

Forensics digest

September 2026

AI-Driven defence in the age of AI-driven threats

Introduction

AI is no longer just a tool for efficiency; it is reshaping both sides of the cybersecurity battlefield. In 2026, organisations are facing a fundamental shift: the same technologies used to strengthen defence are now being weaponised to scale attacks faster, cheaper, and more intelligently than ever before.

Recent developments such as Claude Mythos, which was reported to have autonomously identified, chained and exploited zero-day flaws, show that organisations must move beyond traditional, human-paced security models towards intelligent, adaptive and continuously monitored defence.

AI-driven threat landscape

These global shifts are already materialising across the region. Our investigations show that threat actors are increasingly using AI to adapt malicious scripts to specific target environments, accelerate reconnaissance, and shorten the time between initial discovery and operational impact. For defenders, the central question is no longer whether adversaries are leveraging AI, but where, by whom, and at what pace it is being deployed. The sections that follow address these dimensions, mapping the evolving AI-Driven threat landscape.

1. Autonomous vulnerability discovery and exploitation

AI can scan large codebases, identify weaknesses and develop working exploits in hours rather than weeks. Automated reconnaissance can then move rapidly from an exposed service or weak authentication control to privilege escalation, lateral movement, fraud, disruption or data exfiltration. The practical consequence is a collapsing patch window: organisations have far less time to identify, assess and remediate vulnerabilities before they are weaponised.

2. Automated chaining of vulnerabilities

AI is not only accelerating the discovery of individual vulnerabilities but also enabling attackers to combine multiple weaknesses into coordinated attack paths that can lead to broader system compromise. AI models can analyse how misconfigurations, weak access controls, exposed credentials and software flaws interact, identifying efficient routes to privilege escalation, lateral movement or data exfiltration. Anthropic's internal red-team evaluations of Claude Mythos Preview reported an 83% success rate in weaving low-severity issues into critical control-flow hijacks.

The result is a more sophisticated and realistic form of attack planning at machine speed, where threat actors can move from reconnaissance to exploitation with far greater precision and efficiency. For organisations, this means that vulnerability management can no longer focus solely on individual flaws; it must also consider how separate weaknesses could be exploited in combination to create material business risk.

3. Democratisation of advanced cyber attacks

Based on our work across the region, we have observed instances where threat attackers are leveraging AI to generate malicious scripts, reducing the level of technical expertise required to develop or adapt tools for specific environments.



Threat actors can now generate targeted phishing, clone voices, create deepfake video and adapt malicious scripts for specific environments using capabilities that were once limited to highly skilled hackers or nation-state groups.

Threat actors can now generate targeted phishing, clone voices, create deepfake video and adapt malicious scripts for specific environments using capabilities that were once limited to highly skilled hackers or nation-state groups. This broadens both the volume and sophistication of adversaries organisations must defend against.

This shift is effectively "democratising" cyberattacks, enabling a significant increase in both the volume and sophistication of threats. Interpol's 2026 Global Financial Fraud Threat Assessment reveals that financial fraud has surged into an industrialised, AI-driven global crisis, citing a 54% rise in fraud-related international notices, AI-enabled scams that are 4.5 times more profitable than traditional methods, and the global expansion of forced-labour scam centres. The same industrialisation is visible closer to home: the Communications Authority of Kenya recorded 3.37 billion cyber threat events in Q1 2026 alone, this translates to 433 attacks every second. Communications Authority of Kenya attributes the persistent threat environment to "inadequate system patching, insufficient user awareness of phishing and other social engineering threat vectors, and the increasing exploitation of AI-driven and machine-learning technologies by malicious actors". Consequently, organisations are no longer just defending against a few advanced adversaries, but against a broader and more capable threat landscape with heightened risk exposure.

