# Microsoft
# Microsoft Digital Defense Report 2025

**Lighting the path to a secure future**

A Microsoft Threat Intelligence report

## How to navigate this report

Use the navigation bar above to jump between different sections of the report.

↗  Read more outside this report

+  Read more in this report

<  Previous page

>  Next page

**For more news on cybersecurity, visit:** https://microsoft.com/corporate-responsibility/cybersecurity

**For more report insights, visit:** https://microsoft.com/mddr

**For more news on cybersecurity policy, follow us on LinkedIn:** https://aka.ms/MOILinkedin

# Introduction

**Introductory statement by Amy Hogan-Burney and Igor Tsyganskiy**

# Mobilizing for impact:
## Cybersecurity leadership in a defining era

**Amy Hogan-Burney**
Corporate Vice President,
Customer Security & Trust

**Igor Tsyganskiy**
Corporate Vice President and
Chief Information Security Officer

We are living through a defining moment in cybersecurity. As digital transformation accelerates, supercharged by AI, cyber threats increasingly challenge economic stability and individual safety. Cyber threats are rapidly evolving from technical problems affecting business to events impacting all aspects of our society.

The pace of change in the threat landscape has pushed us to rethink traditional defenses. The growth and adoption of AI by both defenders and threat actors benefits both sides. AI in cybersecurity is already creating new challenges for security organizations as they rush to adapt systems, understand new threats, and equip their people with new knowledge to keep pace.

Cyber threats are also playing an increasingly significant role in geopolitical conflicts and criminal activities, creating both a wide and deep scope of responsibility for defenders. AI will play a critical role in helping security professionals productively address the growing threat landscape, but as an industry we must step into this new paradigm cautiously. With the increased speed of an AI-centric world, the impact of action—whether by security organizations, criminal actors, or nation states—will have faster and potentially greater second, third, or fourth-order effects. It is imperative that defenders consider these ripple effects as they implement new security controls, share security research, fix new security vulnerabilities, and collaborate with each other.

Adversaries, whether nation-states, criminal syndicates, or commercial cyber mercenaries, are leveraging emerging technologies to attack with both greater volume and more precision than ever before, often by exploiting the trust that underpins our digital lives. International collaboration among defenders will be critical to define new coordinated defenses and set new international norms that enforce consequences for cyberattacks targeting the global critical infrastructure or essential services.

For security leaders, the imperative is clear: cybersecurity must be a priority, embedded into the fabric of organizational strategy and addressed regularly as part of risk management. Global partnerships across industry peers and even competitors must be established to coordinate and collaborate on defenses against common adversaries. Traditional perimeter defenses are no longer sufficient. Resilience must be designed into systems, supply chains, processes, and governance. New types of threats will emerge with increasing frequency; being informed and prepared is critical.

## What's new in this year's report

### AI as both a defensive necessity and a target
We're witnessing adversaries deploy generative AI for a variety of activities, including scaling social engineering, automating lateral movement, engaging in vulnerability discovery, and even real-time evasion of security controls. Autonomous malware and AI-powered agents are now capable of adapting their tactics on the fly, challenging defenders to move beyond static detection and embrace behavior-based, anticipatory defense.

At the same time, AI systems themselves have become high-value targets, with adversaries amping up use of methods like prompt injection and data poisoning to attack both models and systems, which could lead to unauthorized actions, data leaks, theft, or reputational damage.

### Diverse vectors for initial access

In today's world, campaigns rely on multi-stage attack chains that mix tactics and techniques such as social engineering and technical exploits. This year, we saw the widespread adoption of "ClickFix," a social engineering technique that tricks users into executing malicious code themselves, bypassing traditional phishing protections. We also saw the incorporation of new access methods like device code phishing by both cybercriminal and nation-state actors.

### The pervasive threat of infostealers

Increasingly, adversaries aren't breaking in, they're logging in. In today's specialized cybercrime economy, access is essential, and infostealers are a way for operators to collect credentials and tokens for sale on the dark web. Follow-on activities by the buyers of compromised credentials can include ransomware, data exfiltration, and/or extortion. Overall, this means that organizations that experience an infostealer infection are at high risk of future breaches.

### Nation-state actors expanding operations

Geopolitical objectives continue to drive a surge in state-sponsored cyber activity, with a notable expansion in targeting the communications, research, and academia sectors. These expansions are mostly within expected scope and volume, and primarily focused on using cyber espionage against typical targets to complement traditional intelligence operations. Building on a trend we first noted last year, nation states continue to accelerate AI use to evolve their cyber and influence operations, making them more scalable, advanced, and targeted.

We urge you to read this report with a bias toward action. It is not just a reflection of the challenges both past and future; it is a call to mobilize, prepare, and confront. Innovation, resilience, and partnership are the pillars of a secure digital future. By embracing these principles, we can navigate uncertainty and build a world where technology empowers and protects us against the rising tide of threats.

