscoring AI’s growing operational impact across cybersecurity functions.

2 Important considerations for AI adoption in cybersecurity

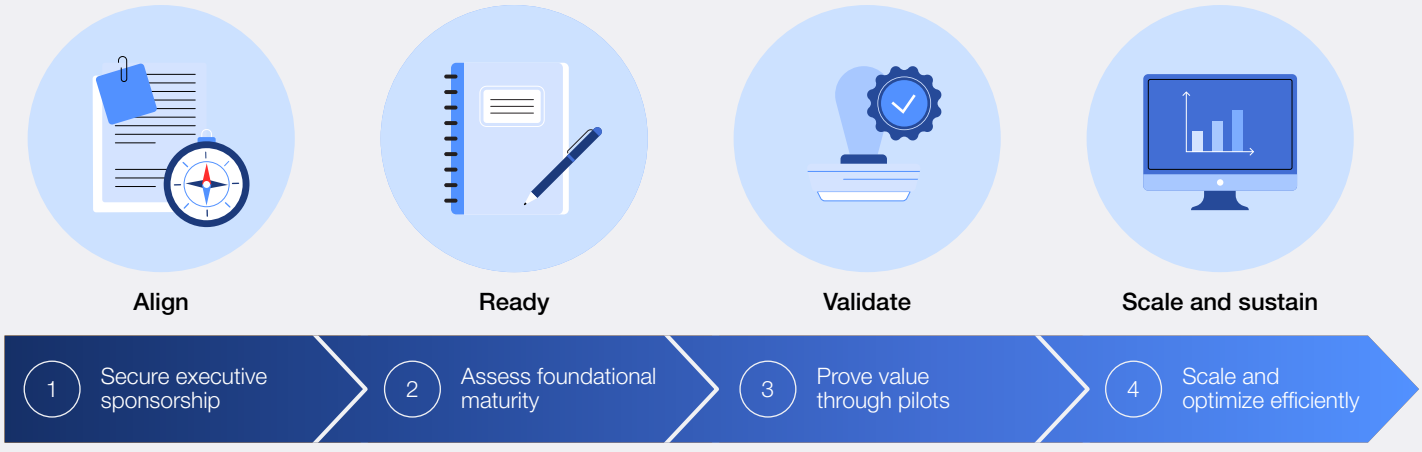
The adoption of AI in cybersecurity requires executive sponsorship, assessment of foundational maturity, validation through pilots and staged scaling.

AI adoption in cybersecurity is a strategic initiative that directly shapes business outcomes, operational resilience and organizational risk. Success depends on, among other things, harmonizing strategy, technology, governance and organizational culture.

While the AI applications discussed in the previous section demonstrate what is possible, this section outlines key questions for C-suite executives to guide successful AI adoption in cybersecurity, along with practical steps that chief information security officers (CISOs) can take to translate them into measurable outcomes.

Although these broad considerations are applicable to most organizations, the path to adoption may vary. While larger enterprises can use greater resources, small and medium-sized enterprises (SMEs) can benefit from simpler ecosystems and leaner decision-making.

FIGURE 2 **The path to adopting AI in an organization's cybersecurity operations**



