FUJITSU

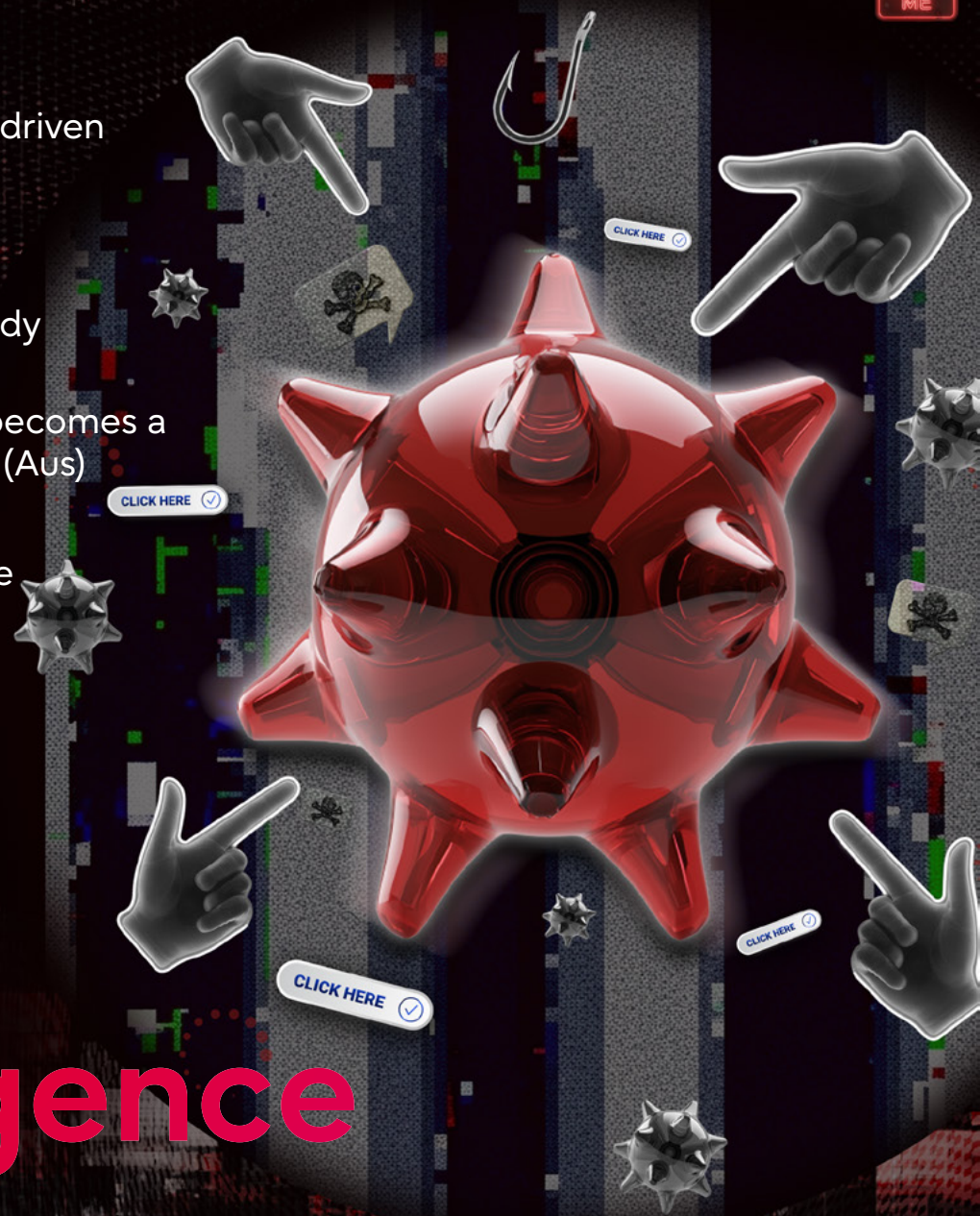
June 2026

ClickFix: Evolution of user-driven malware execution

Quantum is coming: Your encryption may not be ready

June 2026 ISM update: AI becomes a governed security domain (Aus)

Months, not years: Why the Five Eyes cyber warning demands action now



Threat Intelligence Report

A monthly digest of cyber threat activities, insights, and strategies for enhanced cyber resilience

Fujitsu Cyber

Contents

This threat intelligence report has been developed using the insights from the various teams within Fujitsu Cyber. We report on the overarching trends we have recognised in the past few months, with a focus on current events and actionable steps.

At Fujitsu Cyber, we actively take these insights from what we observed and apply them to all the work we do, whether it be with our consulting engagements, our ongoing threat hunting programme, or our managed service client environments. Our constant learning across the business helps us to stay adaptable and on top of our security game, so that we can keep our client systems as safe as possible.

Article one:

ClickFix:

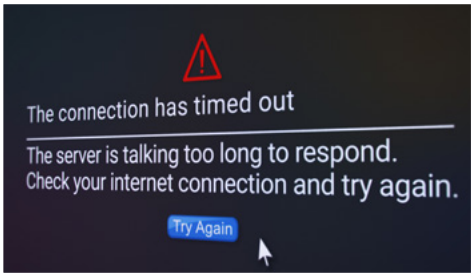
Evolution of user-driven malware execution

This article was written by:
Connor Owens
Senior SOC Analyst
Jack Bartlett
Junior SOC Analyst

ClickFix represents a growing evolution in social engineering malware delivery, first observed in October 2023, moving the focus from traditional malicious attachments and drive-by downloads towards user-driven execution.

Rather than relying on technical exploitation alone, ClickFix campaigns manipulate users into carrying out the attackers' objectives themselves, often by presenting convincing fake security prompts, browser errors, software update messages or CAPTCHA verification steps that instruct the user to copy, paste or execute malicious commands.

The technique has rapidly gained traction across multiple threat actor groups due to the effectiveness against security controls, by abusing legitimate system tools such as Run, Command Prompt and PowerShell. Attackers can bypass traditional email filtering, attachment scanning and some automated detection mechanisms. The result is a highly effective initial access method that relies less on sophisticated malware and more on human trust and urgency. According to Microsoft's 2025 Digital Defense Report, ClickFix was representing 47% of initial access methods, showing how much this method is growing and evolving [1]. This article will cover the changes, the dangers and how to mitigate them.



47%

ClickFix represents 47% of initial access methods [1]

What is ClickFix?

Historically, many phishing and malware campaigns relied on users opening weaponised attachments, enabling macros, or downloading suspicious executables. ClickFix represents a notable shift in attacker methodology by removing the obvious malicious artefacts from the initial stage of compromise.

ClickFix campaigns are increasingly using:

- **Fake browser verification checks**
- **Fraudulent Microsoft, Google, Apple, or IT support prompts**
- **Malicious CAPTCHA verification workflows**
- **Fake "document failed to load" or "plugin missing" messages**
- **Social engineering delivered through compromised websites, malicious adverts, phishing emails and cloud-hosted content**

Recent campaigns have shown to be significantly improving realism, with attackers impersonating legitimate brands and business workflows, as well as compromising local business sites, convincingly enough to deceive technically capable users. In some cases, malicious instructions are hosted on legitimate platforms such as Medium, GitHub, or cloud storage providers. This reduces the likelihood of immediate domain or IOC-based blocks, as these sites are not inherently malicious but can have malicious subdomain sites.

This matters because the technique deliberately shifts execution responsibility onto the user, making many traditional preventative controls less effective. Security tools may still detect post-execution activity, but by that point the initial compromise has already occurred. Organisations should view ClickFix not simply as phishing, but as an adaptive execution technique capable of bypassing security assumptions based solely on malware file detection.

Why is ClickFix so effective?

ClickFix is highly effective because it shifts the responsibility for execution away from the attacker and onto the user, allowing it to bypass many traditional security controls. Rather than delivering a malicious file or exploit directly, the attacker relies on convincing the user to run commands themselves using trusted system tools.