*Amy Hogan-Burney*

**Amy Hogan-Burney**
Corporate Vice President,
Customer Security & Trust

*Igor Tsyganskiy*

**Igor Tsyganskiy**
Corporate Vice President and
Chief Information Security Officer

# About this report

## Commitment to responsible and ethical practices

Our approach to cybersecurity is grounded in our core values of responsibility, transparency, and ethical business conduct. We are dedicated to:

- Upholding the highest standards of privacy and data protection.
- Advancing responsible AI and quantum safety initiatives.
- Collaborating across sectors and borders to harmonize standards and share threat intelligence.
- Supporting global efforts to combat cyber mercenaries, safeguard human rights, and foster trust in digital content.

## Report scope

Microsoft fiscal year 2025 (July 1, 2024-June 30, 2025) unless otherwise stated.

Please note that due to rounding, the percentages in some charts may not total 100%.

## Our commitment to preserving privacy

Any and all data included in this report is presented in alignment to our privacy principles. Microsoft is committed to its focus on preserving customers' control over their data and their ability to make informed choices that protect their privacy. We advocate for strong global privacy and data protection laws requiring companies, including ours, to only collect and use personal data in responsible, accountable ways.

## Setting the stage for stakeholders

As you read this report, you will find actionable insights and recommendations designed to help leaders across government, industry, and civil society navigate the new realities of cybersecurity. Our commitment is clear: to build trust, drive innovation, and secure the digital future through responsible leadership and collaborative action.

We invite you to explore the findings, strategies, and vision outlined in this report—and to join us in shaping a safer, more resilient digital world.

# Our unique vantage point

Our global presence—spanning billions of users, millions of organizations, and a vast network of partners—provides us with an unparalleled perspective on the cybersecurity threat landscape.

Every day, we process more than 100 trillion security signals from across the world, from the broad spectrum of our customers, partners, and platforms. These signals originate from endpoints, cloud services, identity systems, and the intelligent cloud and edge, offering deep visibility into emerging threats, attack techniques, and adversary behaviors.

AI now plays a transformative role in our defense strategy, enabling us to synthesize vast data sets, detect novel threats, and respond in moments, not hours—empowering defenders to anticipate and disrupt attackers, to protect individuals, organizations, or critical infrastructure.

Yet, we recognize that no single organization can see or solve every challenge alone. By sharing our insights, lessons learned, and best practices in this report, we aim to strengthen collective cyber resilience and empower defenders everywhere.

Microsoft remains dedicated to transparency, collaboration, and innovation—helping build a safer digital future for all.



**Our breadth and depth of signals**

**100 trillion**
security signals processed daily

**4.5 million**
net new malware file blocks every day

**38 million**
identity risk detections analyzed in an average day

**15,000+**
Partners in our security ecosystem, making it one of the largest in the world
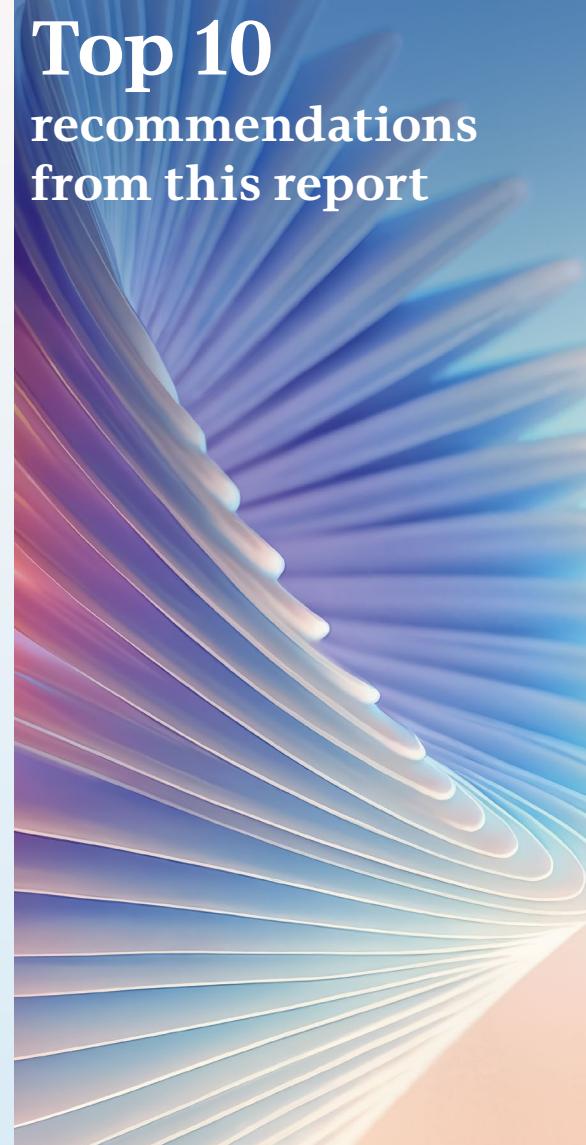
**34,000**
full-time equivalent security engineers employed worldwide