Source: World Economic Forum

2.1 How does the adoption of AI for cybersecurity align with and accelerate strategic priorities?

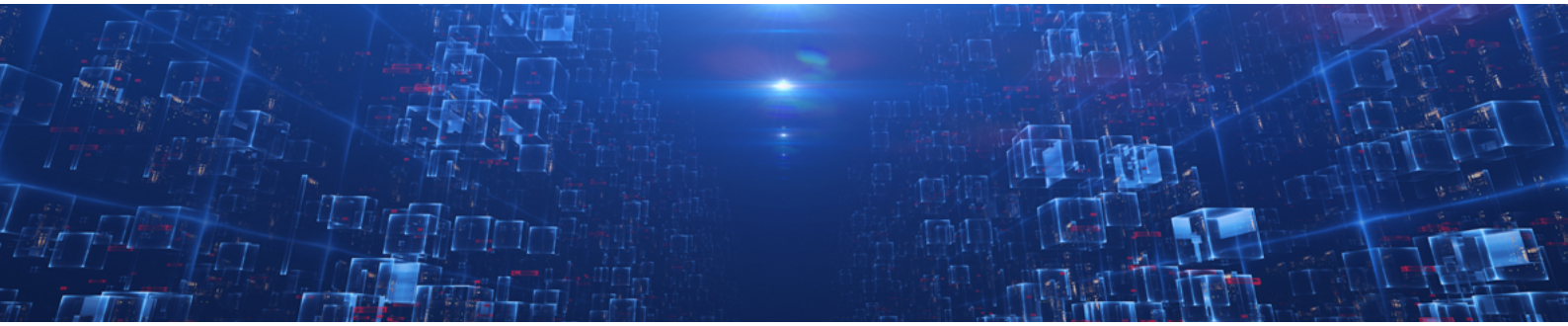
Before investing in AI for cybersecurity, executives need clarity on the business outcomes AI is expected to deliver. AI should not be adopted for its own sake, but as a capability that accelerates priorities such as operational resilience, regulatory compliance, customer trust and cost efficiency. Without clear objectives, AI initiatives risk misalignment with strategic priorities, resulting in failure to secure executive sponsorship and, ultimately, funding.

To achieve this clarity, CISOs should:

- Articulate how AI deployment in cybersecurity is linked to business priorities and demonstrate how it supports key company objectives such as resilience, growth, compliance and cost efficiency.
- Validate AI suitability by assessing whether AI is truly required or if standard automation provides a more cost-effective, lower-risk alternative.
- Communicate clearly what AI can and cannot do, where it adds value and what risks and trade-offs may be involved; CISOs should demonstrate that AI projects comply with regulatory and legal compliance and align with the organization's risk tolerance, ensuring that initiatives operate within acceptable boundaries.
- Communicate success in business terms by translating improvements resulting from AI into metrics that executives care about, such as reduced risk exposure, accelerated incident recovery, faster time-to-market or enhanced service availability; moreover, CISOs should demonstrate early value through quick wins by selecting high-impact, feasible use cases that deliver measurable results quickly.
- Secure executive sponsorship by engaging leadership early and continuously to validate priorities, allocate resources and position AI as a strategic business and risk-management tool.

SME tip

Focus on one or two high-impact use cases that are directly tied to your most critical business risks or operational challenges.



2.2 Is the organization ready to deploy AI in cybersecurity effectively?

AI can strengthen cybersecurity only if the necessary foundational elements are in place. Gaps in processes, data, infrastructure, skills or governance can undermine deployment and waste resources. The C-suite should evaluate organizational readiness before mandating the introduction of AI into security operations.

To position the organization for success, CISOs should:

- Assess process and operational readiness by verifying that security workflows are documented and repeatable; stable processes help identify where AI can add value rather than mask operational gaps.
- Evaluate technical and infrastructure readiness by ensuring that AI integrates seamlessly with existing architecture and security tools, addressing legacy constraints and integration complexities and confirming that identity and access management frameworks can securely handle machine and agent identities with appropriate controls.

- Validate data readiness by verifying that the security datasets required for AI deployment are available, complete, accurate, well-structured and include relevant organizational context; since AI is only as good as the data on which it relies, gaps such as missing historical records or inconsistent formats can lead to incorrect conclusions, missed threats or false alerts.
- Assess skills readiness by evaluating whether security teams have the expertise to manage the full AI life cycle and address skills gaps through reskilling and continuous learning programmes.
- Evaluate governance readiness by verifying that structures and processes exist to oversee AI initiatives, mandate impact and risk assessments, and implement guardrails to ensure protections evolve alongside AI adoption. Clear ownership, accountability and decision-making authority should also be established.
- Promote an AI-ready culture by ensuring that teams are prepared to integrate AI into daily security operations, and build trust in AI by delivering reliable, actionable insights and clearly communicating its capabilities and limitations – this helps move teams from scepticism to active collaboration.

BOX 1 Build-or-buy considerations for AI-driven cybersecurity

Deciding whether to build or buy AI solutions for cybersecurity is an important strategic choice. It affects technology decisions, operating processes, required skills and governance structures. Before committing to either path, organizations should assess their existing tooling, as many security tools now embed AI functionality that may already meet their needs.