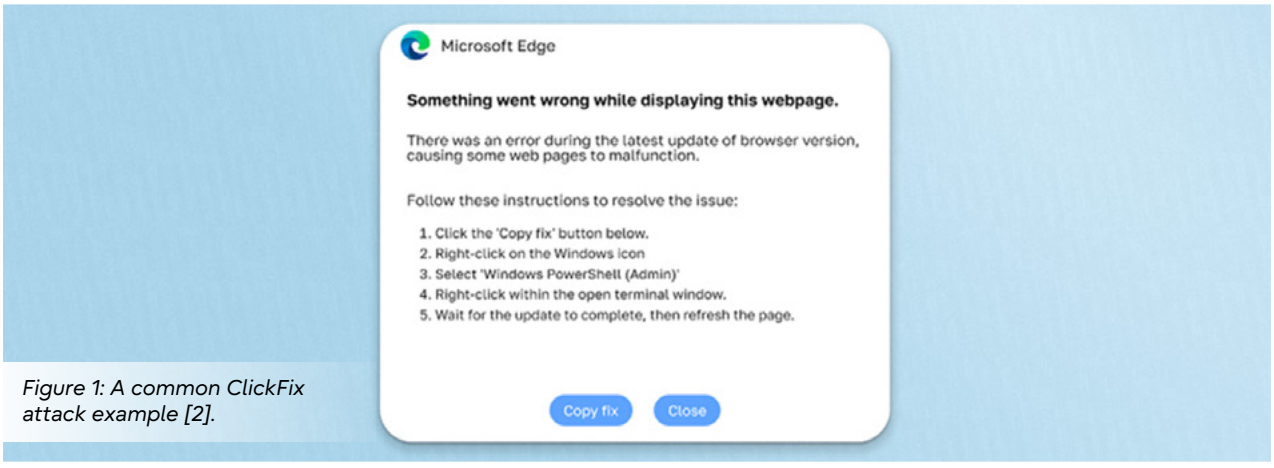


Figure 1: A common ClickFix attack example [2].

This approach enables ClickFix to bypass email security controls, as there may be no malicious attachment or clearly identifiable payload to detect at the point of delivery. Similarly, attachment sandboxing becomes ineffective, as there is no file to detonate or analyse in isolation. Traditional macro-based protections are also avoided entirely, as these campaigns do not rely on Office documents or embedded scripts.

Another key factor in its success is how legitimate the activity appears. Attackers replicate well-known brands, common workflows, and familiar system prompts, making the interaction feel routine rather than suspicious. By using trusted tools such as PowerShell, Command Prompt, and the Run dialog, the resulting activity blends in with normal user behaviour, reducing the likelihood of detection.

ClickFix is also highly adaptable and works across multiple platforms, including Windows, macOS, and browser-based environments. This flexibility allows attackers to reuse the same social engineering techniques across different targets with minimal modification.

Ultimately, ClickFix is effective because it exploits human trust and behaviour rather than technical vulnerabilities. By requiring user interaction to trigger execution, the technique can bypass many conventional and automated security controls that are designed to detect or block malicious files and activity [3].

ClickFix variants seen in the wild

ClickFix campaigns continue to evolve rapidly, with threat actors adapting both the delivery method and the social engineering pretext to better align with real-world user behaviour. While the underlying objective remains consistent - convincing users to execute malicious commands - the way this is presented can vary significantly depending on the target audience and environment.

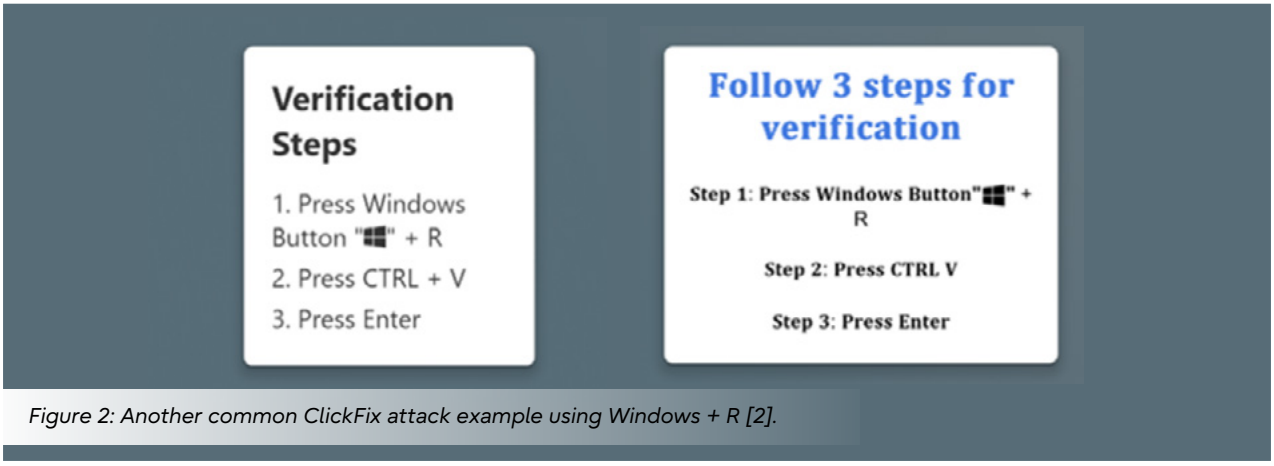


Figure 2: Another common ClickFix attack example using Windows + R [2].

One of the most common approaches is through browser and web-based lures. These often appear as fake verification checks, browser errors, or CAPTCHA challenges that claim the user must complete an additional step to proceed. In many cases, users are instructed to copy and paste a command into the Run dialog or PowerShell under the illusion of "verifying" they are not a bot or resolving a loading issue. Because these prompts are presented within a browser context, they feel routine and are less likely to raise suspicion.

Another frequently observed category is productivity and business workflow lures. These include fake document access errors, secure file sharing prompts, or meeting-related issues such as "fixes" for Teams or Zoom calls. These scenarios are particularly effective in corporate environments, where users are accustomed to interacting with shared documents and collaboration tools. By mimicking common workplace friction points, attackers take advantage of urgency and familiarity to drive user action.

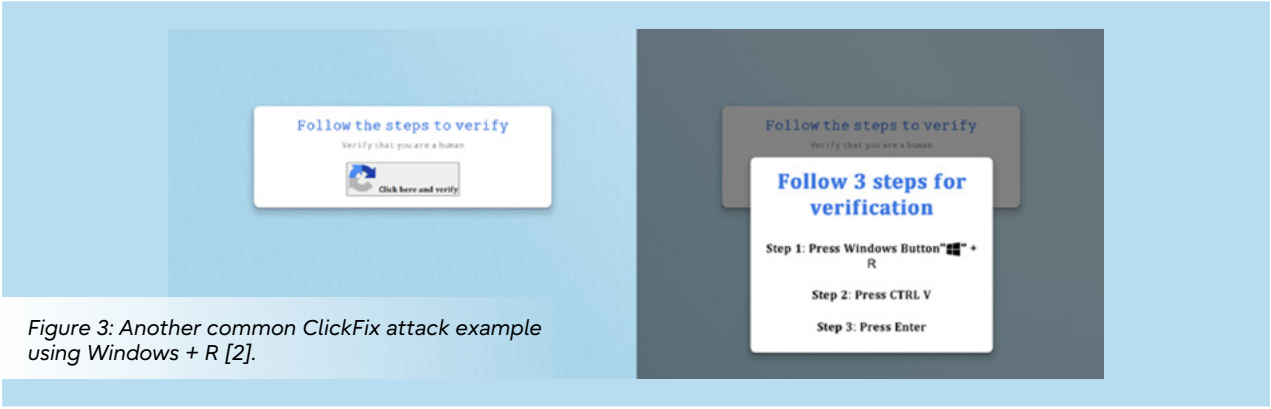


Figure 3: Another common ClickFix attack example using Windows + R [2].

ClickFix has also been seen leveraging platform and operating system impersonation, where prompts imitate legitimate system notifications. Examples include fake macOS security warnings, software verification requests, or generic system alerts claiming that an action is required to continue. These are designed to appear as trusted, system-generated messages, making users more likely to follow the instructions without questioning their validity.

Beyond the lures themselves, the delivery mechanisms play a key role in the success of these campaigns. Attackers increasingly rely on SEO poisoning to surface malicious pages in search engine results, compromised legitimate websites to host attack content, and social media platforms to distribute links or instructions. In some cases, the malicious instructions themselves are hosted on otherwise trusted platforms such as GitHub or cloud storage providers, further reducing suspicion and making traditional blocklisting less effective.

These variations highlight that ClickFix is not a single, static technique, but a flexible and adaptive approach to social engineering. Its strength lies in how easily it can be reshaped to fit different environments, making it difficult to defend against using simple indicators or traditional detection methods alone.

