pwc

Generative AI and the reshaping of cyber-threat tactics

Table of contents

01 The shift in the threat landscape

The shift in the threat landscape

Cyber threats are becoming more sophisticated, scalable and adaptive. Persistent geopolitical disruption has driven a rise in cyber-attacks across the region, including those targeting businesses and critical infrastructure. Generative AI is accelerating this shift, enabling attacks to be launched faster, scaled more easily and carried out with greater credibility.

As the Middle East advances in digital government, smart infrastructure, connected services and AI adoption, the opportunities created by technology are growing, but so are the risks. The potential impact of AI-enabled attacks on critical infrastructure, financial systems and government services is becoming harder to ignore. Nor is that risk static. It reflects a broader shift in cyber threats, from slower, manual attacks to highly automated and increasingly AI-enabled operations. Earlier, cyber-attacks were highly manual. Adversaries conducted reconnaissance using basic network interrogation techniques, including WHOIS lookups,[1] banner grabbing and protocol probing with tools such as Telnet[2] or Nmap.[3] Exploits were painstakingly crafted in assembly or C. Campaigns were slow and limited by the attacker's skill and effort.[4]

In the mid-2000s came the automation phase. 'Exploit' frameworks, such as Metasploit, large-scale botnets and automated scanners like Nmap and Nessus, industrialised what had once been artisanal. Early forms of malware-as-a-service and crimeware kits lowered the barriers to entry, making it possible for less skilled actors to launch campaigns targeting thousands simultaneously.

Social engineering is also being scaled during this time. Attackers moved from hand-crafted phishing to bulk email blasts and automated spear phishing kits. This period shifted the threat landscape from niche, targeted intrusions to industrialised cybercrime.

Now a new phase is underway. GenAI has become a powerful force multiplier, accelerating the scale and perceived authenticity of offensive operations. It can create highly convincing spear-phishing emails and smishing campaigns in a fraction of the time required by a human, making social engineering more efficient and harder to detect.

PwC's Annual Threat Dynamics 2026 report has indicated that AI is now being used across multiple stages of the attack lifecycle, including reconnaissance, social engineering, malware development and data exploitation. The point is not simply that attackers have new tools, but that the time between new AI capabilities emerging and their operational use is narrowing.[5] That has direct implications for defenders: the pace, scale and adaptability of attacks are increasing, while the window for detection and response is shrinking.

This shift is not a single step change but a series of accelerating waves. Each wave of AI capability reduces the time, cost and expertise required to execute attacks, while increasing their scale and precision. For defenders, this means progressively shorter windows to detect, respond and recover.

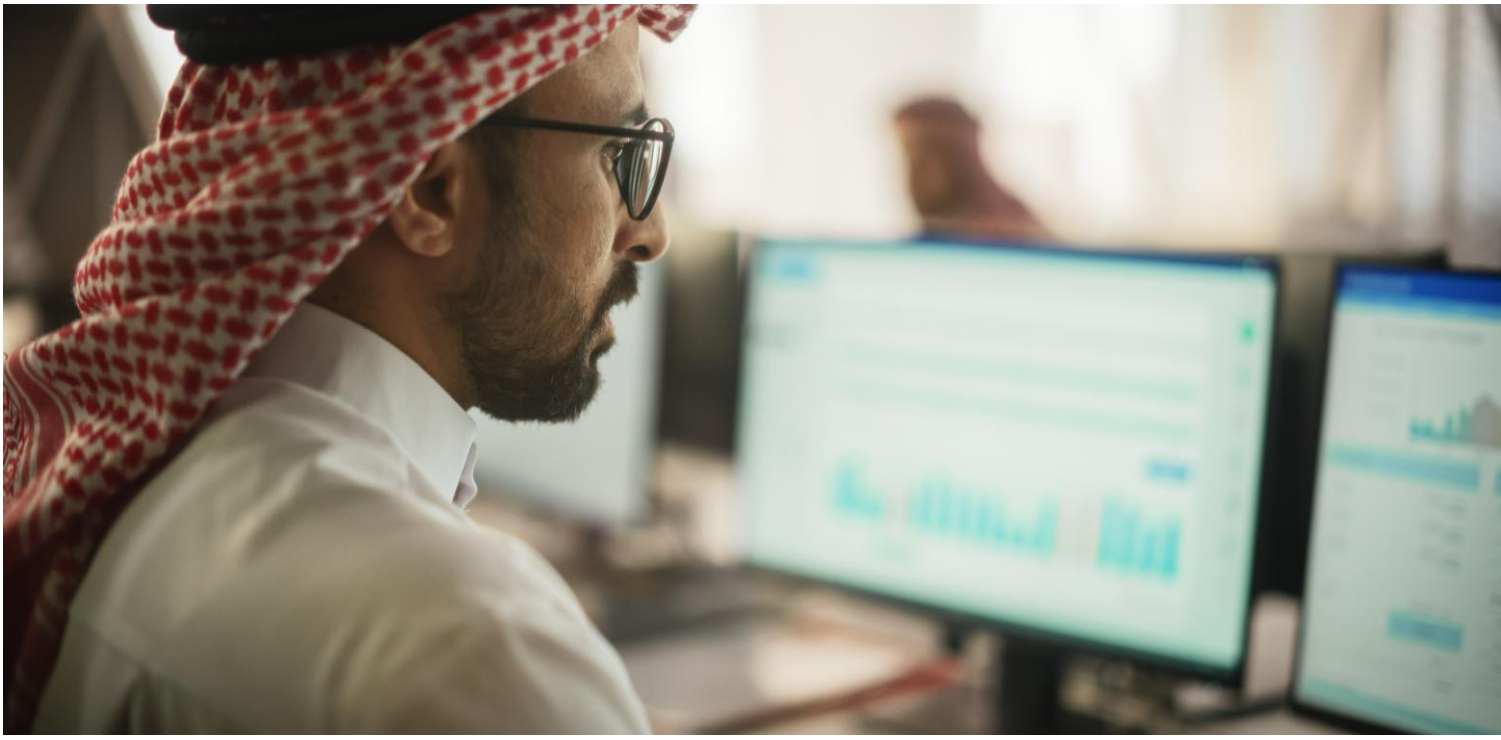
The combination of AI-driven automation and social engineering at scale is changing how attacks are executed and how quickly they spread. Attacks are now faster, more adaptive and more convincing than ever before. Defenders must recognise that this is not an incremental change but a redefinition of offensive capability. Creativity, scale and authenticity can all be manufactured by machine.