Custom-built solutions can offer tailored capabilities but require significant investment and in-house expertise. Commercial solutions enable faster deployment and easier scaling but may offer less flexibility. Table 1 outlines the key factors organizations should consider when evaluating these two options.

In practice, many organizations combine in-house and vendor solutions and adopt a hybrid model to balance speed, control and flexibility. Hybrid approaches allow organizations to leverage vendor solutions for rapid deployment and scale while retaining in-house control and opportunity for customization over critical capabilities.

TABLE 1 **Choosing between custom-built AI solutions and commercial products**

Dimension	Build (in-house)	Buy (vendors)
Deployment	Slower due to development, training and resource allocation	Faster due to ready-made solutions
Sovereignty	Greater control over data, architecture and models	Limited control and third-party dependency
Talent	High internal demand for skilled resources	Lower internal demand, access to specialized expertise
Flexibility	High customization for specialized security workflows	Limited customization; vendor lock-in risk
Risk	Execution risk and maintenance liability	Third-party vendor stability, data portability and compliance risk
Cost	Higher initial total cost of ownership and maintenance overheads, lower long-term costs	Lower initial total cost of ownership, higher cumulative costs from integrations and scaling
Scalability	High control over infrastructure and continuity	Resource scaling limited by vendor constraints
Best suited for	AI strategic differentiator, provides proprietary competitive advantage	Commodity utility, provides rapid time-to-value

BOX 2 | **The next-generation cybersecurity workforce**

The rapid integration of AI is creating a growing mismatch between technological advances and workforce readiness. While AI tools offer significant gains in areas such as threat detection, 54% of organizations identify a shortage of skilled talent as the primary barrier to adoption.[17]

As organizations automate manual tasks, routine and repetitive work is losing its importance. The "single-tool expert" is becoming less central as cybersecurity increasingly seeks versatile professionals who can adapt to multiple tools and functions. More than ever, organizations need cybersecurity professionals who combine strong technical skills with soft skills such as critical thinking, problem-solving and storytelling skills to translate complex AI outputs into actionable guidance. According to the World Economic Forum *Future of Jobs Report 2025*, demand for these skills is expected to rise significantly through to 2030, reinforcing the need to develop well-rounded professionals who can operate effectively in AI-augmented environments.

That said, increasing reliance on AI can reduce opportunities for hands-on practice, leading to skills atrophy, which in turn could weaken organizational resilience when automation falls short.

In response, organizations need a deliberate and structured approach to talent development. This includes defining a clear cybersecurity workforce strategy, investing in continuous learning, enabling reskilling at scale and promoting environments that encourage AI use and experimentation.

SME tip

Ensure that the right cybersecurity expertise is in place by training a small number of team members and use AI capabilities already available in existing tools and solutions.

2.3 How can the organization validate AI solutions before full deployment?

Validating AI through pilots is essential to reduce risk, demonstrate feasibility and prove value before enterprise-wide adoption. Executives need assurance that AI initiatives will deliver measurable operational improvements while protecting core business functions.

To ensure that pilots deliver actionable insights, CISOs should:

- Select pilots that deliver fast, tangible benefits or address high-priority cybersecurity objectives.
- Plan and structure pilots with key milestones, contingency time for unforeseen challenges, success criteria, fall-back strategies and go/no-go decision points to identify and exit non-viable projects quickly.
- Evaluate pilots against business-case validation criteria to determine viability and success.
- Promote cross-functional collaboration among cybersecurity, IT and business teams to ensure pilots address both security objectives and organizational priorities; this prevents solutions from becoming isolated experiments that fail to scale or deliver measurable impact.
- Allocate sufficient resources – including technology, dedicated time and initiatives – to upskill staff in AI for cybersecurity and support effective piloting.

SME tip

Run a fast, low-cost proof of concept to test whether the idea delivers real value in practice – and be ready to exit or pivot if the results don't clearly justify further investment.

2.4 How best to scale and maintain AI solutions while ensuring continuous optimization?

Executive leadership must have the confidence that AI can be scaled efficiently through a phased approach that balances innovation with operational stability. Since AI models can lose their effectiveness over time without ongoing monitoring and refinement, executives also need assurance that AI remains effective and aligned with organizational goals.