Impact of ClickFix

ClickFix is often framed as an initial access technique, but its true impact lies in what follows once a user executes the attacker's instructions. It serves as the entry point into a broader compromise, enabling threat actors to move quickly from a single user interaction to persistent access, credential theft, and wider organisational risk. Microsoft threat research highlights that ClickFix campaigns are actively used to deliver payloads at scale, often resulting in information theft and data exfiltration across both Windows and macOS environments [3].

A common first-stage outcome is the deployment of infostealer malware, which is specifically designed to harvest sensitive data from compromised systems. These payloads focus on extracting browser-stored credentials, saved passwords, authentication tokens, cryptocurrency wallets, and other high-value information. Microsoft research shows that modern infostealers associated with ClickFix-style lures are capable of rapidly collecting large volumes of data while blending into legitimate system activity, often completing their objective before detection mechanisms are triggered [4].

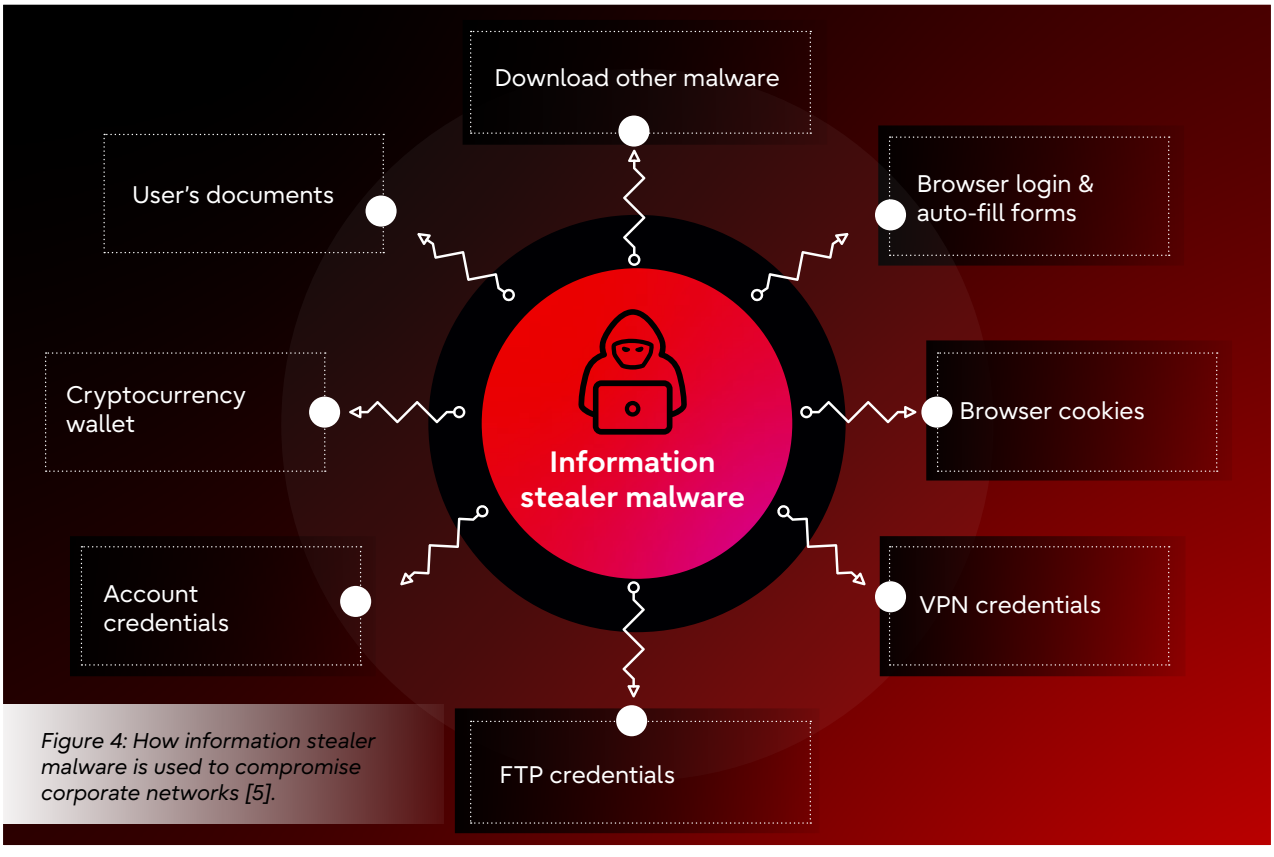


Figure 4: How information stealer malware is used to compromise corporate networks [5].

In addition to credential theft, attackers increasingly target session data and authentication tokens, which can provide immediate access to user accounts without requiring credentials or re-authentication. This significantly increases the risk of account takeover, as access can appear legitimate within logs and monitoring systems. Recent reports on identity threat data highlight the scale of this issue, including widespread exposure of session cookies that enable attackers to bypass multi-factor authentication protections entirely [6].

Once initial access is achieved, threat actors commonly move to establish persistence and deeper control within the environment. Rather than relying solely on traditional malware techniques, this may include the deployment of remote access tools or the abuse of legitimate authentication mechanisms. More recent ClickFix developments have also expanded into identity-driven techniques such as OAuth consent abuse, where attackers gain persistent access through authorised applications rather than stolen credentials. Research into emerging ClickFix variants demonstrates that this approach allows attackers to maintain access even after password resets or other defensive actions [7].

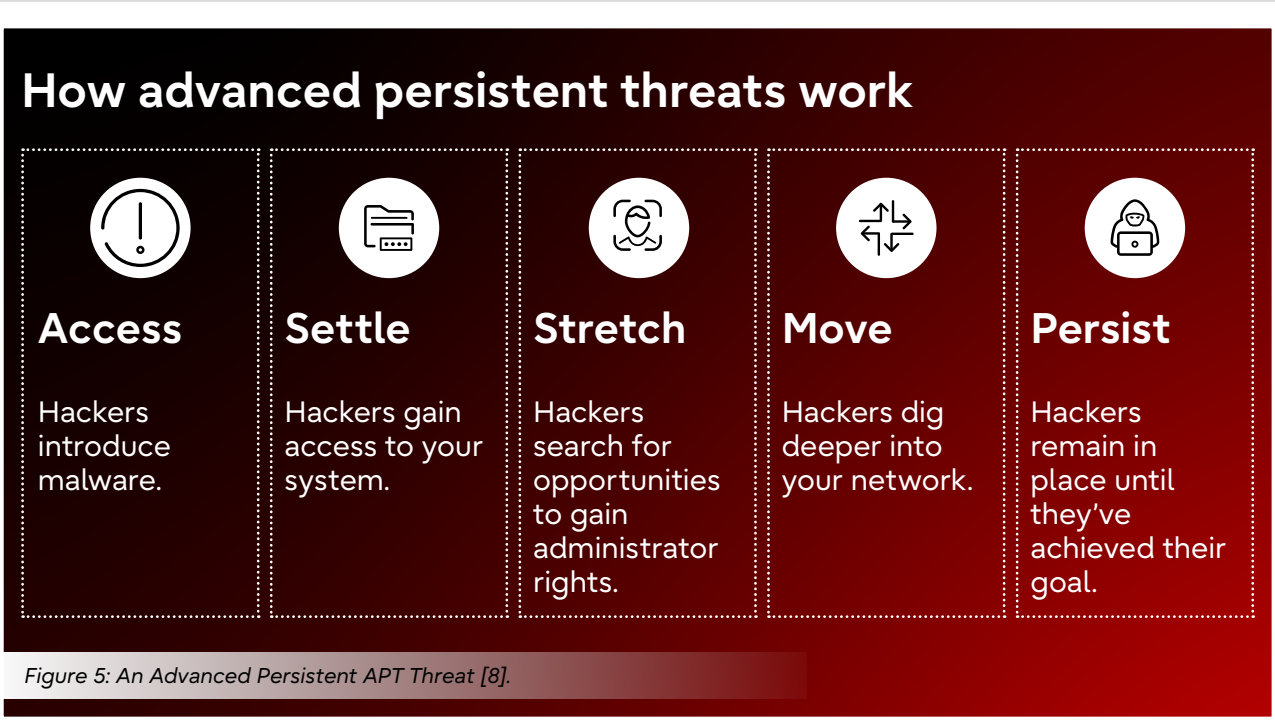


Figure 5: An Advanced Persistent APT Threat [8].