The threat evolution timeline below (see Figure 1) captures this progression from the manual era to automation and now to AI-powered operations, underscoring why defenders must rethink how they prepare for attacks shaped by machine speed and scale.

Figure 1

Threat evolution timeline

Manual hacking era	Automated tools era	Generative AI era
Pre -2000s - Early 2000s	Mid-2000s - 2020	2021 - Present
• Human-driven attacks • Limited scale • Low automation	• Tool-based attacks • Increased scale • Higher automation	• AI-powered attacks • Personalised scale • Adaptive automation



Across the region, governments are strengthening cyber resilience through a mix of national strategy, regulatory controls and institutional coordination.

Qatar	Qatar is embedding cybersecurity within wider state-led digital transformation through its Digital Agenda 2030[7] and the Qatar Digital Government NextGen Strategy.[8]
Saudi Arabia	The National Cybersecurity Authority (NCA) has shaped a unified cybersecurity posture in the Kingdom through the development of national policies, essential cybersecurity controls and sector-wide compliance frameworks. The Kingdom has helped raise security maturity across critical infrastructure and government entities by prioritising governance, risk management, capacity building and public-private coordination.
UAE	The UAE's National Cybersecurity Strategy sets out a coordinated national approach to strengthening cyber resilience. It reflects a broader push to improve national preparedness, protect critical systems and build greater resilience against a rapidly evolving threat landscape.

02 Types of AI-powered offensive attacks

Types of AI-powered offensive attacks

Rather than simply automating traditional attack methods, AI has enabled entirely new categories of threats that exploit machine learning capabilities to create more sophisticated, scalable and evasive attack vectors. For security leaders and practitioners, understanding these AI-enhanced offensive tactics is essential to anticipating attacker behaviour, adapting defensive models and strengthening defences.

2.1 AI-enhanced social engineering and phishing

Modern AI-powered phishing campaigns leverage large language models (LLMs) to create messages that are grammatically accurate, contextually appropriate and highly personalised.

These systems can analyse vast amounts of publicly available data – including social media, corporate communications and professional networks – to craft messages that appear to come from trusted sources and reference specific, current information relevant to the target.

AI-generated phishing emails can improve engagement rates by incorporating personal touches such as recent purchases, upcoming business deals, or even referencing the target's recent social media activity.[9]

2.2 Deepfakes and synthetic identity abuse

Deepfake technology is pushing social engineering into a more convincing and harder-to-verify phase. Attackers now clone voices using as little as three to 15 seconds of audio grabbed from public sources to convincingly impersonate CEOs, colleagues or family members. Evidence across regulated industries illustrates the scale of this challenge. In the financial sector, 38% of incidents targeting individuals involved phishing, smishing or vishing techniques, while 36% of attacks against credit institutions relied primarily on social engineering rather than technical exploitation.[10] These patterns reinforce that human trust, identity and verification controls have become primary targets in AI-enabled attack campaigns.

2.3 AI-accelerated vulnerability discovery and exploitation

AI is now transforming how vulnerabilities are found and exploited, automating tasks that once took even advanced teams weeks to complete. Machine learning-powered code analysis tools can scan complex applications at scale, identifying bugs and security flaws across multiple languages and frameworks. In many cases, they can even uncover previously unknown vulnerabilities that human analysts might miss.

AI-driven analysis and fuzzing[11] can reduce exploit development time from around 168 hours to less than 24 hours. Some research frameworks have demonstrated automated exploit generation in under a minute for known vulnerability types.

GenAI and LLMs are now being integrated into offensive toolchains, such as PwnGPT, which connects vulnerability identification modules to language models capable of finding flaws then build, test and refine proof-of-concept exploits automatically. This has lowered technical barriers to exploitation, enabling attackers with limited skills to launch sophisticated code-reuse or memory corruption attacks that were formerly reserved for elite threat actors and nation-state teams.[12, 13]

The implication is that zero-day exploitation windows are likely to compress, increasing pressure on patch governance and continuous monitoring. Recent advances in so-called “Mythos-class” AI capabilities demonstrate the direction of travel. These systems have shown the ability to autonomously identify vulnerabilities across large codebases, generate working exploits without human guidance and orchestrate elements of the attack chain at scale. In controlled environments, thousands of vulnerabilities have been identified across operating systems and browsers, with exploit generation occurring in a fraction of the time previously required. This represents a structural shift. The time between vulnerability discovery and weaponisation is collapsing, from weeks or days to hours, creating a persistent asymmetry where attackers can act faster than organisations can patch or respond.

2.4 Adaptive AI-powered malware and evasion

AI-powered malware is beginning to incorporate more dynamic mutation techniques, altering elements such as file hashes, imports, headers and instruction sequences each time it replicates or executes. Unlike traditional polymorphic malware that relied on static encryption or packers, AI-driven variants dynamically rewrite their code structure using machine learning, creating countless unique variants that retain the same functionality while evading static signatures.