To scale AI and sustain its effectiveness and reliability, CISOs should:

- Validate that infrastructure and data environments can support AI workloads without security risks or cost spikes to ensure sustainable scaling.
- Assign clear ownership and accountability for the deployment, with dynamic governance that continuously monitors the evolving risks and adapts security guardrails and accountability models as technology evolves.
- Monitor performance and deterioration against established criteria to trigger proactive refinements, such as recalibrating algorithms and updating data pipelines to prevent model drift.
- Continue to build trust with executive leadership through transparent reporting, regular updates and clear evidence of AI-driven security improvements, and implement change management with staff, along with targeted training to accelerate AI adoption and scaling.
- Encourage participation in AI knowledge-sharing networks to strengthen situational awareness, accelerate learning and enhance organizational resilience through collaboration.
- Continuously evaluate emerging AI capabilities and technological developments to identify opportunities to enhance and optimize existing deployments.

SME tip

Scale only the initiatives that demonstrate clear, measurable benefits, and as solutions expand, regularly reassess their performance, cost implications and potential risks to ensure that they continue to support strategic business objectives.

3 The evolving landscape of AI in cybersecurity

By leveraging agentic AI, organizations can shift from a reactive to a pre-emptive approach in cybersecurity.

The future of cybersecurity is defined by a transition from reactive tools to autonomous collaborators, and organizations are embracing this shift rapidly. Some 88% of enterprises are actively investing in AI agents[18] and 92% of technology executives view AI agent management becoming a non-negotiable skill for the cybersecurity workforce within the next five years.[19]

3.1 The opportunity of agentic AI

Agentic AI allows cybersecurity to move from reacting to attacks towards more pre-emptive protection,[20] like an autonomous cyber immune system, identifying, disrupting and containing threats before they fully materialize.

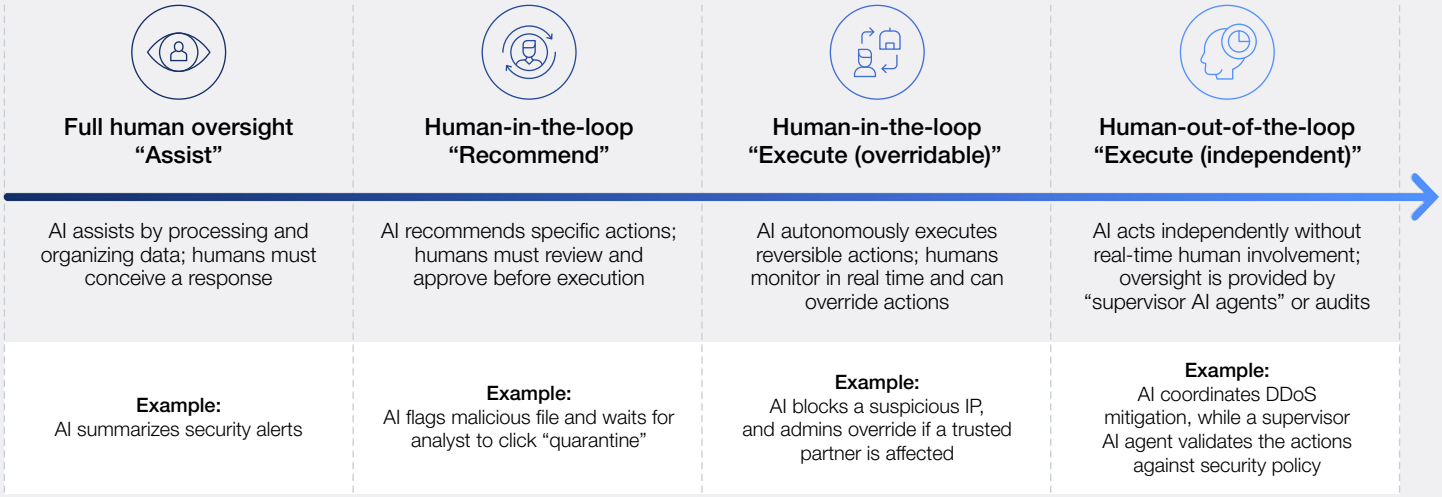
Gartner predicts that by 2028, 15% of day-to-day work decisions will be made autonomously by AI agents.[21] For cybersecurity, this means specialized agents collaborating across different functions such as threat intelligence, vulnerability management and incident response to achieve shared security objectives. These agents will increasingly plan, coordinate and execute complex cybersecurity tasks at machine speed and scale.