As the intrusion progresses, the focus typically shifts toward data access, lateral movement, and operational impact. With valid credentials or persistent access in place, attackers can move across internal systems, identify high-value assets, and access sensitive data without immediately triggering alerts. Real-world incident response cases reinforce this progression. CERT Polska analysis demonstrates how a single ClickFix-style fake CAPTCHA execution can initiate a full attack chain, leading to malware deployment, internal reconnaissance, and sustained attacker presence within the network [9].

From a business perspective, the downstream effects can be significant. Compromised credentials and persistent access enable data exfiltration, financial fraud, and further stages of attack such as ransomware deployment. In many cases, ClickFix is not the objective but the enabler, providing the foothold required for more impactful activity to take place. This shift in attacker strategy, from delivering malware to enabling user-driven execution, means that by the time suspicious behaviour is detected, the attacker may already be operating with a high degree of legitimacy within the environment.

Ultimately, ClickFix should be viewed not simply as a social engineering technique, but as a highly effective gateway into modern attack chains. Its ability to bypass preventative controls and accelerate progression from initial access to full compromise makes it a particularly high-risk threat, especially in environments where user-driven execution of commands is not tightly controlled.

Remediation

Organisations should adopt a layered defence strategy that combines user awareness, preventative controls, and effective monitoring. As ClickFix relies on users executing attacker-provided commands, reducing opportunities for user-driven execution is critical.

User awareness

Users should be educated that legitimate websites, software vendors, and internal IT teams will not ask them to manually execute commands through PowerShell, Command Prompt, Terminal, or the Run dialog.

Security awareness training should include examples of fake CAPTCHA verification pages, browser verification scams, fraudulent software update prompts, and impersonation of trusted brands or services.

Restrict script execution

Where operationally feasible, organisations should restrict PowerShell usage to approved administrators, enforce PowerShell Constrained Language Mode (CLM), implement AppLocker or Windows Defender Application Control (WDAC), and apply the principle of least privilege.

These controls reduce the ability of attackers to abuse legitimate system tools following successful social engineering.

Network and identity controls

Organisations should implement URL filtering, DNS protection, and rapid IOC blocklisting to reduce exposure to malicious infrastructure. Newly observed malicious domains, URLs, and payload locations should be blocked quickly across the environment where possible.

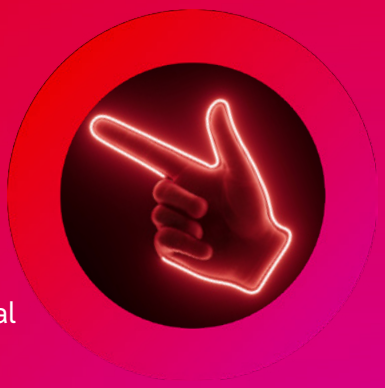
Identity controls should also be reviewed, including multi-factor authentication, anomalous sign-in monitoring, OAuth consent grant reviews, and minimising standing privileged access.

Detection and monitoring

Security teams should monitor for suspicious use of:
powershell.exe | cmd.exe | mshta.exe | wscript.exe | cscript.exe | rundll32.exe

Additional monitoring should focus on encoded PowerShell commands, PowerShell downloading remote content, usage of tools such as curl, bitsadmin, certutil, wget and Invoke-WebRequest, and browser processes spawning scripting engines. As ClickFix often bypasses preventative controls, behavioural detection remains a critical defensive layer.

Conclusion



ClickFix represents a significant evolution in social engineering, shifting malware execution responsibility from the attacker to the user. By abusing trusted tools and familiar workflows, attackers can bypass many traditional security controls and gain initial access with minimal technical exploitation.

As the technique continues to evolve, organisations should combine user awareness, preventative controls, identity protection, and behavioral detection to reduce the likelihood and impact of compromise. Defending against ClickFix is no longer just a technical challenge; it requires users to be recognised as a critical part of the security perimeter.

References (including image references)

[1] Microsoft, "Microsoft digital defense report 2025," 2025. [Online]. Available: https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/msc/documents/presentations/CSR/Microsoft-Digital-Defense-Report-2025.pdf.

[2] Keep Aware, "ClickFix: The what, why, where and how of it all," 2025. [Online]. Available: https://keepaware.com/blog/clickfix-the-what-why-where-and-how-of-it-all.

[3] Microsoft, "Think before you Click(Fix): Analyzing the ClickFix social engineering technique," 2025. [Online]. Available: https://www.microsoft.com/en-us/security/blog/2025/08/21/think-before-you-clickfix-analyzing-the-clickfix-social-engineering-technique/.

[4] Microsoft, "ClickFix variant 'CrashFix' deploying Python RAT trojan," 2026. [Online]. Available: https://www.microsoft.com/en-us/security/blog/2026/02/05/clickfix-variant-crashfix-deploying-python-rat-trojan/.

[5] Australian Cyber Security Centre (ACSC), "The silent heist: Cybercriminals use information stealer malware to compromise corporate networks," 2024. [Online]. Available: https://www.cyber.gov.au/about-us/view-all-content/alerts-and-advisories/silent-heist-cybercriminals-use-information-stealer-malware-compromise-corporate-networks.

[6] Recorded Future, "2025 identity threat landscape report: Inside the infostealer economy," 2026. [Online]. Available: https://www.recordedfuture.com/blog/identity-trend-report-march-blog.

[7] Revel8, "ClickFix attacks in 2026: 7 variants, real attack data & defence guide," 2026. [Online]. Available: https://www.revel8.ai/blog/threat-research-clickfix-attacks-2026.

[8] Okta, "Advanced persistent threat: Definition, lifecycle and defense," 2024. [Online]. Available: https://www.okta.com/en-in/identity-101/advanced-persistent-threat-apt/.

[9] CERT Polska, "ClickFix in action: How fake CAPTCHA can lead to a company-wide infection," 2026. [Online]. Available: https://cert.pl/en/posts/2026/02/fake-captcha-in-action/.

[10] Huntress, "ClickFix attack: Variants, detection & how it works," 2025. [Online]. Available: https://www.huntress.com/blog/dont-sweat-clickfix-techniques.

[11] Microsoft, "Infostealers without borders: macOS, Python stealers and platform abuse," 2026. [Online]. Available: https://www.microsoft.com/en-us/security/blog/2026/02/02/infostealers-without-borders-macos-python-stealers-and-platform-abuse/.

[12] Todyl, "ClickFix: The evolution of copy-paste social engineering," 2026. [Online]. Available: https://www.todyl.com/blog/clickfix-evolution-copy-paste-social-engineering

[13] HKCERT, "Phishing alert: ClickFix tactics evolve – now attacking both Windows and macOS," 2026. [Online]. Available: https://www.hkcert.org/security-bulletin/phishing-alert-clickfix-tactics-evolve-now-attacking-both-windows-and-macos_20260331.

Article two:

Quantum is coming:

Your encryption may not be ready

This article was written by:
Haley Southgate
Senior Consultant

Emerging quantum capabilities present a long-term confidentiality risk for sensitive information collected today under “harvest now, decrypt later” (HNDL) activity.

In this model, adversaries collect encrypted data now with the intention of decrypting it later once quantum capabilities mature.

For organisations in Australia and New Zealand, the priority is to identify high-value data such as critical infrastructure data, intellectual property or health information, understand cryptographic dependencies, and begin transition planning before migration timeframes become compressed and high risk.

Quantum readiness should be treated as a governance and data protection issue, not just a future technical upgrade.

Threat overview

The shift from immediate to delayed exploitation

The adversary approach is shifting from immediate misuse to delayed exploitation, where nation-state actors, intelligence services, and advanced persistent threat (APT) groups collect and retain encrypted data for later exploitation. The risk increases when sensitive data outlasts the cryptographic controls protecting it.

Quantum computing is expected to undermine widely used public-key cryptographic systems such as RSA and elliptic-curve cryptography, which protect communications and data. While secure today, sufficiently advanced quantum systems could break them in the near future, exposing previously protected data. [1]

There is no single agreed "Q-day" at which quantum computers will suddenly break widely used public-key cryptographic systems. However, recent expert analysis indicates that the expected timeline for cryptographically relevant quantum capability has tightened. The Global Risk Institute's 2025 Quantum Threat Timeline Report estimates a 28-49% probability of such capability emerging within the next 10 years, and a 51-70% probability within 15 years, reflecting a material risk horizon for organisations relying on classical cryptography. [9] Organisations should act before this capability exists, as cryptographic migration often takes years, and delay forces a compressed, high-risk transition.