**5 billion**
emails screened daily on average to protect users from malware and phishing

# Top 10 recommendations from this report

### 1. Manage cyber risk at the boardroom level

Treat cybersecurity as a business risk on par with financial or legal challenges. It is important that corporate boards and CEOs understand the security weaknesses of their organization. Track and report metrics like multifactor authentication (MFA) coverage, patch latency, incident counts, and incident response time to develop a comprehensive understanding of both your organization's potential vulnerabilities and its preparedness in the event of a cybersecurity incident.

### 2. Prioritize protecting identities

Since identity is the top attack vector, enforce phishing-resistant multifactor authentication across all accounts, including administrative accounts.

### 3. Invest in people, not just tools

Cybersecurity is a whole-of-organization effort. Find ways to upskill your workforce and consider making security part of performance reviews. Culture and readiness—not just technology—are primary factors in both an organization's defenses and its resilience.

### 4. Defend your perimeter

A third of attackers use crude tactics as the easy path into an organization's exposed footprint, often looking beyond what you deploy to the vendors and supply chain you trust, including perimeter web-facing assets (18%), external remote services (12%), and supply chains (3%). Knowing the full scope of your perimeter, auditing the accesses you grant to trusted partners, and patching any exposed attack surface forces attackers to work harder to be successful.

### 5. Know your weaknesses and pre-plan for breach

Combine knowledge of the organization's exposure footprint with organizational risk awareness to develop a proactive plan for responding to future breach. Tie security controls to business risks in terms the board can understand. Since a breach is a matter of when, not if, develop, test, and practice your incident response (IR) plan—including specific scenarios for ransomware attacks, which remain one of the most disruptive and costly threats to operations. How fast can you isolate a system or revoke credentials?

### 6. Map and monitor cloud assets

Since the cloud is now a primary target for adversaries, conduct an inventory on every cloud workload, application programming interface (API), and identity within the organization, and monitor for rogue virtual machines, misconfigurations, and unauthorized access. At the same time, work proactively to enforce app governance, conditional access policies, and continuous token monitoring.

### 7. Build and train for resiliency

If breaches are all but inevitable, resilience and recovery become key. Backups must be tested, isolated, and restorable, and organizations should have clean rebuild procedures for identity systems and cloud environments.

### 8. Participate in intelligence sharing

Cyber defense is a team, not individual, sport. By sharing and receiving real-time threat data with peers, industry groups, and government, we can make it harder for cyber adversaries to achieve their goals.

### 9. Prepare for regulatory changes

It's more important than ever for organizations to align with emerging laws like the European Union (EU) Cyber Resilience Act or United States (US) critical infrastructure mandates, which may require reporting cyber incidents within a certain timeframe or Secure by Design practices. These regulations reinforce the importance of timely incident reporting and stronger internal oversight of an organization's cybersecurity practices.

### 10. Start AI and quantum risk planning now

Stay ahead of emerging technologies. Understand both the benefits and risks of AI use within an organization and adjust your risk planning, attack surface exposure, and threat models appropriately. Prepare for a post-quantum cryptography (PQC) world by taking the time to inventory where encryption is used and create a plan to upgrade to modern standards as they evolve.

# Part I
# The threat landscape

# Key takeaways

**What every leader needs to know about today's threat landscape**

## 1. Phishing-resistant MFA is the gold standard for security

No matter how much the cyber threat landscape changes, multifactor authentication (MFA) still blocks over 99% of unauthorized access attempts, making it the single most important security measure an organization can implement. Phishing-resistance provides an even more secure solution.

⊞ Read more on p23

## 2. Adversaries are targeting identities that enable access to data

Government organizations, information technology (IT) companies, and research and academic institutions were the most impacted by cyber threats this year. Among other data they hold that might interest adversaries, these organizations store vast amounts of personally identifiable information (PII), whose theft enables future attacks. Accessing organizational data has become a primary objective for threat actors. Government, NGO, and academic entities using legacy systems or operating with small IT teams and limited incident response capabilities should prioritize securing data and identity-facing assets.

⊞ Read more on p17

## 3. Adversaries are using diverse—but well-known—initial access routes

Incident response investigations found that 28% of breaches were initiated through phishing or social engineering, 18% were via unpatched web assets, and 12% leveraged exposed remote services. Not only are adversaries heavily leveraging the ClickFix social engineering method to deliver malware this year, but threat actors are incorporating exploits for known vulnerabilities faster than ever.