This dynamic mutation renders most signature-based antivirus solutions ineffective, as no persistent patterns exist for reliable detection.

Furthermore, AI-enhanced malware incorporates adaptive evasion techniques such as anti-virtualisation, anti-debugging, environmental awareness and intelligent obfuscation to further complicate detection efforts. Notably, some malware uses adversarial machine learning to trick AI-driven defence tools by embedding prompt injections that mislead the analysis process into misclassifying malicious code as benign.

These advancements underscore the urgent need for security strategies that go beyond signatures, focusing on behavioural and anomaly detection supported by AI-driven defences.[14, 15, 16]

2.5 Adversarial attacks against AI systems

Beyond exploiting systems and users, attackers are increasingly targeting AI itself. Adversarial AI attacks target the very foundations of machine learning systems by poisoning training data or crafting deceptive inputs that cause AI models to make incorrect decisions. Through data poisoning, attackers inject subtle manipulated samples into training datasets that systematically degrade model accuracy or embed hidden backdoors triggered only under specific conditions.

These alterations can lead to models consistently misclassifying critical inputs, such as ignoring fraudulent transactions or misdiagnosing diseases, while appearing normal in routine validation. Furthermore, adversarial examples exploit vulnerabilities in model decision boundaries by adding imperceptible perturbations to inputs, misleading AI systems across domains like healthcare, finance and autonomous vehicles.

Defending against these sophisticated attacks requires robust data validation, adversarial training and anomaly detection focused on preserving model integrity in increasingly hostile environments.[17]

03 Offensive team applications

Offensive team applications

3.1 Using AI to increase realism in offensive testing

‘Red teams’ or authorised cybersecurity specialists can use many of the same AI capabilities as adversaries, but under strict controls, to increase test realism and strengthen defences. The distinction is not technical, it’s procedural. In a legitimate offensive engagement, AI is applied within an explicit scope, with clear security parameters and with measurable objectives, so that increased speed and realism translate into better detection, stronger responses and fewer blind spots.

When applied well, AI does not simply make penetration testing faster. It makes testing more representative of the current threat environment, where scale, personalisation and iteration are no longer limiting factors.

A practical approach is to treat AI as an integrated layer within the penetration testing workflow, rather than a standalone tool. Established guidance, such as NIST SP 800-115,[18] presents security testing as a structured process involving planning, execution, analysis and mitigation.

AI can support each phase of testing, but it delivers most value when embedded as a set of targeted capabilities that improve transitions between steps. By using AI for translation, triage and documentation, teams can reduce the time spent converting raw tool outputs into concrete plans while maintaining human review for decisions involving operational risk.

In engagement planning, AI is most effective when it structures complexity rather than introducing additional challenges. With a clearly defined scope, constraints and business objectives, AI can help generate candidate scenarios, test hypotheses and repeatable runbooks for refinement by the red team. This process includes translating objectives, such as validating identity controls or assessing the detection of post-compromise behaviour, into executable, measurable test paths.

A significant advantage is the consistency achieved through AI-assisted planning, which enables rapid creation of repeatable runbooks, checklists and documentation templates. The established consistency facilitates comparison across business units, environments, or time periods. While the red team retains responsibility for the final plan, AI reduces the effort required to transform intent within an organised exercise design.

3.2 AI-assisted reconnaissance and analysis

Reconnaissance often takes a lot of time, as teams must sift through large volumes of data to identify what is relevant. AI is especially effective at condensing and organising this information into prioritised leads. For example, AI can summarise open-source intelligence (OSINT) findings to identify potential attack surfaces, categorise discovered services by risk and highlight anomalies warranting further validation.

Recent research in automated penetration testing has demonstrated the use of LLMs to support intelligence gathering and subsequent phases, frequently employing modular designs to minimise context loss and enhance task continuity. Importantly, AI does not replace traditional reconnaissance expertise. It reduces the mechanical workload of parsing, correlating and documenting findings, enabling testers to concentrate on validating material threats.

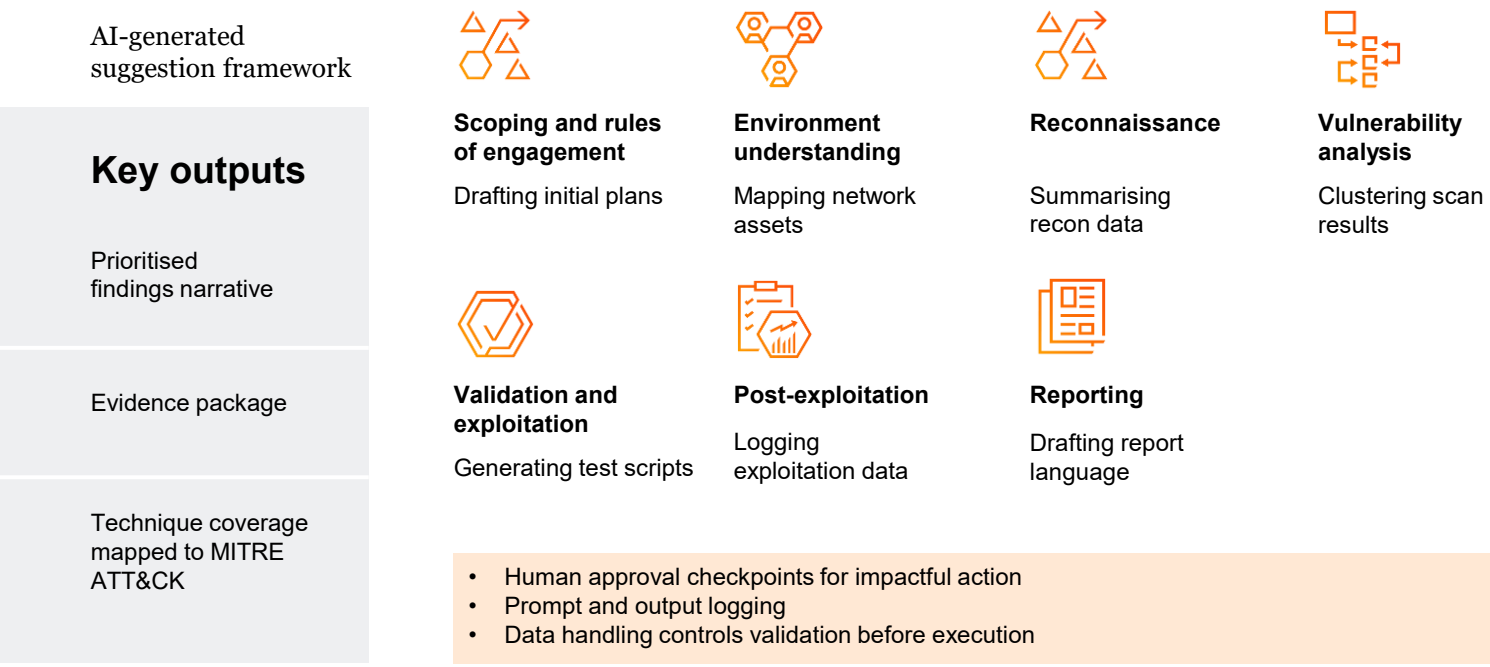
This approach also applies to vulnerability analysis and exploitation preparation. While most organisations can generate vulnerability data at scale, the primary challenge lies in identifying findings that accurately reflect attacker behaviour and safely demonstrating their impact.