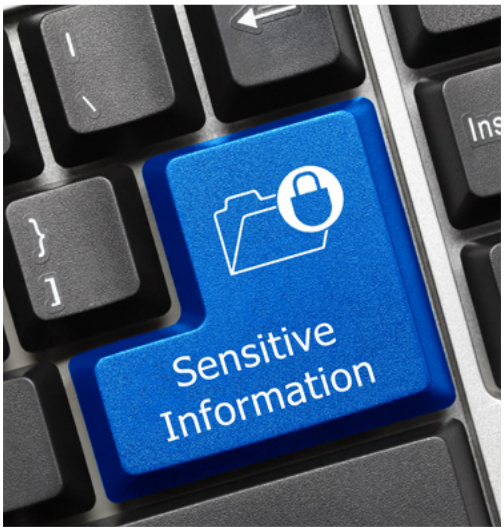
In this model, compromise occurs at the point of collection, even if the data cannot be decrypted until later.

Consider data encrypted in 2019 using RSA-2048: it could be intercepted and stored today, remain unreadable at the time of collection, and become exposed later once advances in cryptanalysis and quantum computing enable decryption. While not yet observed at scale, this risk is widely recognised in cyber security guidance. [4]

Adversary activity "Harvest now, decrypt later"

Typical adversary lifecycle under HNDL:

Collection	Storage	Future exploitation
Encrypted data is gathered via interception, endpoint compromise, or server access.	Data is retained long-term in adversary-controlled repositories.	Decrypted when quantum capabilities become viable.



The immediate risk is data collection today, especially where information has long confidentiality lifetimes such as government records, financial data, intellectual property, and defence-related information. [2]

For example, identity records and payment data may remain sensitive for years, while defence plans, government records, intellectual property, or strategic planning material may require confidentiality over decades.

The delayed risk is future exposure and exploitation, particularly by government and nation-state actors able to conduct long-term intelligence operations.

Mapping to MITRE ATT&CK

This activity aligns most directly with MITRE ATT&CK Exfiltration (TA0010), with the key distinction that exfiltrated data may not be immediately usable by the adversary. Relevant behaviours may include exfiltration over command-and-control channels (T1041), use of alternative encrypted protocols (T1048), and fragmented or low-volume transfers designed to evade detection. [5]

Importantly, this activity may not trigger immediate incident response, as the data remains encrypted at the time of theft.

Risk implications

For Australia and New Zealand, this risk is most relevant to sectors that hold long-lived sensitive information, including government, defence, financial services, healthcare, critical infrastructure, and data storage or processing providers. These environments are already subject to strict regulations around data protection, cryptography, information asset management, and risk-based security assurance through regional cyber security guidance and critical infrastructure obligations. [7][8]

1

Long-term data exposure

The problem is simple, sensitive data may need to remain confidential for longer than the encryption protecting it remains reliable. [2]

2

Invisible compromise

Data may be stolen without triggering alerts, limiting the organisation's ability to respond, report, or manage risk.

3

Strategic timing of exploitation

By the time the business impact is visible, the compromise may have already happened, allowing adversaries to time disclosure or use of decrypted data for strategic, financial, or operational effect.

4

Cryptographic obsolescence risk

NIST guidance indicates that current cryptographic standards will require replacement due to quantum threats, with migration recommended to begin now. [1]

Historical note: The Data Encryption Standard (DES), first standardised in the 1970s, illustrates why cryptographic resilience matters. It provided sufficient protection at the time but became inadequate as computing capability and cryptanalysis advanced. [6]

Why acting now matters

Even where some encryption remains resistant longer than public-key systems, exposure persists through vulnerable key exchange, stored encrypted data, and long migration timelines. Organisations that delay action risk entering a forced, time-compressed migration cryptographic transition can take years where critical systems must be upgraded rapidly, increasing cost, operational disruption, and security risk. [4]

Recommended actions for quantum readiness

Immediate risk reduction

Identify and prioritise high-risk data to reduce long-term confidentiality risk.

Priority action: Focus on data with long confidentiality requirements, intellectual property, regulated data, authentication systems, and identity/PKI/signing systems.

Conduct a cryptographic inventory to enable informed transition planning.

Priority action: Document where encryption is used, which algorithms are deployed, and dependencies on vulnerable public-key cryptography covering TLS, VPNs, PKI and certificate services, identity/authentication systems, and data-at-rest encryption.

Structural readiness

Implement cryptographic agility to reduce future upgrade cost and disruption.

Priority action: Design systems so algorithms can be replaced without full redesign and updated as standards evolve.

Future transition

Adopt quantum-resistant algorithms to maintain long-term data confidentiality.

Priority action: Begin phased adoption for key exchange, digital signatures, and long-term storage encryption where feasible.

Use hybrid cryptography during transition to reduce migration risk.

Priority action: Combine current and quantum-resistant methods to maintain interoperability and reduce single-point failure risk.

Supporting controls

Strengthen data minimisation and retention to reduce exposure of sensitive data.

Priority action: Limit sensitive data storage and enforce retention and deletion controls.

Monitor for exfiltration behaviour to improve early detection of data theft.

Priority action: Detect low-and-slow transfers, unusual outbound traffic, and data staging aligned to MITRE ATT&CK Exfiltration behaviours.

Governance

Maintain a quantum-readiness roadmap to ensure coordinated and cost-effective transition.

Priority action: Define migration timelines, risk priorities, and integration into existing Governance, Risk & Compliance (GRC) frameworks.

Conclusion

The key decision for organisations is whether to start preparing while migration can be planned and prioritised, or defer action until cryptographic change becomes compressed, disruptive, and harder to govern.

Delay increases the risk of:

- Permanent loss of confidentiality
- Strategic exploitation of sensitive data
- Being unable to retroactively secure already stolen information.

Transitioning to quantum-resistant security is not simply a technical upgrade. It is **a long-term risk mitigation strategy** that must begin now.

References

[1] National Institute of Standards and Technology (NIST), "What Is Post-Quantum Cryptography?", 2024. [nist.gov]
[2] Palo Alto Networks, "Harvest Now, Decrypt Later (HNDL)", 2026. [paloaltonetworks.com]
[3] NIST, "New Draft White Paper: PQC Migration", 2025. [nist.gov]
[4] CISA, NSA, NIST, "Quantum Readiness: Migration to Post-Quantum Cryptography", 2023. [media.defense.gov]
[5] MITRE, "Exfiltration (TA0010)", MITRE ATT&CK®, 2025. [attack.mitre.org]
[6] NIST, "The Cornerstone of Cybersecurity – Cryptographic Standards and a 50-Year Evolution," 2022. [nccoe.nist.gov]
[7] Australian Signals Directorate, "Guidelines for cryptography," Australian Government Information Security Manual, 2026. [Guidelines for cryptography | Cyber.gov.au]
[8] Cyber and Infrastructure Security Centre, "Security of Critical Infrastructure Act 2018," 2024. [Security of Critical Infrastructure Act 2018 (SOCI)]
[9] Global Risk Institute, "Quantum Threat Timeline Report 2025," 2026. [globalriskinstitute.org]

Article three:

June 2026 ISM update:

AI becomes a governed security domain

This article was written by:
Shreya Dhoopad
Consultant

The Australia Cyber Security Centre's latest ISM release is not a routine update but one that represents a deliberate and targeted evolution of the framework.

This evolution is driven by the need to bring emerging technologies, particularly Artificial Intelligence (AI) into clearer regulatory focus. Rather than treating AI as an adjacent concern, the update embeds itself across the landscape inturn expanding governance expectations, reinforcing accountability, and integrating AI-specific risks into multiple security domains.

In doing so, the ISM signals a broader integration of AI considerations into operational security expectations, guiding organisation to proactively address AI within their existing security, risk, and assurance models.

The new updates appear to respond to three converging pressures regarding accelerated adoption of Artificial Intelligence, the increased sophistication of modern threat actors and the need for organisations to embed more adaptive and resilient security controls.

The most significant change in June 2026 is the explicit integration of AI into the ISM control framework. For the first time the ISM includes dedicated AI controls as well as embedded requirements across:

- **Software development**
- **Security operations**
- **Testing and assurance processes**

ASD is clear in its signal – AI must be secure, controlled, monitored, and governed in the same way as any critical system.