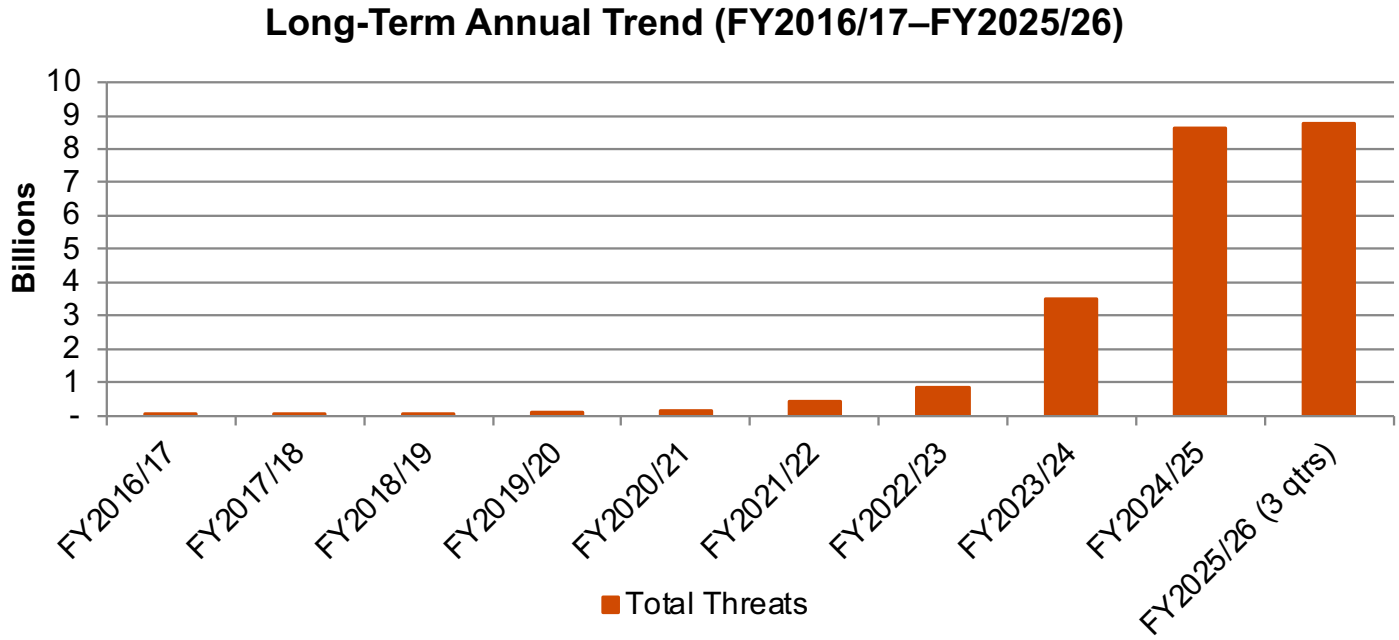


Figure 1: Kenya cyber-threat detections, annual trend (FY2016/17 to FY2025/26).
Source: Communications Authority of Kenya

AI has transformed cyber fraud and social engineering into repeatable, scalable operations. Localised and personalised phishing is harder to distinguish from legitimate communication, while deepfake audio and video can impersonate executives, authorise payments or bypass identity checks.

4. Industrialisation of Fraud and Social Engineering

AI has transformed cyber fraud and social engineering into repeatable, scalable operations. Localised and personalised phishing is harder to distinguish from legitimate communication, while deepfake audio and video can impersonate executives, authorise payments or bypass identity checks. Fraud-as-a-Service platforms add phishing kits, malware, stolen data, document templates and infrastructure, allowing criminal groups to buy or rent capabilities.

Based on our experience, these platforms are already in use within the Eastern Africa region as reflected in a specific investigation we conducted, where the threat actor used a phishing website that enabled them to tailor a phishing email to bypass internal control mechanisms.

As a result, fraud campaigns are becoming more coordinated, repeatable, and commercially efficient, while investigations are made more difficult by the distributed nature of the actors, infrastructure, and financial flows involved.



Key Challenges for Organisations

Most traditional security frameworks were designed for relatively static environments, such as endpoints, users, servers, and applications. AI systems, by contrast, are dynamic, adaptive, and continuously evolving. This creates new risks that many organisations are struggling to understand, monitor, and control. As AI changes the cyber threat landscape, organisations are facing four key gaps: speed, skills, visibility, and trust:

a. Speed gap

AI-enabled attacks can identify targets, generate convincing content, test defences and adapt in real time. Thousands of personalised phishing messages can be produced within minutes and refined as users respond. Security processes must therefore detect, decide and respond faster.

b. Skills gap

AI-enabled attacks can identify targets, generate convincing content, test defences and adapt in real time. Thousands of personalised phishing messages can be produced within minutes and refined as users respond. Security processes must therefore detect, decide and respond faster.