AI can assist by clustering related weaknesses, suggesting plausible attack chains and generating structured hypotheses for tester verification. This capability is especially valuable in complex environments where identity misconfigurations, exposed management interfaces and cloud control plane permissions interact in intricate ways.

Research initiatives such as PentestGPT[19] and PentestAgent[20] demonstrate that large-language-model-driven frameworks can automate aspects of intelligence gathering, vulnerability analysis and exploitation. In operational settings, the most effective use of AI is to accelerate the reasoning process, while ensuring that all AI-generated suggestions are subject to standard validation before acceptance (see Figure 2).

Figure 2

AI-generated suggestion framework



3.3 Automated adversary emulation

AI delivers most value when it moves offensive testing from a manual process to a repeatable, automated workflow capable of executing significant parts of the attack chain.

This shift is not only about speed. It also allows testing to better reflect how real intrusions unfold. MITRE's CALDERA,[21] for example, is an automated red-team system that reduces the resources needed for routine testing by using a decision engine, agents and ATT&CK-based profiles to replicate adversary behaviours.[22]

Such capabilities change red team operations. Rather than treating each engagement as a unique exercise, teams can turn validated behaviours into repeatable profiles, automate execution and reserve human expertise for tasks requiring creativity, judgment and advanced exploitation skills.

This matters because modern attacks are rarely linear. They involve branching decisions, trade-offs and pivots. Defenders must therefore be assessed on behaviours rather than static indicators. Automated adversary emulation supports this by chaining post-compromise techniques and identifying detection or response failures.

AI can also introduce controlled variability by adjusting timing, sequencing and environmental conditions, helping prevent exercises from becoming predictable. While the red team still aligns each action with ATT&CK techniques and approved objectives, AI reduces the operational burden that limits how often these simulations can be run.

3.4 APT simulation and campaign-based testing

Simulating advanced persistent threat (APT) behaviour is a logical progression of automation, but it demands rigour to prevent the term from becoming vague. MITRE's ATT&CK resources and adversary emulation plans provide a structured basis for modelling documented threat behaviours.

The APT3[23] emulation plan, based on a documented threat group associated with state-sponsored cyber activity, shows how an adversary lifecycle can be represented from initial compromise to exfiltration. For red teams, such plans are valuable because they shift testing from isolated techniques to campaign-level sequences, allowing persistence, privilege escalation, discovery, lateral movement, collection and exfiltration to be assessed together.

AI can accelerate the conversion of these mappings into executable plans tailored to the environment, while keeping technique selection anchored in documented behaviour patterns.

3.5 AI-enhanced phishing and exploit validation

AI also enhances social engineering exercises by removing many of the indicators that once made phishing attempts easier to detect. The value lies in testing whether personnel and processes remain effective when confronted with polished, contextually relevant and psychologically convincing messages.

IBM X-Force found that a GenAI model could produce convincing phishing emails in around five minutes, compared with roughly 16 hours for the team's traditional process, while performing comparably to human-crafted versions in A/B testing.[24] This makes it easier to test controls such as out-of-band verification, identity and access safeguards and reporting mechanisms under conditions that more closely resemble AI-enabled attacks.

For red teams, the value lies in the capacity to generate scalable, plausible variants within ethical and approved boundaries. That makes it easier to test procedural controls such as out-of-band verification, identity and access safeguards and reporting mechanisms, under conditions that closely resemble those created by AI-enabled adversaries.

AI is also beginning to affect exploit development and validation in controlled environments.

Most credible advances remain research-led and benchmark-tested rather than proven in production. Frameworks such as PwnGPT suggest that modular, verification-focused approaches can improve performance on benchmark exploit challenges, reducing the iteration cost of some exploit development tasks in controlled settings.

For red teams, these capabilities should be used as accelerators for analysis and proof-of-concept work under strict authorisation, not as autonomous exploitation engines. Applied carefully, AI can shorten the time between identifying a potential weakness and validating the risk, improving both the quality of findings and the credibility of recommendations.