Key AI controls that were introduced

1. Restrict AI data exposure

Recommended to disable direct access to external public data sources.

Why this matters:

This reduces the risk of data leakage through external model interactions which is a known concern with generative AI platforms.

2. Human oversight

Recommended to flag high-risk or defined actions for human approval before execution.

Why this matters:

This introduces human-in-the-loop governance hence ensuring AI decisions cannot operate unchecked in critical environments.

3. Establishing behaviour baselines

Organisations are recommended that baselines of expected behaviour for AI applications are established and monitor for deviations and anomalies.

Why this matters:

This aligns AI controls with traditional security monitoring, introducing concepts such as

- Model drift detection
- Output validation
- Performance monitoring

4. Securing AI data and interactions

New requirements recommend that prompts and outputs are securely deleted when AI sessions are removed.

Why this matters:

This directly addresses risks associated:

- Prompt leakage
- Sensitive data persistence
- Training data exposure

5. Event log monitoring, vulnerability assessments and penetration tests

The ISM now recommends use of suitable AI models for the detections of:

- Detect cyber security events
- Identity incidents
- Perform vulnerability assessments and penetration testing
- Enhance software security testing

Why this matters:

AI is positioned not just as a risk, but as a force multiplier for defence. It encourages organisations to adopt intelligent security capabilities.

6. AI in software development

The introduction of 'Software Development Fundamentals' clarifies the ISM that it applies to human, AI-assisted, AI-powered, and AI-driven software development. Additionally, "Developers" now include AI agents and tools.

Why this matters:

This fundamentally changes how organisations must approach secure development lifecycles (SDLC), code assurance and validations, and DevSecOps governance. AI-generated code must now meet the same assurance standards as human-developed code.

Key implications for organisations

AI introduces a new risk domain

Artificial intelligence is no longer an experimental capability. It is now a regulated component of the enterprise environment. Organisations are recommended to treat AI systems as formal assets which are subject to the same governance, assurance, and control expectations as traditional systems.

Failure to do so introduces a new class of risk including unintended data leaks, model manipulation, prompt injection attacks, and unreliable or unexplainable outputs like hallucinations. Exposure from this risk could result in material risk, including unintended data exposure, opaque or invalidated decision-making, and increasing compliance gaps across regulatory obligations. It is further compounded by supply chain dependencies within AI models hence creating additional exposure across third-party ecosystems.

Security operation must materially modernise

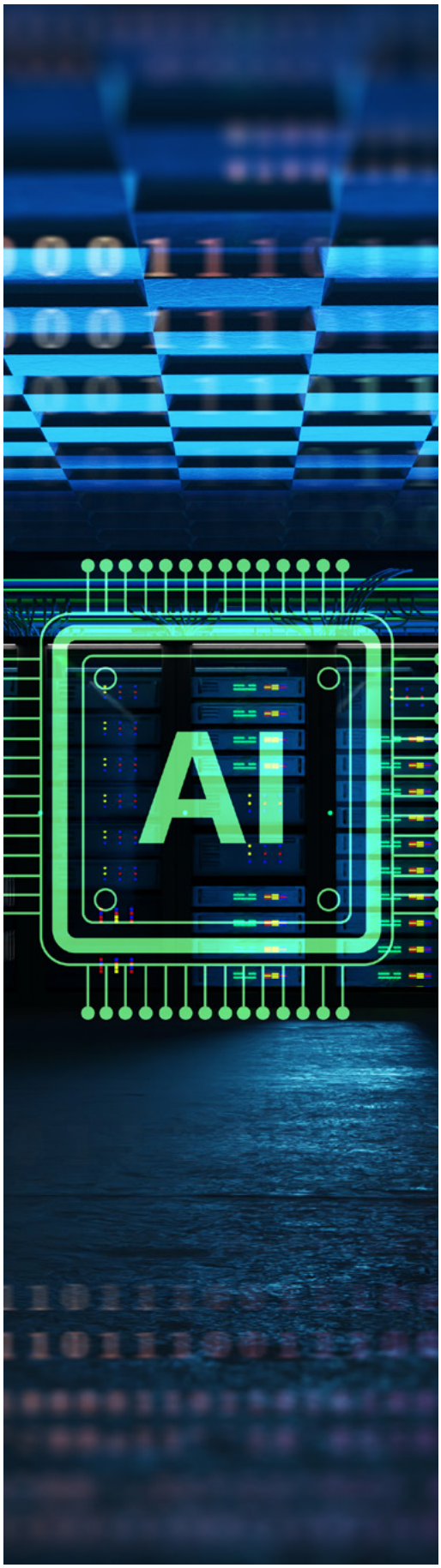
The clear shift in the ISM shows more adaptive and intelligence-driven security models. It is expected from organisations to transition from static controls towards continuous monitoring and assurance practices.

Governance models must expand to encompass AI

Executive accountability is no longer limited to traditional IT and cyber domains. It extends explicitly into AI usage and risk management. Organisations must establish formal governance structures that address AI lifecycle oversight, risk ownership, and decision accountability ensuring AI is deployed in a controlled and auditable manner.

Development practices must evolve alongside AI adoption

As AI continues to be embedded within development pipelines, secure development practices must adapt accordingly. This includes recognising AI-generated outputs as part of the code base subject to assurance. Organisations must enhance validation process to ensure that both human and machine generated components meet security, integrity, and compliance expectations.



What this means for compliance and assurance

AI introduces a new assurance scope

The update brings AI firmly into scope for assurance and compliance activities. Organisations are to treat AI systems as regulated assets, ensuring they are identified, risk assessed and likely to become increasingly relevant to System Security Plans (SSP) documentation, security assessments and assurance activities. This elevates AI to be treated with the same level of scrutiny as traditional systems, requiring clear governance, documentation and demonstratable control implementation.

Assurance shifts to monitoring AI behaviour

From the changes it is clear to see a key shift from static compliance to behaviour-based assurance. Organisations are recommended to not only demonstrate that the controls are implemented but the AI systems are operated within defined behavioural baselines over time with monitoring in place to detect drifts or anomalies. This introduces a more dynamic assurance model focused on outcomes rather than just implementation.

Human oversight and data governance

The update introduces new auditable requirements around human oversight and data handline Organisations are recommended to show that high risk AI actions are subject to human approval. The controls exist to manage data exposure including restrictions on external access and secure handling of prompts and outputs.

AI as both risk and control enabler

The ISM recommends organisations to assess AI from a dual perspective: as a potential risk and as a security enabler. This includes validating AI-generated outputs within development processes and ensuring AI-driven security capabilities are reliable, monitored, and fit for purpose. Collectively this drives a broader shift towards continuous assurance where organisation maintain real-time visibility or AI risks, behaviours, and controls effectiveness to sustain compliance.

Recommendations on how to respond

Strengthen AI governance and control

- Define accepted AI use cases and risk thresholds
- Implement 'human-in-the-loop' controls
- Establish monitoring for AI behaviour and outputs.

Secure AI data handling

- Restrict external integrations for sensitive environments
- Define retention and deletion policies for AI outputs
- Monitor prompt usage and data exposure risks.

Integrate AI into security operations

- Leverage AI for threat detection and analysis
- Enhance vulnerability management with AI assisted tools
- Incorporate AI into testing and assurance processes.

Update development and assurance practices

- Apply ISM controls to AI assisted development
- Validate AI generated code and outputs
- Embed AI considerations into DevSecOps pipelines.

Enhance monitoring and assurance capabilities

- Move from periodic audits to continuous assurance
- Implement real-time monitoring and alerting
- Align detection strategies with evolving AI-driven threats.

Conclusion

The latest ISM updates represent a targeted shift in how emerging technologies, particularly Artificial Intelligence, are governed, secured and operationalised within enterprise environments. Explicitly bringing AI into scope, the ACSC is signalling that these technologies are no longer experimental but that they are core to the organisations attack surface. It must be treated accordingly.

For organisations, this is not simply an incremental compliance uplift. It is a catalyst to modernise security operations, expand governance models, and embed stronger control assurance across increasingly complex and dynamic environments.

Those who respond proactively will not only strengthen their security posture but also position themselves to safely and confidently leverage AI at scale. Achieving this will require executive ownership, uplift across security and development practices, and a more integrated approach to managing risk across the full technology lifecycle.