However, this increased autonomy introduces new risks. As cybersecurity agents assume more complex responsibilities, organizations must establish robust governance, clear accountability and meaningful human oversight.

3.2 Unpacking the spectrum of AI autonomy

The success of agentic AI in cybersecurity depends on matching the right level of autonomy to the task at hand. As agents transition from simple assistants to autonomous executors, organizations must define where AI can safely act independently and where human judgement remains a critical safeguard. Each level, as shown in Figure 3, involves a fundamental trade-off – machine-speed actions enable cybersecurity professionals to counter AI-driven threats, but reduce the human accountability and oversight needed to catch errors before they cause damage, raising ethical questions about reliability and responsibility. The appropriate choice is determined by the risk and reversibility of an action. High autonomy delivers efficiency gains for low-risk, reversible decisions while human oversight remains essential for high-stakes actions with lasting consequences, even at the cost of a slower response.

FIGURE 3 | **From helper to independent action: Four levels of AI autonomy in cybersecurity**



Source: World Economic Forum

3.3 Agentic AI risks and guardrails

While agentic AI in cybersecurity enables greater speed and efficiency for defenders, its adoption amplifies existing risks and introduces new ones that cannot be addressed through traditional security controls alone.

With the adoption of agentic AI, organizations face the following main challenges:

- An expanded attack surface where AI agents can introduce new entry points for exploitation or can be hijacked for destructive actions.
- Unintended agent behaviours driven by hallucinations, external manipulation or misconfigured objectives that can cascade across multi-agent environments at machine speed.
- Governance gaps where AI agents can be deployed rapidly without proper approval or validation, potentially taking undesirable actions with no clear accountability for the outcomes.

Together, these challenges underscore the need for new technical guardrails and governance models, including ethical considerations, tailored specifically to the realities of agentic AI. The World Economic Forum's report *AI Agents in Action: Foundations for Evaluation and Governance* offers concrete controls to support secure AI deployments.

Conclusion

AI has transitioned into a core enabler of modern cybersecurity, driven by the growing volume, speed and sophistication of threats that outpace traditional defences. For organizations that deploy it strategically, AI can augment human expertise, automate and accelerate security operations and help address some of the structural challenges in cybersecurity such as talent shortages, resource constraints and increasing regulatory demands.

The case studies in this white paper illustrate that AI adoption is most advanced in threat intelligence, vulnerability management, risk and anomaly detection and incident response. However, translating AI capabilities into value requires more than technological investment. Securing executive sponsorship is critical and depends on clearly linking AI initiatives to business objectives and measurable outcomes such as reduced risk exposure and improved service availability. Organizations must also assess foundational readiness, ensuring that data, infrastructure and governance frameworks can support AI deployment and that skills gaps are addressed.

Before full deployment, AI solutions should be validated through structured pilots with clear success criteria. Once deployed, continuous monitoring and refinement remain essential, with risks and guardrails regularly reassessed as threats and technologies evolve.

Looking ahead, agentic AI enables autonomous systems to detect and respond to threats before they fully materialize. That said, organizations must carefully determine the appropriate level of human oversight, from human-in-the-loop to fully autonomous operations, based on risk and reversibility of actions. At the same time, agentic AI introduces new risks that require robust guardrails throughout the agent life cycle.

AI offers unprecedented opportunities to strengthen cybersecurity, but realizing its potential requires a strategic approach, realistic expectations and vigilance against over-reliance, ensuring resilient workflows and preserving human expertise and judgement.