3.6 Ethical, governance and operational safeguards for AI-enabled testing

Given AI's potential to enhance offensive capabilities, ethical and operational safeguards must be integral to the methodology. The NIST AI Risk Management Framework[25] underscores the importance of governance and lifecycle risk management to ensure trustworthiness, while NIST's GenAI profile extends these principles to generative systems.[26]

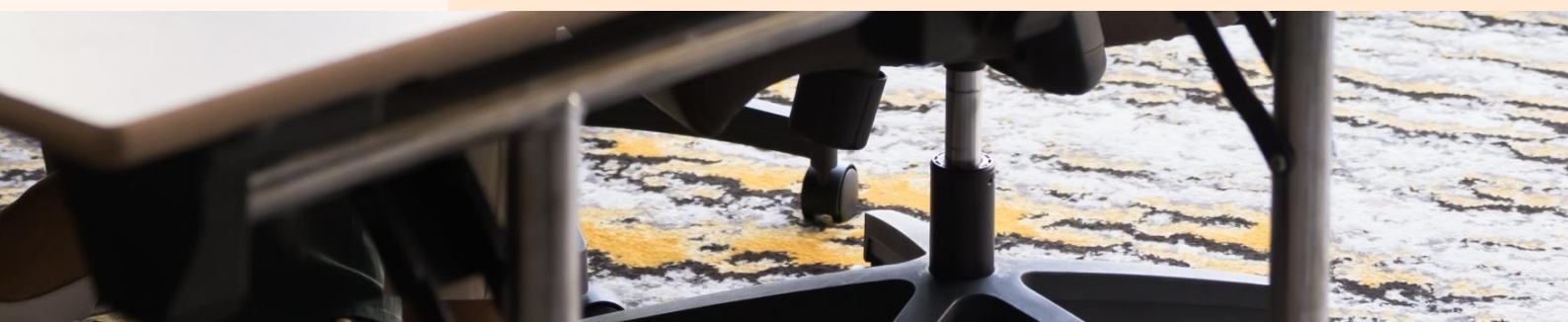
For red teams, this requires explicit decisions about data sharing with AI systems, output logging, approval processes for high-impact actions and safety controls to prevent scope drift. Disciplined transparency is also essential, including documentation of AI usage, outputs, human validation or rejection of results and the management of AI-related risks during engagements.

When these safeguards are integrated, AI supports a more advanced offensive operating model. Incorporating AI into workflows reduces translation and documentation time, enabling more thorough validation within the same engagement period.

Automation also enables repeatable execution of attack chains, allowing detection and response to be tested against actual behaviours rather than simply the presence of vulnerabilities.

APT-style emulation also becomes more practical when published frameworks and emulation plans provide structure and AI minimises the effort required to adapt these models to specific environments. The objective is not automation for its own sake. It is to identify where a real intrusion could succeed, then give the organisation time and evidence to act before an adversary does.

04 Real-world insights



Real-world insights

A credible discussion of offensive AI requires grounding in both documented incidents and controlled research studies. This section brings together three complementary forms of evidence: documented cases of deepfake-enabled fraud, findings from a controlled phishing study and the emergence of malicious large language models such as FraudGPT. Together, they show how offensive AI is already reshaping deception, scale and accessibility in cyber-attacks.

4.1 Evidence A: Deepfake-enabled fraud in practice

One of the most well-documented incidents occurred in early 2024 at a global engineering consultancy. A finance employee in the company's Hong Kong office joined what appeared to be a legitimate video call with the firm's CFO and other colleagues, only to discover later that every participant except the employee had been AI-generated.[27] Convinced by the authenticity of the meeting, the employee executed 15 transfers to five separate accounts, resulting in losses of approximately HK$200m (around US$25m), shown below (Figure 3).