References

[1] ASD – ISM June 2026 Change Document - ISM June 2026 changes (June 2026).pdf
[2] ASD – Information Security Manual (June 2026) – Information security manual (June 2026).pdf
[3] AI Policy Update: Strengthening responsible use across government | Digital Transformation Agency

Next article

Article four:

Months, not years:

Why the Five Eyes cyber warning demands action now

This article was written by:
Daniel Broad
Head of Managed Security Operations

The recent joint statement from the Five Eyes cyber security agencies should not be treated as another industry warning that is read, circulated and then quietly filed away.

The language was unusually direct.

Artificial intelligence is changing the cyber threat environment now, and organisations have months, not years to prepare.

That message came collectively from the cyber security leaders of Australia, Canada, New Zealand, the United Kingdom and the United States. These are agencies that sit close to national security, intelligence, critical infrastructure protection and major cyber incidents.

When they choose to make a public statement together, organisations should pay attention.

This does not mean every Five Eyes defence or intelligence capability is being brought together under a single command. It does mean that the five countries have reached a shared view of the threat. They are seeing the same direction of travel.

AI is reducing the time, cost and expertise required to conduct cyber operations. It is helping capable threat actors move faster and giving less capable actors access to techniques they would previously have struggled to use. The warning is not that AI will suddenly invent an entirely new form of cyber attack.

The warning is that it will make the attacks we already understand faster, cheaper, more convincing and easier to scale.

The real issue is speed

From a security operations perspective, the most important part of the Five Eyes statement is the compression of time. Many organisations still operate on security processes built for a slower threat environment.

A vulnerability is identified. A ticket is created. A risk assessment is completed. The affected business area is consulted. A change window is scheduled. Remediation may then occur days or weeks later.

That model becomes increasingly dangerous when an attacker can use AI to identify exposed infrastructure, analyse a vulnerability, produce supporting code, tailor social engineering and begin targeting organisations within a much shorter period.

This does not mean AI is independently conducting every stage of a sophisticated intrusion. It does not need to.

Even where a human remains responsible for the attack, AI can support reconnaissance, coding, translation, target research, impersonation, data analysis and decision-making.

A threat actor who can complete each of those activities faster can pursue more targets and adapt more quickly when security controls block the initial approach.

For defenders, this changes the operational equation. The question is no longer simply whether the organisation can detect an attack.

It is whether it can identify, understand and contain the attack before the adversary reaches a critical system, identity or business process.

Why a Five Eyes statement carries weight

The Five Eyes partnership is built on long-standing intelligence cooperation between Australia, Canada, New Zealand, the United Kingdom and the United States.

Most of that activity does not occur publicly. That is why a coordinated public warning from the leaders of these agencies matters.

They are not commenting on a hypothetical future risk. They are signalling that they believe the capability of AI systems is developing quickly enough to change defensive priorities today.

The Five Eyes agencies have highlighted both sides of the equation. AI can assist attackers with vulnerability discovery, social engineering, exploitation and operational scale.

It can also support defenders by improving analysis, identifying weaknesses, accelerating investigations and helping security teams process large volumes of data.

That creates a race between offensive adoption and defensive adoption.

The organisations that benefit will not simply be those that purchase the most AI products. They will be the organisations that use AI to improve the speed and consistency of decisions while maintaining strong human oversight.

> The organisations most at risk will be those that automate fragmented or poorly understood processes and assume technology alone will compensate for weaknesses in governance, architecture or operational readiness.

What the security community is saying

The response across the Australian and New Zealand cyber security community has broadly reinforced the seriousness of the statement. Much of the commentary has focused on the need for faster defence, stronger resilience and better prioritisation. That is understandable.

Most organisations are not short of security data. They are short of time, context and experienced people who can decide what matters.

Security teams may be managing thousands of vulnerabilities, large volumes of alerts, multiple cloud environments, ageing infrastructure and an expanding number of third-party services. The challenge is not finding more things that could be wrong. The challenge is determining what creates the greatest risk and acting before an attacker does.

Industry commentary has also connected the Five Eyes statement with broader warnings from Australia and New Zealand intelligence and national security community.

That connection is important. Cyber security, espionage, foreign interference, fraud, critical infrastructure disruption and geopolitical risk are not isolated problems.
They often intersect.

A compromised identity may be used to steal information. The information may support fraud or espionage. Access may be retained for future disruption. Stolen data may then be used for impersonation, coercion or influence activity. AI can accelerate each part of that chain. Organisations therefore need to move beyond narrow technical assessments of individual alerts and consider the wider operational and strategic consequences of compromise.

Estimated attack timeframes

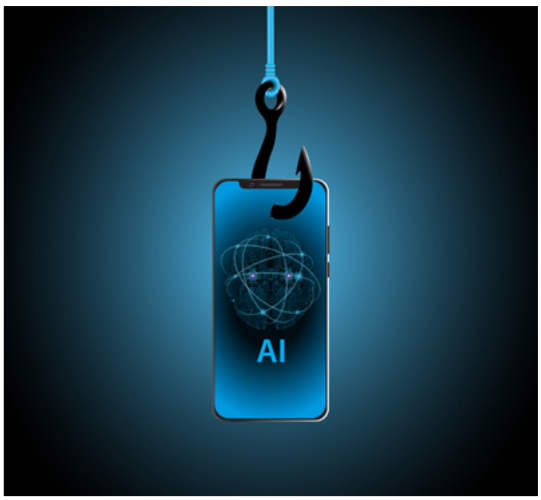
There is no single authoritative dataset that measures every stage of a traditional cyber attack against an AI-enabled equivalent. In practice, most modern attacks are hybrid. Human operators remain responsible for the overall objective and key decisions, while AI and automation are used to accelerate reconnaissance, develop phishing content, analyse vulnerabilities, identify attack paths and process stolen information.

The figures below should therefore be treated as indicative rather than definitive. Actual attack speed will depend on the complexity of the target, the access method used, the maturity of the attacker and the effectiveness of the organisation's defensive controls.

Stage	Traditional attack Est.	AI-enabled attack Est.
Reconnaissance	1–14 days	Minutes–4 hours
Initial-access preparation	Hours–7 days	5–60 minutes
Establish foothold	30 minutes–4 hours	Seconds–30 minutes
Privilege escalation and lateral movement	1–24 hours	5 minutes–2 hours
Data discovery and collection	4–48 hours	10 minutes–2 hours
Exfiltration or impact	1–48 hours	Minutes–several hours

Even with those limitations, the direction is clear. A traditional intrusion may take days or weeks to progress from initial reconnaissance through to data theft, disruption or fraud. An AI-enabled attack can compress much of the same activity into hours and, under favourable conditions, minutes.

Reconnaissance is one of the clearest examples of this acceleration. A traditional attacker may spend days manually researching employees, technology platforms, suppliers and exposed infrastructure. AI can analyse large amounts of public information, identify likely access points and generate detailed target profiles in a much shorter period.



The same compression applies to initial access. Attackers can use AI to create convincing phishing messages, imitate writing styles, translate content and tailor lures to specific individuals or business functions. AI can also support exploit development and help less experienced attackers troubleshoot technical problems that would otherwise slow them down.

Once access has been gained, AI can assist with analysing permissions, identifying valuable accounts, mapping internal systems and determining possible pathways toward critical assets. It can also help process large volumes of files and communications, allowing the attacker to locate valuable information more quickly.

Recent industry reporting demonstrates how little time defenders may have. CrowdStrike reported an average eCrime breakout time of 29 minutes, measuring the period between initial access and lateral movement. Its fastest observed breakout occurred in 27 seconds. Unit 42 found that the fastest quarter of incidents reached data exfiltration within 1.2 hours, while an AI-assisted attack simulation progressed to exfiltration in approximately 25 minutes.

These figures should not be interpreted as evidence that every AI-enabled attack will succeed within minutes. Complex environments, strong identity controls, segmentation and effective security monitoring can significantly delay or stop an attacker.

The more important change is that AI can remove much of the manual delay between attack stages. Reconnaissance, exploit analysis, credential assessment and data discovery can be conducted in parallel rather than as a slow, sequential process.

This changes the operational challenge for defenders.

Security teams may no longer have several hours to investigate an unusual login before the attacker moves deeper into the environment. By the time an initial alert has been reviewed, the adversary may already have established persistence, accessed additional accounts or begun identifying sensitive data.

The comparison can be summarised simply:

Traditional attack:
Days to weeks

AI-enabled attack:
Minutes to hours

The central risk is not that AI has made every cyber attack fully autonomous. It is that AI allows human-led attacks to move faster, operate at greater scale and adapt more quickly when defensive controls are encountered.