Contributors

Lead authors

Carlos Anastasiades
Senior Manager, Cybersecurity, KPMG

Chiara Barbeschi
Specialist, Technology and Innovation, Centre for Cybersecurity, World Economic Forum

Matthias Bossardt
Partner, Cybersecurity, KPMG

Lina Gehri
Senior Consultant, Cybersecurity, KPMG

Akshay Joshi
Head of the Centre for Cybersecurity; Member of the Executive Committee, World Economic Forum

Nataša Perućica
Lead, Capacity Building, Centre for Cybersecurity, World Economic Forum

Akhilesh Tuteja
Global Cyber Security Leader, KPMG

The World Economic Forum extends its gratitude to the members of the AI and Cyber: Empowering Defenders initiative for their valuable insights and expertise shared through a series of workshops and one-on-one interviews. Additionally, the Forum wishes to acknowledge the Chief Information Security Officer Community, the Partnership Against Cybercrime and Bridging the Cyber Skills Gap initiatives as well as the AI Global Alliance Safe Systems and Technologies Working Group.

Acknowledgements

Shakeel Ahmed
Chief Technology Officer and Group Head, Information Technology and Special Projects, Pathfinder Group

Ali Al-Amri
Director, Cyber Defense Operations, Aramco

Bushra AlBlooshi
Director, Governance and Risk Management, Dubai Electronic Security Center (DESC)

Abdullah T. Alessa
Head, Global Cyber Threat Intelligence, Aramco

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

Nebahat Arslan
Director, Group General Counsel and Partnership Officer, Women in AI

Romain Aviolat
Group Chief Information Security Officer, Kudelski Group

Lori Bailey
Head of Global Cyber & Technology, AXIS Capital

Eduardo Barbaro
Head of Security Analytics, ING Group

Kerry-Ann Barrett
Cybersecurity Section Chief, Organization of American States (OAS)

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

Nik Bartholomew
Vice-President, IT Cybersecurity & Risk Management, Occidental

Mauricio Benavides
Co-Founder and Chief Executive Officer, Batuta

Karine Ben-Simhon
Vice-President, Partnerships, Dream

Janus Friis Bindslev
Chief Digital Risk Officer, PensionDanmark

Ellen Boehm
Senior Vice-President IoT and AI Identity Innovation, Keyfactor

Stefan Braun
Chief Information Security Officer, Henkel

Ian Buffey
Chief Information Security Officer, AtkinsRéalis

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

Jorge Eduardo Ahumada Calbo
Head of Cyber Automation & AI, Banco Santander

Nicole Carignan
Senior Vice-President, Security and AI Strategy, Field CISO, Darktrace

Denise Cassidy
EMEA Security HR Lead, Accenture

Daniele Catteddu
Chief Technology Officer, Cloud Security Alliance (CSA)

Sebastian Cesario
Co-Founder and Chief Technology Officer, BforeAI PreCrime

Sonya Chan
Deputy Director of Emerging Technologies, Cyber Security Agency of Singapore (CSA)

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

David Chayer
Managing Director and Deputy Chief Information Security Officer, Depository Trust and Clearing (DTCC)

Piotr Ciepiela
Partner, EMEIA Cybersecurity Leader, EY

Giuseppe Cinque
Principal Architect, Cisco Digital Impact Office, Cisco

Ben Colman
Co-Founder and Chief Executive Officer, Reality Defender

Melonia da Gama
Director of Training and Learning Programs, Fortinet

Michael Daniel
President and Chief Executive Officer, Cyber Threat Alliance

Ayomide Daniels
Chief Scientist and Co-Founder, Cybervergent

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

Fabio Di Franco
Cybersecurity Officer, ENISA

James Dolph
Chief Information Security Officer, Guidewire

Yawen Duan
Technical Program Manager, Concordia AI

Nelia Argaz Durango
Head of Business Resilience and Cyber Advisory, Europe, Marsh

Marielle Ehrmann
Chief Security Compliance & Risk Officer

Gregory Eskins
Head, Global Cyber Insurance Center, Marsh

Charles Finlay
Founding Executive Director, Rogers Cybersecure Catalyst

Jon France
Chief Information Security Officer, ISC2

Sabrina Feng
Chief Risk Officer, Technology, Cyber and Resilience, LSEG

Laurent Gobbi
Partner, Global Head of Cyber & Tech Risk, KPMG

Vinayak Godse
Chief Executive Officer, Data Security Council of India (DSCI)