⊞ Read more on p13

## 4. Most attacks are for money, not espionage

More than half of cyberattacks with known motives had financial objectives such as extortion or ransom, while only 4% were motivated solely by espionage.

⊞ Read more on p11

## 5. Data exfiltration is the norm

Regardless of adversary motivations, accessing organizational data is now a primary goal for attacks. In the past year, we observed data collection in 80% of reactive engagements.

⊞ Read more on p29

## 6. Workload identities are under threat

As organizations implement phishing-resistant MFA and conditional access policies, adversaries are pivoting to targeting identities and elevated privileges granted to service-to-service workloads like apps, services, and scripts that access cloud resources because service-based workloads are often implemented with elevated privileges but weak security controls.

⊞ Read more on p17

## 7. Adversaries are conducting destructive attacks in the cloud

We have seen an 87% increase in campaigns aimed at disrupting Azure cloud customer environments through destructive actions such as ransomware or mass deletion. Additionally, over 40% of ransomware attacks now involve hybrid components.

⊞ Read more on p41

## 8. Adversaries are already using AI as a multiplier

Adversaries have begun implementing AI across a range of malicious activities, including for automated vulnerability discovery or phishing campaigns, malware or deepfake generation, data analysis, and to craft highly convincing fraudulent messages.

⊞ Read more on p52

## 9. Using AI can be both a benefit and a vulnerability

AI is driving rapid, substantial change. While it offers many benefits for organizations, particularly in cyber defense, AI can be attacked as well. As organizations implement the strengths of AI, they should also manage the weaknesses and potential exposure of sensitive data by protecting against threats like prompt injection, malicious tool invocation, and training data poisoning.

⊞ Read more on p52
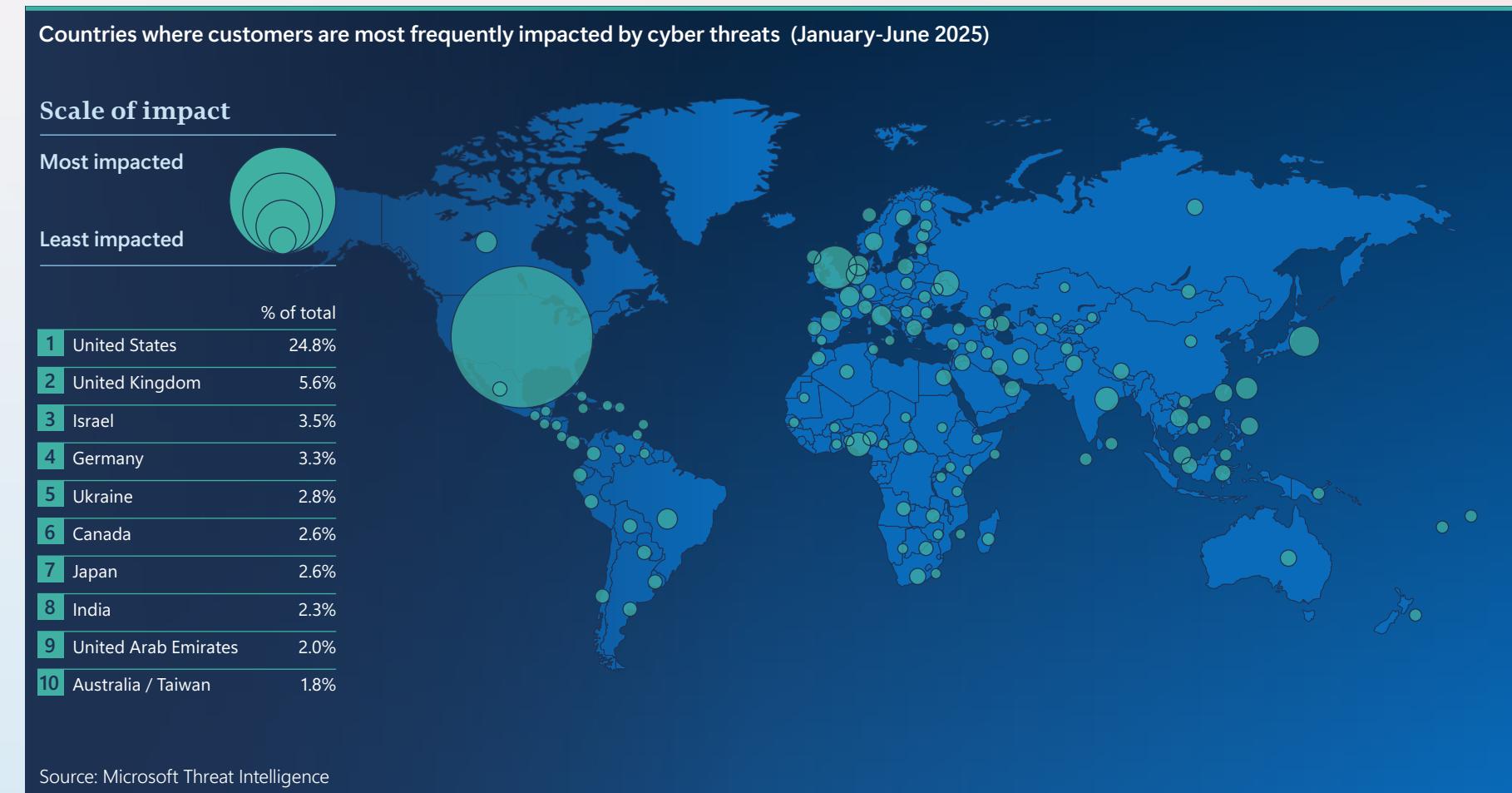
## 10. Quantum computing could challenge cybersecurity