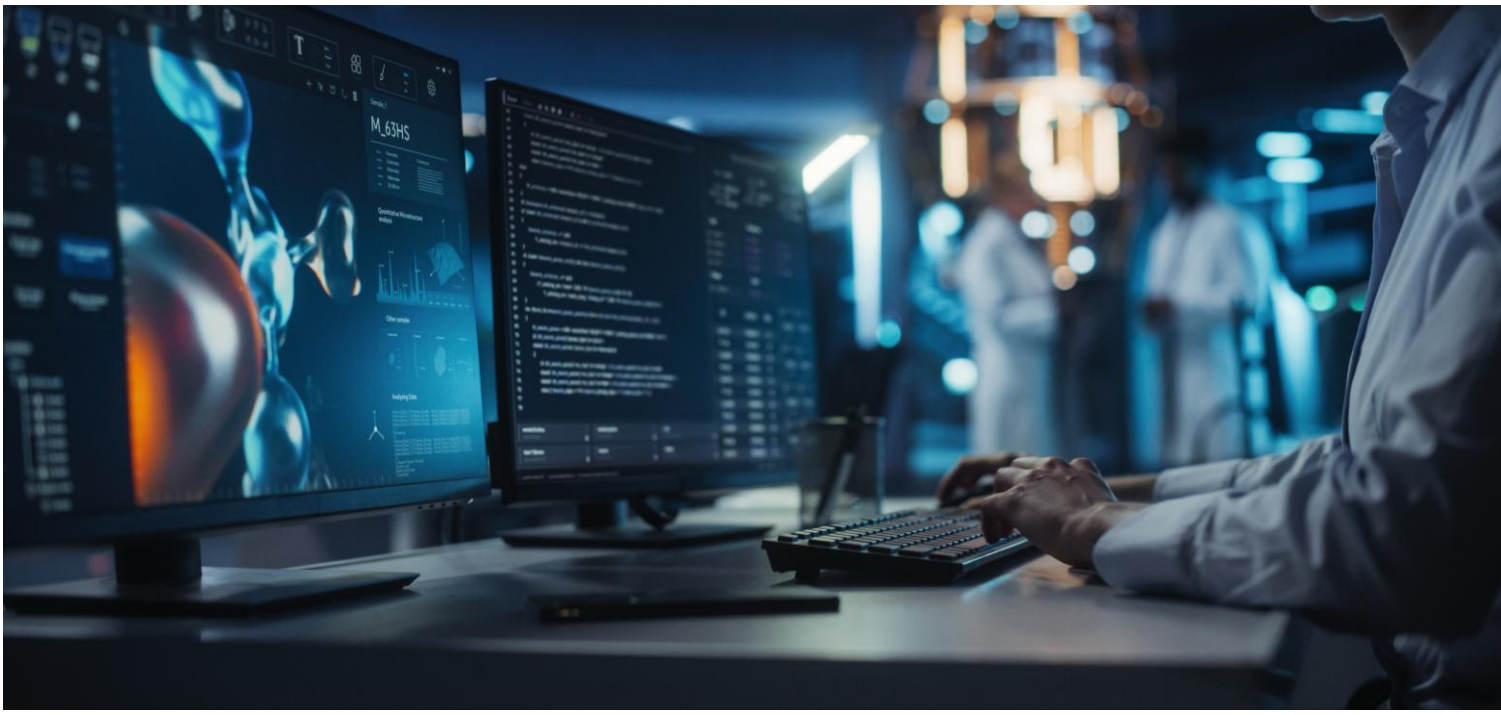
Figure 3

AI-generated deepfake video call fraud

The fraud was uncovered only after the funds had been moved, when the employee followed up with headquarters to confirm the instructions.

Similar tactics appeared in a March 2025 case disclosed by Singapore Police, in which a finance director at a multinational company joined a Zoom call featuring deepfake impersonations of senior executives, signed a non-disclosure agreement and transferred more than US$499,000 before the funds were traced and withheld.[28] Other attempted cases have followed the same pattern, using AI-generated voice or video to impersonate senior leaders and induce payments or disclose sensitive information.

These incidents suggest that deepfake-enabled fraud is becoming a repeatable form of executive impersonation. They show how trusted communication channels can be manipulated to create urgency, suppress challenge and exploit approval workflows.

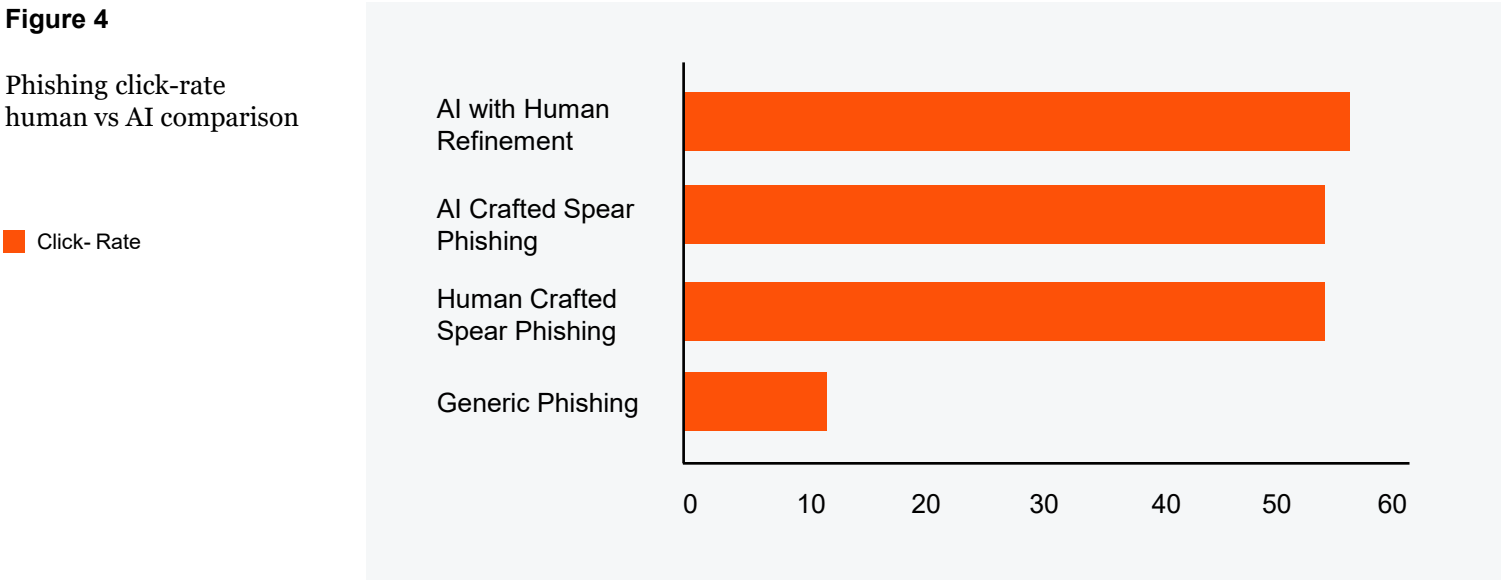


4.2 Evidence B: Evidence from human-subject phishing studies

While these incidents demonstrate AI's ability to deceive in real time, controlled studies show that similar risks also exist in text-based channels such as phishing. A 2024 human-subject study exposed 101 participants to four categories of phishing emails: generic templates, human-crafted spear phishing, fully AI-automated messages and AI content refined with limited human input (see Figure 4).[29]

Figure 4

Phishing click-rate human vs AI comparison



The results show that AI-generated phishing can match the persuasiveness of skilled human attackers while scaling far beyond human capacity, making phishing over three times more effective than generic attempts. The lesson for defenders is that traditional training around obvious errors in phishing messages is no longer sufficient and organisations must adapt by incorporating behavioural email analysis and awareness programmes that account for sophisticated, AI-driven lures.