These timeframes are illustrative and will vary depending on the target environment, access method, attacker capability, available automation and the strength of the organisation's defensive controls.

AI creates new risks inside the organisation

There is another side to this discussion. Organisations are rapidly adopting AI internally, often faster than their security and governance controls can keep up.

AI tools are being integrated into software development, customer service, document management, analytics, cyber security and business operations. Some of these systems are being given access to sensitive data, enterprise applications and automated workflows. **That access creates risk.** An AI agent that can read email, search internal documents, modify cloud environments or initiate transactions should not be treated like a basic software application.

It should be treated as a non-human identity.	**It needs an owner.**	**It needs a defined purpose.**
It should have minimum necessary access.	**Its actions should be logged and monitored.**	**There should be clear limits on what it can do without human approval.**

Organisations also need to account for risks such as prompt injection, sensitive data leakage, malicious or poisoned data sources, model theft, insecure integrations and the use of unapproved AI tools by employees.

> **The security of AI cannot be separated from the wider cyber security strategy.**

What organisations should do now

The Five Eyes statement provides a clear reason for boards and executives to revisit their current level of preparedness. The first step is to reassess the threat model. Organisations should consider how AI changes the capability of the threat actors most likely to target them. That includes nation-state actors, organised cybercriminals, fraud groups, insiders and opportunistic attackers.

The next priority is attack surface reduction. Every unnecessary internet-facing service, dormant identity, unsupported platform and excessive permission creates an opportunity. Reducing exposure is not always the most visible security initiative, but it remains one of the most effective.

Vulnerability management also needs to become more intelligence-led. Organisations cannot patch everything at once. They should prioritise based on evidence of exploitation, internet exposure, business criticality, attack paths and the consequences of compromise, not severity scores alone.

Identity security must remain a central focus. Phishing-resistant multifactor authentication, privileged-access controls, removal of dormant accounts, stronger service-account governance and just-in-time access should be treated as core resilience measures.

Security operations should also be reviewed. Organisations need to know whether their monitoring covers the systems and identities that matter most. They should understand which attack techniques they can detect, which they cannot, and whether their teams have clear authority to act.

Threat intelligence must be converted into operational outcomes. That may include:

- New detection logic
- Focused threat hunting
- Exposure reviews
- Control testing
- Changes to incident response procedures

AI can support these activities, but it should not replace accountability.

Automated investigation and response must operate within defined boundaries, confidence levels and escalation paths.

Resilience must be tested, not assumed

One of the most important themes in the Five Eyes statement is the need to verify that controls will work during a real incident. Too many organisations rely on plans that have never been tested under pressure.

Tabletop exercises should no longer focus only on traditional ransomware scenarios.

They should include AI-enabled impersonation, rapid exploitation after a vulnerability disclosure, compromise of an enterprise AI agent, manipulation of internal data sources and coordinated cyber and fraud activity. These exercises should involve more than the security team.

Executives, legal, communications, technology, risk, fraud, business continuity and operational leaders all need to understand their roles. The objective is not to produce a perfect response. It is to expose assumptions and weaknesses before a real attacker does.

The questions boards should be asking

Boards do not need to become cyber security engineers. They do need to understand whether the organisation can make effective decisions at the speed the threat now requires. The most useful questions are practical.

Q.

How quickly can we identify that a critical service is under attack?

Q.

How long does it take to disable a compromised account?

Q.

Which systems would cause the greatest operational damage if they were unavailable?

Q.

Where are AI systems and agents being used across the organisation?

Q.

What information and systems can they access?

Q.

Can we recover our critical services without relying on compromised infrastructure?

Q.

When did we last test our response under realistic conditions?

Q.

Are cyber security, fraud, intelligence, resilience and crisis management operating from the same picture of risk?

These questions provide a far stronger indication of preparedness than simply counting security products or reviewing compliance dashboards.

The warning has already been given

The Five Eyes statement does not make established cyber security principles obsolete. It reinforces them.

Secure-by-design technology, strong identity controls, timely patching, defence in depth, effective monitoring and tested incident response remain essential. What has changed is the amount of time organisations have to make those controls work.

AI is increasing the speed at which adversaries can identify and exploit weaknesses. It is lowering the barrier to some forms of malicious activity. It is also creating new enterprise systems, identities and dependencies that must be protected.

The organisations best positioned to respond will not necessarily be those with the largest security budgets. They will be the organisations that can connect intelligence, technology, people and decision-making into a coherent operational capability.

The message from the Five Eyes agencies is clear. This is not a future problem. The warning has been issued, and the time for organisations to test their readiness is now.

References

[1] Australian Cyber Security Centre — Five Eyes cyber security agencies statement
[2] UK National Cyber Security Centre — Impact of AI on cyber threat from now to 2027
[3] UK National Cyber Security Centre — Why cyber defenders need to be ready for frontier AI
[4] CrowdStrike — 2026 Global Threat Report
[5] Palo Alto Networks Unit 42 — Global Incident Response Report
[6] Anthropic — Disrupting the first reported AI-orchestrated cyber espionage campaign
[7] Anthropic — What we learned mapping a year's worth of AI-enabled cyber threats
[8] Snyk — What are AI attacks?
[9] DeepStrike — AI cyber attack statistics 2025
[10] SecurityBrief Australia — Five Eyes AI cyber warning prompts calls for faster defence
[11] Peter A. Clarke — Five Eyes release statement on cyber security: A call to action
[12] OneStep Group — The Five Eyes warning on AI cyber risk
[13] ASPI Strategist — Five Eyes, ASIO warnings are not separate. Nor should our response be
[14] The Guardian — AI models capable of devastating attacks months away, Five Eyes warns
[15] The Guardian — Once, cyber-attacks required great skill. AI is changing that

FUJITSU

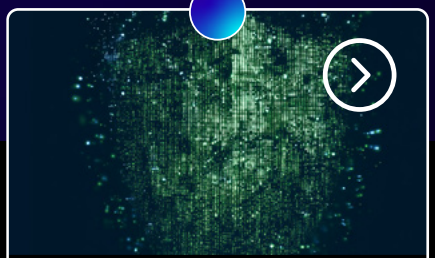
Fujitsu Cyber

Built to enhance cyber resilience

We are a Trans-Tasman team providing **end-to-end cyber security solutions designed to protect, enable, and transform organisations in Oceania.** We help you align with best practices, strengthen your defences, and ensure your systems are resilient and compliant. **Our cyber security services are structured around three core pillars:**

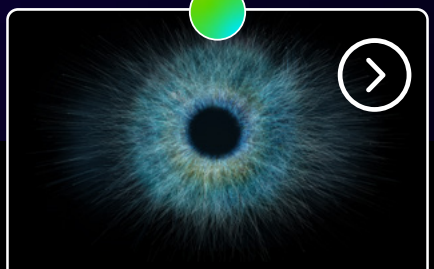
Advisory & Assurance

Delivering tailored consulting, strategic roadmaps, and hands-on support to help you identify risks, align with standards, and build resilience - empowering confident, secure business growth.



Technical Consulting

Uncover vulnerabilities and validate your defences through expert-led assessments and security testing. We deliver solutions aligned with business objectives while designing and implementing robust, future-ready security architectures.



Managed Security Operations

Providing 24/7 monitoring, proactive threat detection, and swift incident response to safeguard your organisation from evolving cyber threats.

Visit our website

Fujitsu Cyber draws on all parts of the business to identify key trends and changes with relevance to companies operating in New Zealand and Australia, both now and in the future. These threats are not solely technical. They can also arise from business operations, regional conditions in New Zealand and Australia, or global events that influence the cyber security environment in both countries.

Our research is the result of collaboration across the entire Australia and New Zealand team, including detection engineers, threat intelligence analysts, threat researchers, automation engineers, digital forensics and incident response specialists, as well as training and awareness professionals.

View all of our previous Threat Intelligence Reports here

Authors:

Connor Owens
Senior SOC Analyst

Jack Bartlett
Junior SOC Analyst

Haley Southgate
Senior Consultant

Shreya Dhoopad
Consultant

Daniel Broad
Head of Managed Security Operations

Curated by:
Hilary Bea, Thomas Hacker

Compiled by:
Ed Goodacre
Digital Content Specialist

Contact us

Ready to strengthen you cyber resilience, get in touch