Quantum computing has vast economic potential, but if used by malicious actors, it could threaten the encryption of sensitive data.

⊞ Read more on p57
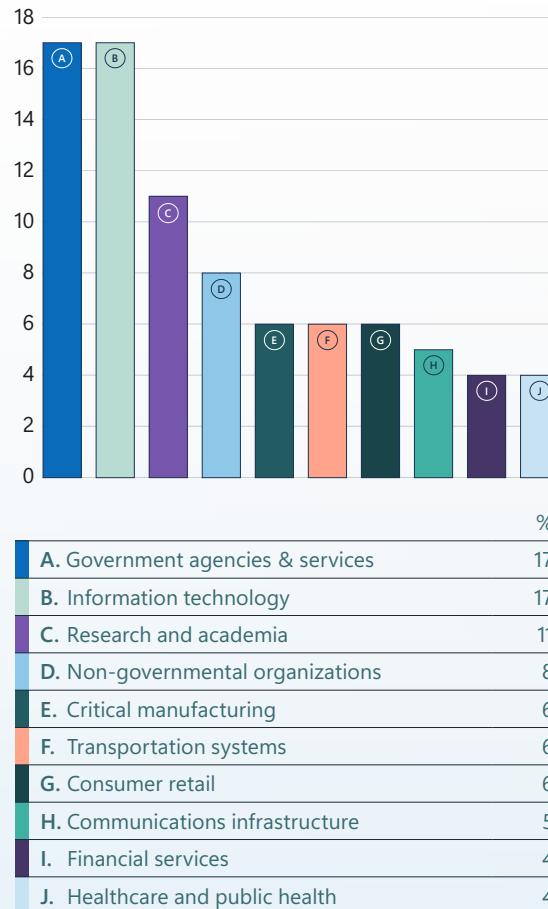
# How threat actors are shaping the cyber risk environment

Looking back over the past year, we've continued to see actors accelerate their development of new and novel techniques to challenge the defenses organizations are implementing to detect and prevent them. However, the daily threats organizations face largely remain the same: attacks by opportunistic threat actors targeting known security gaps. While users globally are at risk, we've observed most attacks in the last six months focused on the United States, the United Kingdom, Israel, and Germany.

**Countries where customers are most frequently impacted by cyber threats (January–June 2025)**

**Scale of impact**

Most impacted

Least impacted

| | | % of total |
|---|---|---|
| 1 | United States | 24.8% |
| 2 | United Kingdom | 5.6% |
| 3 | Israel | 3.5% |
| 4 | Germany | 3.3% |
| 5 | Ukraine | 2.8% |
| 6 | Canada | 2.6% |
| 7 | Japan | 2.6% |
| 8 | India | 2.3% |
| 9 | United Arab Emirates | 2.0% |
| 10 | Australia / Taiwan | 1.8% |

Source: Microsoft Threat Intelligence

**How threat actors are shaping the cyber risk environment** continued

## Ten global sectors most impacted by threat actors (January-June 2025)



| | % |
|---|---|
| A. Government agencies & services | 17 |
| B. Information technology | 17 |
| C. Research and academia | 11 |
| D. Non-governmental organizations | 8 |
| E. Critical manufacturing | 6 |
| F. Transportation systems | 6 |
| G. Consumer retail | 6 |
| H. Communications infrastructure | 5 |
| I. Financial services | 4 |
| J. Healthcare and public health | 4 |
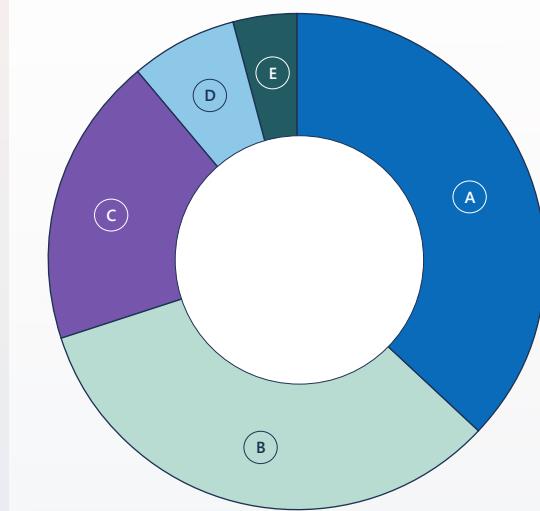
Source: Microsoft Threat Intelligence