4.3 Evidence C: The emergence of malicious generative AI tools

Beyond incidents and research, the rise of malicious GenAI models highlights how the cybercrime ecosystem itself is evolving. FraudGPT and its predecessor WormGPT are LLMs marketed on dark-web forums and Telegram channels, stripped of the ethical safeguards embedded in mainstream AI platforms (Trustwave). These tools are advertised as capable of generating phishing campaigns and offensive content, though many claims about exploit generation remain exaggerated.[30]

These models illustrate how quickly the cybercrime market adopts technologies that reduce cost, effort or skill requirements but developments are not occurring in isolation. Across the finance sector alone, nearly 500 reported cyber incidents were recorded over an 18-month period, with credit institutions accounting for 46% of cases.[31] Social-engineering-led attacks consistently ranked among the most damaging, highlighting how AI-enabled deception is now a dominant risk vector rather than an edge case.

Tool	Advertised on	Community size	Subscription pricing
WormGPT	Telegram	~5,000 subscribers within days	€60-100 per month[32]
FraudGPT	Underground forums	Not publicly disclosed	From US$200 per month[33]

Although no public breach has yet been definitively attributed to FraudGPT or WormGPT, their subscription-based business models demonstrate the commoditisation of offensive AI. Even if exaggerated, their rapid uptake points to a wider distribution of offensive capability. Advanced attack capabilities, once the preserve of highly skilled adversaries, are now accessible to less experienced actors. Adversaries therefore no longer require the same depth of technical expertise, making proactive resilience and AI-specific controls a business imperative.[34]

Taken together, these cases highlight common patterns. AI enhances deception by making deepfakes and phishing campaigns more persuasive, as seen in the Hong Kong incident and the controlled phishing study. Metrics confirm the scale of the risk, ranging from a US$25m financial loss to phishing campaigns that are more than three times as effective as generic ones.

Defences must evolve accordingly, with organisations adopting out-of-band verification, anomaly detection and workforce training. Most importantly, the rise of tools like FraudGPT suggests that offensive AI is shifting from isolated cases toward systematic and scalable campaigns capable of industrialising cybercrime.

Aspect	AI-powered attacks	Traditional attacks
Effectiveness	Highly convincing (deepfakes, AI phishing ~54), large-scale fraud[35]	Less convincing (generic phishing ~12-20% click-through), smaller losses36
Detection	Often detected after damage, bypasses filters and human suspicion	Easier to spot, spam filters and training effective against poor-quality scams
Recovery	Often slow or impossible (Funds unrecovered, AI ransomware caused weeks of downtime)	Sometimes recoverable (insurance payouts and backups for ransomware)
Barrier to entry	Low. Tools like FraudGPT and WormGPT enable unskilled attackers to create malware/phishing	Higher. Requires technical skills or strong social engineering
Lesson	Out-of-band verification, AI-aware phishing training, proactive detection	Standard awareness training, spam filtering, incident response

Modern cyber-attacks increasingly span identities, networks and applications simultaneously, eroding the effectiveness of perimeter-based and signature-driven controls. As AI accelerates attack speed and adaptability, defenders face a shrinking window to detect, verify and contain incidents before material damage occurs.

Despite this shift, organisational detection models and response playbooks often remain calibrated to a pre-AI threat model. Incident data shows that social engineering remains a dominant entry point, yet formal governance for AI use, AI-enabled incidents and deepfake-driven fraud is still immature across many sectors. Without changes to detection models, testing approaches and response playbooks, organisations risk facing AI-enabled attacks with controls designed for a different threat era.

05 Recommendations and next steps

From reactive defence to proactive resilience

GenAI is changing how cyber-attacks are created, scaled and delivered. Speed, scale and authenticity are no longer meaningful constraints for adversaries and traditional security controls are proving insufficient. To remain resilient, organisations must move beyond incremental controls and redesign how they detect, test and respond to modern threats. Organisations should prioritise five actions:

01 Focus detection on behaviour, not indicators

Invest in detection capabilities that focus on anomalous behaviour, identity misuse and lateral movement rather than static signatures, recognising that AI-generated attacks are designed to evade traditional controls

02 Institutionalise continuous adversary simulation

Move from periodic penetration testing to regular, automated adversary emulation that reflects real attacker behaviour, enabling security teams to identify gaps in detection and response before incidents occur

03 Embed out-of-band verification for high-risk actions

Introduce mandatory verification controls for sensitive transactions and access changes, particularly where AI-generated voice, video or messaging could be used to impersonate trusted individuals

04 Strengthen governance for AI use across the organisation

Define clear policies for AI access, data sharing and logging, aligned to recognised frameworks such as NIST and ensure accountability for AI-enabled decisions across security, IT and the business

05 Tighten red-blue collaboration through purple teaming

Use purple teaming to ensure lessons from offensive testing translate immediately into defensive improvements, reducing response times and improving organisational readiness

Together, these actions move organisations beyond reactive defence towards proactive resilience. As AI drives cybercrime at scale, continuous verification, detection and testing will define who keeps up – and who falls behind.