Prikshit Goel
Vice-President, Cyber Security and GRC Services, HCLTech

Randy Harold
Chief Information Security Officer, ManpowerGroup

Mark Hughes
Global Managing Partner, Cyber Security Services, IBM

Paul J
Technical Director for Cyber AI Research, National Cyber Security Centre (NCSC)

Manish Jain
Associate Vice-President, Information Security Group, Infosys

Lawrence Jarvis
Chief Information Security Officer, Iron Mountain Information Management

Terje Jensen
Senior Vice-President, Global Business Security Officer, Telenor Group

Rita Jonusaite
Government Affairs & Public Policy Manager, Google

Sam Kaplan
Director and Senior Global Policy Counsel, Palo Alto Networks

Steven Kelly
Chief Trust Officer, Institute for Security and Technology

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

Robert Kerby
Chief Technology Officer, Cisco Security and Trust Organization, Cisco

Dee Kimata
Cybersecurity Thought Leadership Director, Schneider Electric

Sigmund Kristiansen
Chief Cyber Security Officer, Aker BP

Erich Kron
CISO Advisor, KnowBe4

Aamir Lakhani
Global Strategist and Architect, Fortinet

Sebastian Lange
Chief Security Officer, SAP

Troy Leach
Chief Strategy Officer, Cloud Security Alliance (CSA)

Luigi Lenguito
Founder and Chief Executive Officer, BforeAI PreCrime

Anat Lewin
Global Lead, Digital Safeguards, World Bank

Oumou Ly
Non-Resident Research Fellow, AI Security Initiative, University of California, Berkeley

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

Nada Madkour
Non-Resident Research Fellow, University of California, Berkeley

Vanja Madzgalj
Senior Project Manager, Western Balkans Cyber Capacity Centre

Pilar Manchón
Senior Director, Engineering, Google

Danny Manimbo
Principal, Schellman Compliance

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

Sanjeev Mehrotra
Global Head, Cyber Security and Risk Management, Tech Mahindra

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

Michael Mestrovich
Chief Information Security Officer, Rubrik

Adam Meyers
Senior Vice-President of Counter Adversary Operations, CrowdStrike

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

Gilles Montagnon
Vice-President, Head of Security Awareness & Enablement, SAP

Jakub Olszewski
Head of Skills Academies, ICS, Standard Chartered

Ade Omotosho
Chief Executive Officer and Co-Founder, Cybervergent

Barbara O'Neill
Global Chief Information Security Officer, EY

Mark Orsi
Chief Executive Officer, Global Resilience Federation

Keith O'Sullivan
Senior Vice-President, Head of Cyber Defense, Tokio Marine

Tom Parker
Global Lead, Growth and Strategy, Cybersecurity Services, IBM

Haider Pasha
Chief Security Officer, EMEA & LATAM, Palo Alto Networks

Keri Pearlson
Executive Director, Cybersecurity, MIT Sloan Research Consortium, MIT – Sloan School of Management

Cezary Piekarski
Group Chief Information Security Officer, Standard Chartered Bank

Marco Pineda
Director, Cybersecurity, EY

Sarah Powazek
Program Director, Center for Long-Term Cybersecurity, University of California, Berkeley

Vijayeendra Purohit
Senior Vice-President and Chief Information Security Officer, Infosys

Thelma Quaye
Director, Infrastructure, Skills and Empowerment, Smart Africa Secretariat

Javier García Quintela
Chief Information Security Officer, Repsol

Cyril Reol
Group Chief Information Officer, Mercuria Energy Group

Humberto Luiz Ribeiro da Silva
Head, Center for Cyber Incident Prevention, Ciberlab, University of Brasilia

Reza Rooholamini
Chief Scientific and Artificial Intelligence Officer, CCC Intelligent Solutions

Jason Ruger
Chief Information Security Officer, Lenovo

Amir Abdul Samad
Head, Cyber Security (Chief Information Security Officer), PETRONAS (Petroliam Nasional)

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

Pascal Steichen
Chief Executive Officer, Luxembourg House of Cybersecurity (LHC)

Ina Steyn
Head, Security Education & Awareness, Absa Group

Mark Swift
Chief Information Security Officer, Trafigura Group

Sumit Taneja
Senior Vice-President; Global Head, Artificial Intelligence (AI) Consulting and Implementation, EXL Service