IT and government bodies were the most impacted by cyber threats this year, from national to local entities. These organizations manage critical public services—for example, healthcare, research and academia, transportation, and public safety—and store vast amounts of sensitive data, including personally identifiable information (PII), tax records, and voting information. Additionally, many local governments operate on legacy systems that are difficult to patch and secure, and budget constraints and small IT teams often mean delayed updates, minimal threat monitoring, and limited incident response capabilities. This makes them high-value targets for both nation-state actors and financially motivated cybercriminals.

While attacks on IT, manufacturing, transportation, finance, energy, and healthcare can have both digital and physical consequences, attacks on industries like research and academia and telecommunications could additionally serve as a launchpad for attacks on other entities.

## Identified motivations in incident response engagements



| | % |
|---|---|
| A. Data theft | 37 |
| B. Extortion | 33 |
| C. Destruction/human-operated ransomware | 19 |
| D. Infrastructure building | 7 |
| E. Espionage | 4 |

Source: Microsoft Incident Response, Detection and Response Team
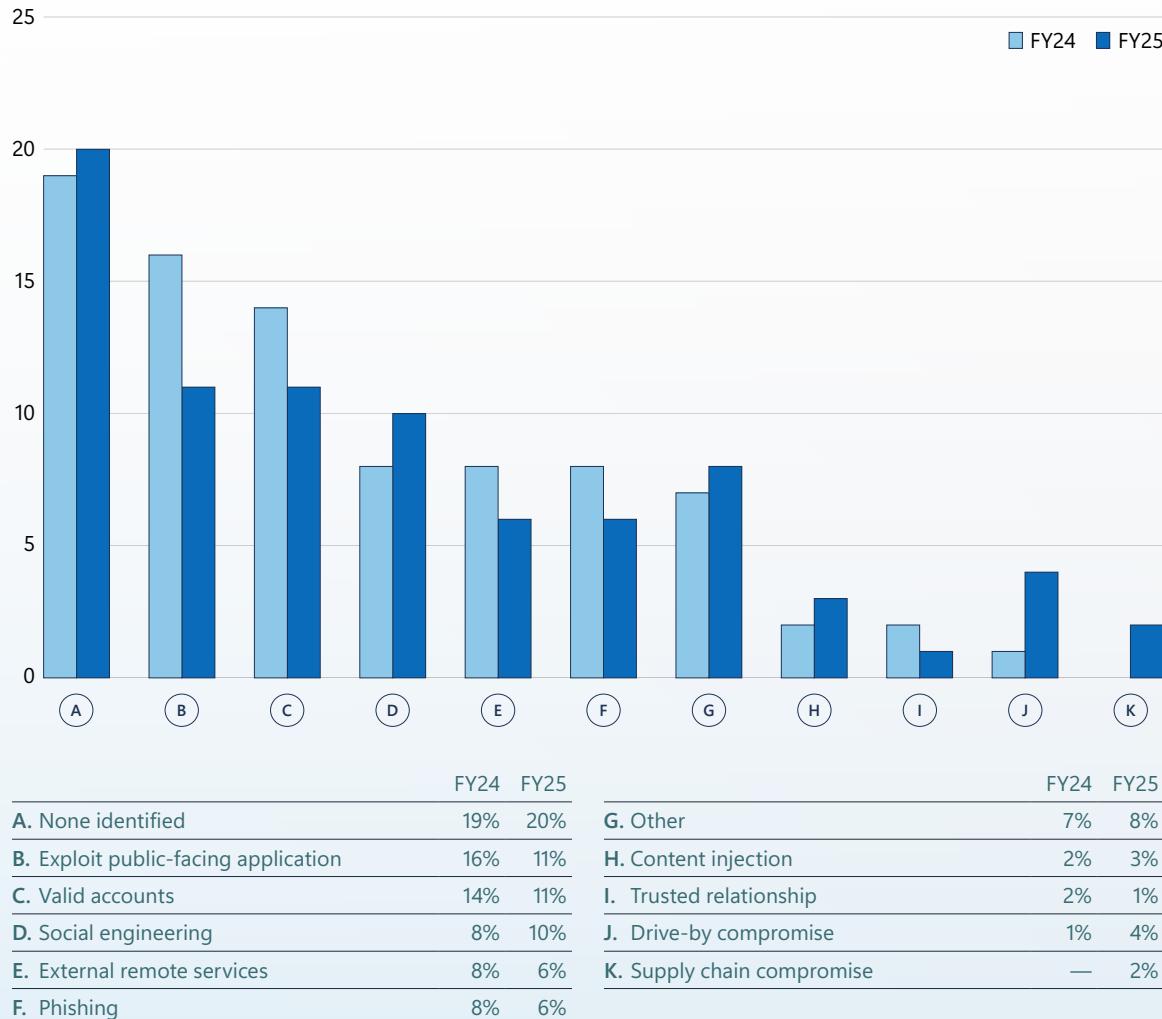
Attacker motivations and goals range from stealing sensitive information—such as personally identifiable information (PII), intellectual property (IP), or financial records—to disrupting business operations. The most common actions observed post-compromise are financially motivated extortion and ransomware operations. In incidents where we were able to determine threat actor objectives, we found that 33% involved extortion, while 19% used attempted destructive or human-operated ransomware attacks. We observed the deployment of a ransomware payload in 8% of engagements. In contrast, threat actors were motivated solely by espionage in only 4% of engagements. Notably, 7% of organizations were impacted by infrastructure building. This means threat actors might be taking advantage of organizations' unmanaged digital assets to stage attacks against other third-party targets downstream.