Contact Us

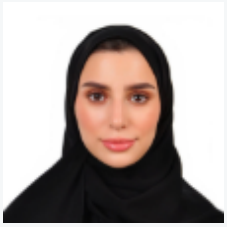
Haitham Al-Jowhari
Partner, Cybersecurity,
PwC Middle East
E: haitham.al-jowhari@pwc.com
LinkedIn

Contributors

Mohammed Ayesh
Director, Cybersecurity,
PwC Middle East
E: mohammed.ayesh@pwc.com
LinkedIn

Waad Albayyali
Senior Manager, Cybersecurity,
PwC Middle East
E: Waad.albayyali@pwc.com
LinkedIn

Nayef Alaqeel
Senior Consultant, Cybersecurity,
PwC Middle East
E: Nayef.alaqeel@pwc.com
LinkedIn



Raghad Al Sagga
Senior Consultant, Cybersecurity,
PwC Middle East
E: raghad.al.sagga@pwc.com
LinkedIn

Noura Alluhaidan
Consultant, Cybersecurity,
PwC Middle East
E: noura.alluhaidan@pwc.com
LinkedIn

Wafi Alabdulkarim
Consulting Intern,
Cybersecurity, PwC Middle East
E: wafi.alabdulkarim@pwc.com
LinkedIn

References

1. WHOIS is a protocol used to query databases for information about domain names, IP addresses and network ownership
2. Telnet is an early remote access protocol that allows command-line interaction with networked systems, often used historically for testing connectivity.
3. Nmap is a widely used network scanning tool that maps systems, identifies open ports and detects running services.
4. Assembly language and C are older, low-level programming languages that allow precise control over computer systems, but usually require more specialist skill and time to use than newer tools.
5. https://www.pwc.com/gx/en/issues/cybersecurity/cyber-threat-intelligence/annual-threat-dynamics.html
6. UAE – National Cybersecurity Strategy https://u.ae/en/about-the-uae/strategies-initiatives-and-awards/strategies-plans-and-visions/strategies-plans-and-visions-untill-2021/national-cybersecurity-strategy-2019
7. Qatar – Digital Agenda 2030 https://www.mcit.gov.qa/en/digital-agenda-2030/
8. Qatar – Qatar Digital Government NextGen Strategy https://services.hukoomi.gov.qa/assets/documents/digitalprojects/QDG%20NextGen%20Strategy.pdf
9. https://www.strongestlayer.com/blog/ai-generated-phishing-enterprise-threat
10. https://www.enisa.europa.eu/sites/default/files/2025 -02/Finance%20TL%202024_Final.pdf
11. https://layerxsecurity.com/generative-ai/fuzzing/
12. https://layerxsecurity.com/generative-ai/fuzzing/
13. https://www.paloaltonetworks.com/blog/2024/05/ai-generated-malware/
14. https://www.paloaltonetworks.com/blog/2024/05/ai-generated-malware/
15. https://www.sentinelone.com/cybersecurity-101/threat-intelligence/what-is-polymorphic-malware/
16. https://cybelangel.com/blog/data-model-poisoning/
17. NIST Special Publication 800-115, Technical Guide to Information Security Testing and Assessment, outlines a phased approach covering planning, execution, analysis and post-assessment activities.
18. The National Institute of Standards and Technology (NIST), a US government standards body, published SP 800-115 as a guide to planning, conducting and assessing information security testing. https://csrc.nist.gov/pubs/sp/800/115/final
19. PentestGPT – an LLM-based penetration testing framework designed to guide and automate security testing workflows. Available at: https://github.com/GreyDGL/PentestGPT
20. PentestAgent – a large language model–based automated penetration testing framework that uses multi-agent design to perform reconnaissance, analysis and exploitation tasks with reduced human intervention. https://arxiv.org/abs/2411.05185
21. MITRE's, CALDERA: Automated Adversary Emulation Platform, describes CALDERA as a system that "emulates adversary behaviour within networks" using the MITRE ATT&CK framework. Available at: https://caldera.mitre.org/
22. MITRE ATT&CK (Adversarial Tactics, Techniques and Common Knowledge) is a publicly available framework that catalogues known cyber adversary behaviours, including tactics and techniques observed in real-world attacks. https://attack.mitre.org/
23. APT3 (Advanced Persistent Threat 3) is a cyber espionage group documented in the MITRE ATT&CK framework and commonly used to model real-world adversary behaviour. See: https://attack.mitre.org/groups/G0022/
24. https://www.ibm.com/think/x-force/ai-vs-human-deceit-unravelling-new-age-phishing-tactics
25. The NIST AI Risk Management Framework is a voluntary NIST framework for identifying and managing AI risks and promoting trustworthy AI. https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-ai-rmf-10
26. NIST's Generative AI Profile applies the AI Risk Management Framework to generative AI and helps organisations identify and manage its distinct risks. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf
27. https://digitalcommons.unomaha.edu/ncitereportsresearch/136/
28. https://www.channelnewsasia.com/singapore/deepfake-scam-impersonate-ceo-company-finance-director-5048706
29. https://arxiv.org/abs/2412.00586
30. https://blog.talosintelligence.com/cybercriminal-abuse-of-large-language-models/
31. https://www.enisa.europa.eu/sites/default/files/2025-02/Finance%20TL%202024_Final.pdf
32. https://www.theregister.com/2025/11/25/wormgpt_4_evil_ai_lifetime_cost_220_dollars
33. https://www.infosecurity-magazine.com/news/dark-web-markets-fraudgpt-ai-tool/
34. https://www.rapid7.com/blog/post/ai-goes-on-offense-how-llms-are-redefining-the-cybercrime-landscape/
35. https://arxiv.org/abs/2412.00586
36. https://arxiv.org/abs/2412.00586

About PwC

At PwC, we help clients build trust and reinvent so they can turn complexity into competitive advantage. We're a tech-forward, people-empowered network with more than 364,000 people in 136 countries and 137 territories. Across audit and assurance, tax and legal, deals and consulting, we help clients build, accelerate, and sustain momentum. Find out more at www.pwc.com.

With over 11,000 people across 12 countries in 30 offices, PwC Middle East combines deep regional insight with global expertise to help clients solve complex problems, drive transformation, and achieve sustained outcomes. Learn more at www.pwc.com/me.

PwC refers to the PwC network and/or one or more of its member firms, each of which is a separate legal entity. Please see www.pwc.com/structure for further details.