Josh Tomkiel
Managing Director, Schellman Compliance

Stanley Tsang
Engineer and Senior Director, Cyber Security Agency of Singapore (CSA)

Khalid Waheed
Manager of Artificial Intelligence, Dubai Electronic Security Center (DESC)

Fabian Willi
Head Cyber Key Accounts, Swiss Re

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

Jonathan Zanger
Chief Technology Officer, Check Point Software Technologies

Adam Zoller
Chief Information Security Officer, CrowdStrike

Production

Bianca Gay-Fulconis
Designer, 1-Pact Edition

Simon Smith
Editor, Astra Content

Endnotes

1. World Economic Forum. (2026). *Global cybersecurity outlook 2026*. https://reports.weforum.org/docs/WEF_Global_Cybersecurity_Outlook_2026.pdf
2. Darktrace. (2025). *The state of AI cybersecurity 2025*. https://www.darktrace.com/the-state-of-ai-cybersecurity-2025
3. IBM. (2025). *Cost of a data breach report 2025*. https://www.ibm.com/forms/mkt-53830
4. Palo Alto. (2026). *Global incident response report 2026*. https://www.paloaltonetworks.com/content/dam/pan/en_US/assets/pdf/unit42/Unit42-Global-Incident-Response-Report.pdf
5. Google. (2024). *Secure, empower, advance: How AI can reverse the defender's dilemma*. https://services.google.com/fh/files/misc/how-ai-can-reverse-defenders-dilemma.pdf
6. Sophos. (2025). *The human cost of vigilance: Addressing cybersecurity burnout in 2025*. https://assets.sophos.com/X24WTUEQ/at/n8gx8gk3p9tzrtbrv8gx7srh/sophos-the-human-cost-of-vigilance-addressing-cybersecurity-burnout-2025.pdf
7. ISACA. (2025). *State of cybersecurity 2025 report*. https://www.isaca.org/resources/reports/state-of-cybersecurity-2025
8. National Institute of Standards and Technology (NIST). (2024). *The NIST cybersecurity framework (CSF) 2.0*. https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.29.pdf
9. Ibid.
10. STRIDE (spoofing, tampering, repudiation, information disclosure, denial of service, elevation of privilege): a model for identifying and categorizing security threats.
11. OWASP. (n.d.). *OWASP Application Security Verification Standard (ASVS)*. Retrieved April 9, 2026, from https://owasp.org/www-project-application-security-verification-standard/
12. A globally recognized knowledge base of adversary tactics and techniques. MITRE ATT&CK is a globally recognized knowledge base of adversary tactics and techniques. MITRE (n.d.). *ATT&CK*. Retrieved April 9, 2026, from https://attack.mitre.org/
13. Sandboxing is a security technique that runs suspicious files or code in an isolated environment, separate from the main system, to safely analyse their behaviour without risk to the wider network.
14. Telemetry hunting is the process of actively searching through system and network data to identify threats that have not triggered automated alerts.
15. Canadian Centre for Cyber Security (n.d.). *Assemblyline4*. Retrieved March 23, 2026, from https://cybercentrecanada.github.io/assemblyline4_docs/
16. Samaroo, R., & Vigneault, J-P. (2024). *Supercharge your malware analysis workflow*. Virus Bulletin – Canadian Centre for Cyber Security. https://www.virusbulletin.com/uploads/pdf/conference/vb2024/papers/Supercharge-your-malware-analysis-workflow.pdf
17. World Economic Forum. (2026). *Global cybersecurity outlook 2026*. https://reports.weforum.org/docs/WEF_Global_Cybersecurity_Outlook_2026.pdf
18. KPMG. (2026). *KPMG global tech report 2026*. https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2026/01/global-tech-report.pdf
19. Ibid.
20. Gartner. (2025). *Gartner says that in the age of GenAI, preemptive capabilities, not detection and response, are the future of cybersecurity*. https://www.gartner.com/en/newsroom/press-releases/2025-09-18-gartner-says-that-in-the-age-of-genai-preemptive-capabilities-not-detection-and-response-are-the-future-of-cybersecurity
21. Gartner. (2025). *Gartner predicts over 40% of agentic AI projects will be canceled by end of 2027*. https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027

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