> " Threat actors were motivated solely by espionage in only 4% of our engagements.

**How threat actors are shaping the cyber risk environment** continued

## Comparing initial access vectors across two years



Legend: FY24 FY25

| | | FY24 | FY25 |
|---|---|---|---|
| A. | None identified | 19% | 20% |
| B. | Exploit public-facing application | 16% | 11% |
| C. | Valid accounts | 14% | 11% |
| D. | Social engineering | 8% | 10% |
| E. | External remote services | 8% | 6% |
| F. | Phishing | 8% | 6% |

| | | FY24 | FY25 |
|---|---|---|---|
| G. | Other | 7% | 8% |
| H. | Content injection | 2% | 3% |
| I. | Trusted relationship | 2% | 1% |
| J. | Drive-by compromise | 1% | 4% |
| K. | Supply chain compromise | — | 2% |

Source: Microsoft Incident Response, Detection and Response Team

# A ransomware attack with potential global impact stopped in under two minutes

In February 2025, the global economy narrowly averted catastrophe after a global shipping company experienced a ransomware attack. Had the company's systems been taken offline for even a few hours, the cascading effect would have impacted trade and industry around the world. Prolonged downtime would have ground maritime commerce to a crawl.

The attack epitomizes the risk of our interconnected world: a ransomware attack against just one private company can have global implications. Supply chains—both physical and digital—increase our attack surface, and organizations and industries halfway around the world can feel the knock-on effects of a single successful compromise. Malicious cyberactivity is not just a problem for individual victims to handle, but a whole-of-society problem.

As daunting as today's cyber threat landscape feels, this is a success story—proof that investing in cybersecurity pays off. Because the shipping company committed to protecting its digital assets, the attack was quickly stopped. **The time from observation to disruption was a mere 14 minutes, with encryption stopped one minute and eight seconds after it began.**

If the right protections are enabled, ransomware attacks can be contained at the onset of the attack, with no encryption at all.

**How threat actors are shaping the cyber risk environment** continued

While attack tactics, techniques, and procedures (TTPs) continue to evolve at a rapid pace, over the past year, attackers nevertheless persisted in targeting well-known pain points, regardless of targeted industry or attacker motivation. According to our incident response engagements, a significant portion of attacks begin by targeting an organization's exposure footprint: perimeter web-facing assets (18%) and external remote services (12%) as well as—to a lesser degree—supply chains (3%).

Threat actors are incorporating exploits for known vulnerabilities at a faster pace than before to target misconfigured or vulnerable web-facing applications and remote services. The rapid weaponization of exploits has increasingly impacted the windows between vulnerability disclosure, patch availability, and patch deployment. Ransomware operators and botnet distributors often choose targets of opportunity, using scanning tools or services to identify unpatched systems or copying the successful publicized attacks of other threat actors. Sophisticated threat actors are also targeting supply chains and trusted third-party relationships, which can affect downstream customers. By compromising a less secure partner or vendor in the supply chain, attackers could potentially impact more hardened targets in multistage attacks.