AI adoption is outpacing organisations' ability to secure and govern it. During development, AI models may be exposed to risks such as dataset poisoning, adversarial attacks, or insecure training practices. During deployment, inadequate controls can lead to data privacy breaches, regulatory non-compliance, unauthorised access to sensitive information, and misuse of AI-enabled systems.

The challenge is not limited to headcount, but extends to whether existing teams have the specialist capabilities required to assess AI security, govern its use and respond to AI-enabled threats. The 2026 Fortinet Cybersecurity Skills Gap Report found that 91% of organisations are using or experimenting with AI-powered cybersecurity solutions, yet the cybersecurity skills gap continues to grow as AI changes the capabilities required across security teams. Without targeted upskilling in areas such as AI risk assessment, secure model deployment, threat detection and incident response, organisations may adopt AI faster than they can secure or oversee it.

c. Visibility gap

Unlike traditional applications, AI systems process natural language, interact with users, access data sources, and integrate with business workflows, creating additional pathways for exploitation. Many organisations lack visibility into where AI is deployed, what data it can access, which services it connects to, and what permissions it has, resulting in security blind spots and misconfigurations.

Key risks include prompt injection, where attackers manipulate AI behaviour through crafted inputs and data exfiltration, where attackers use AI tools to access, summarise, or transfer sensitive information.

Limited visibility is also linked to excessive agency, where AI systems are granted broad permissions with inadequate oversight or privilege management and shadow AI, where employees use unapproved AI solutions outside governance frameworks, potentially exposing sensitive data and creating regulatory, privacy and intellectual property risks.

d. Trust gap

Deepfakes and synthetic identities undermine document, facial and voice verification. By blending real and fabricated information, criminals can pass onboarding or account-recovery checks and create credible personas for fraud. Organisations must assume that seeing or hearing a familiar person is no longer sufficient proof of identity and reinforce verification through independent, layered controls.



Implications of AI-Driven Threats

AI has significantly increased the sophistication, scale, and effectiveness of cyberattacks. Unlike traditional cyber threats, AI-enabled attacks use automation, adaptive learning, and data-driven precision to operate faster and more convincingly. The implications extend beyond technical risk and can affect financial stability, operational resilience, regulatory exposure, stakeholder trust, and significantly increase the costs associated with cybersecurity, risk management, compliance, and incident response.

a. Increased financial losses due to high-precision fraud and automated attacks

Based on our experience conductingl fraud investigations across the region, we have observed that AI-enabled fraud is increasing financial exposure through synthetic identities, manipulated documents, biometric fraud, and increasingly sophisticated social engineering techniques.

These capabilities enable threat actors to scale attacks against digital onboarding, verification, and payment processes, increasing both the likelihood and impact of fraud losses.

b. Increased cybersecurity, risk management and compliance costs

As AI-enabled attacks become more sophisticated, organisations are being forced to invest in advanced security capabilities, including AI-enabled monitoring and detection, behavioural analytics, identity verification, threat intelligence and continuous security testing. Evolving expectations for cyber resilience, data protection and AI governance also increase the resources required for compliance, oversight and incident response. Consequently, organisations face rising expenditure as they seek to maintain resilience against increasingly automated and adaptive threats.

c. Higher risk of data breaches and system compromise at scale

AI-enabled tools can scan networks, identify weaknesses, test credentials, and adapt attack strategies in real time. This increases the likelihood of successful compromise, particularly where organisations rely on basic or static security controls. AI-assisted credential stuffing attacks, for instance, can test large volumes of stolen credentials across multiple platforms, including streaming services, financial applications, and enterprise systems.

d. Operational disruption as attacks become continuous and automated

Automated campaigns can operate continuously, test defences and refine tactics without direct human intervention, prolonging downtime, disrupting supply chains and increasing pressure on already constrained security teams.

e. Significant reputational damage driven by trust-based exploits (e.g., deepfake scams)

AI-enabled threats increasingly exploit trust rather than just technical vulnerabilities. Deepfakes and synthetic media can trigger unauthorised transactions, false public statements and loss of confidence before they are disproved. Customers and investors may interpret a successful trust-based attack as a failure of governance and control.



AI-Driven Defence: What Organisations Must Do

Building resilience in this new era requires organisations to rethink how they detect threats, manage cyber exposure and establish trust across their digital environments.

To respond effectively, organisations must adopt adaptive, intelligence-driven security models;

AI-Powered Fraud Risk Management; Evolve beyond periodic control assessments to continuous, intelligence-led fraud risk management. AI and advanced analytics can detect anomalous patterns, assess inherent risk against existing controls and prioritise mitigation before losses materialise.

At PwC, we leverage an AI-enabled Fraud Risk Management Tool as part of Fraud Risk Assessments to support organisations in strengthening fraud risk governance and enhancing the effectiveness of their control environment. By assessing inherent fraud risks against existing controls, organisations gain actionable insights that strengthen governance, enhance internal controls and improve organisational resilience against increasingly sophisticated fraud schemes.

AI-Powered Detection and Response; Strengthen Security Operations Centres with AI-enabled SIEM, behavioural analytics and anomaly detection. Human analysts can then focus on investigation and judgement rather than manual triage, enabling faster containment of attacks operating at machine speed.

Continuous Threat Exposure Management; Adopt Continuous Threat Exposure Management: continuous scanning, risk-based patching, attack-path analysis and ongoing control validation.

Identity and Trust Reinforcement; Rebuild identity assurance on the premise that video and voice can be fabricated. Use multi-factor authentication, behavioural or biometric checks and independent out-of-band verification for high-risk payments, privileged access and changes to payment instructions.

AI-driven threats are already redefining cybersecurity. Organisations that combine intelligent detection, continuous exposure management, secure-by-design systems, strong identity controls,s and informed human judgement and timely incident investigations will be better positioned to protect operations and stakeholder trust.

Security by Design; Embed security in applications, payment workflows, identity systems, onboarding, APIs, third-party integrations and AI systems from the outset rather than treating it as an afterthought or a compliance layer added after deployment.

Human + AI Collaboration; Train staff across finance, procurement, customer service, executive offices and security teams to recognise personalised phishing, deepfakes and synthetic identities. Resilience depends on combining AI-enabled detection with human judgement, independent verification and disciplined approval controls.

Incident response and investigation; Establish clear and tested incident response procedures supported by competent incident response and forensic teams equipped with the appropriate technology. Following an incident, organisations should preserve relevant evidence, contain the threat, determine the attack path and root cause, assess the full business and regulatory impact, and identify control improvements required to prevent recurrence. AI-enabled forensic and analytical tools can accelerate evidence review and anomaly detection, but findings should remain subject to validation and professional judgement.

Conclusion

AI-driven threats are already redefining cybersecurity. Organisations that combine intelligent detection, continuous exposure management, secure-by-design systems, strong identity controls,s and informed human judgement and timely incident investigations will be better positioned to protect operations and stakeholder trust.

For further guidance and support, don't hesitate to reach out to us. We are here to provide expert insights and dedicated support tailored to your institution's unique needs. Let's navigate through these challenges together and ensure your business remains resilient and compliant in the face of evolving cyber fraud risk.



Contact us



John Kamau
Partner, Forensics Services
john.kamau@pwc.com



Isaac Gachigua
Manager, Forensics Services
isaac.gachigua@pwc.com



Josephat Mwanzia
Manager, Forensics Services
josephat.mwanzia@pwc.com



Evelyne Mwende
Senior Associate, Forensics Services
evelyne.mwende@pwc.com



Angela Nadine
Senior Associate, Forensics Services
angela.nadine@pwc.com

This publication has been prepared as general information on matters of interest only, and does not constitute professional advice. You should not act upon the information contained in this publication without obtaining specific professional advice.

© 2026 PricewaterhouseCoopers Limited. All rights reserved. In this document, “PwC” refers to PricewaterhouseCoopers Limited which is a member firm of PricewaterhouseCoopers International Limited, each member firm of which is a separate legal entity.