Managing an organization's footprint has become increasingly complex due to difficulties in understanding true exposure. Organizations can guard the wrong assets, lack a complete picture of their exposure, or struggle to address vulnerable devices. For cloud environments in particular, organizations might struggle to properly determine who has what access within their cloud tenant across the trust chain of software-as-a-service (SaaS) applications, guest accounts, and delegated privileges. This complexity is compounded by the fragmented nature of many security measures, as the lack of integration between security tools adds to the complexity and creates potential blind spots for attackers to exploit.

The abuse of valid accounts is also a frequently observed technique (17%). This can be the result of several types of attacks that maliciously gain access to user credentials—for example, theft, phishing, brute force, or social engineering—and use them to infiltrate systems without triggering traditional security alerts. As will be discussed later in this report, Microsoft has seen attackers acquiring stolen credentials on underground criminal forums to sign in directly to networks. And in cloud environments, we have identified multiple criminal and nation-state actors conducting entire end-to-end attacks as legitimate users or resources, with the ability to manipulate any resource or process that the compromised identity is trusted to access, including email, other cloud services, or the on-premises environment.

## Logging in: Today's playbook for initial access

This year, Microsoft Defender Experts observed a sharp change in how threat actors achieve initial access. Campaigns are no longer dominated by simple phishing and instead rely on multi-stage attack chains that mix technical exploits, social engineering, infrastructure abuse, and evasion through legitimate platforms. Specific initial access tactics observed include:

- ClickFix, an approach in which users are tricked into copying and pasting malicious code into their systems themselves (discussed more on page 36).
- Attacks combining email bombing, voice phishing (vishing) calls, and Microsoft Teams impersonation to convincingly pose as IT support and gain remote access.
- The deployment of rogue virtual machines using the open-source Quick Emulator (QEMU), giving attackers an isolated space to operate completely out of sight from traditional security tools.
- The exploitation of zero-day vulnerabilities in commonly used tools like SimpleHelp, BeyondTrust, Fortinet, Cleo, and Apache Tomcat.
- Use of malvertising, especially where malware is hosted on trusted platforms like GitHub and Discord and delivered through deceptive ads on high-traffic sites.
- The deployment of commodity infostealers to harvest credentials for future intrusions or resale by access brokers (discussed more on page 24).

These trends reveal a key shift in attacker strategy. Threat actors are no longer trying to force their way in—they're blending in. By abusing legitimate tools, platforms, and user behaviors, they gain access quietly, often without tripping standard detection mechanisms.

What these trends mean for defenders:

- **Trust is not enough.** Familiar platforms and tools could be—and are—abused. It's crucial to apply zero trust principles in your security model.
- **Endpoint visibility is insufficient.** Critical attack activity might occur outside the reach of endpoint detection and response (EDR).
- **Early-stage threats are critical signals.** Infostealers and credential theft should trigger investigation, not be dismissed as routine.

In 2025, initial access is no longer a single event—it's an extended process, carefully staged and tailored to avoid detection at every layer.

How threat actors are shaping the cyber risk environment continued

# Emerging threats: What's next from attackers

While attackers' motives don't change over time, the methods they use do, as they continually pursue new approaches to access, evasion, and persistence. Given the rapid advancement of AI, the decentralization of malicious actor infrastructure, and the rise of commercialized cyber capabilities, Microsoft believes the following emerging threats will play an increasing role in the next year.

## 1  AI-enhanced social engineering and attacks

The integration of generative AI into adversarial operations has significantly elevated the persuasiveness and scale of social engineering campaigns. As organizations improve their hardening against traditional cybersecurity threats, threat actors will increasingly turn to AI-enabled social engineering to achieve initial access. For example, these threat actors will leverage AI to improve the speed and effectiveness of their attacks by deploying autonomous malware capable of lateral movement, vulnerability discovery, and privilege escalation without human intervention.

Or they could use AI-powered agents capable of adapting in real time to defensive environments, rerouting command and control channels or rewriting payloads dynamically to evade EDR systems. This level of autonomy could enable them to conduct scalable, multi-vector intrusions across sectors with little operational overhead.

## 2  More supply chain compromise

For years, threat actors have increasingly exploited the interconnectedness of modern software ecosystems and operational structures to conduct malicious activity. Microsoft continues to observe threat actors targeting the trusted relationships with upstream managed service providers (MSPs), remote access services like virtual private network (VPN) or virtual private server (VPS) systems, remote monitoring and management (RMM) solutions, cloud backups, continuous integration/continuous delivery (CI/CD) pipelines, and third-party deployment vendors to gain access through trusted or commonly deployed IT systems. These intrusions generally compromise privileged vendor accounts, exploit unpatched software, or insert malicious code into legitimate components.

The persistent danger posed by supply chain threats highlights the need for organizations to audit access privileges, validate software bills of materials (SBOM), maintain dependency hygiene, and perform runtime integrity checks.

## 3  Expansion of covert, decentralized networks
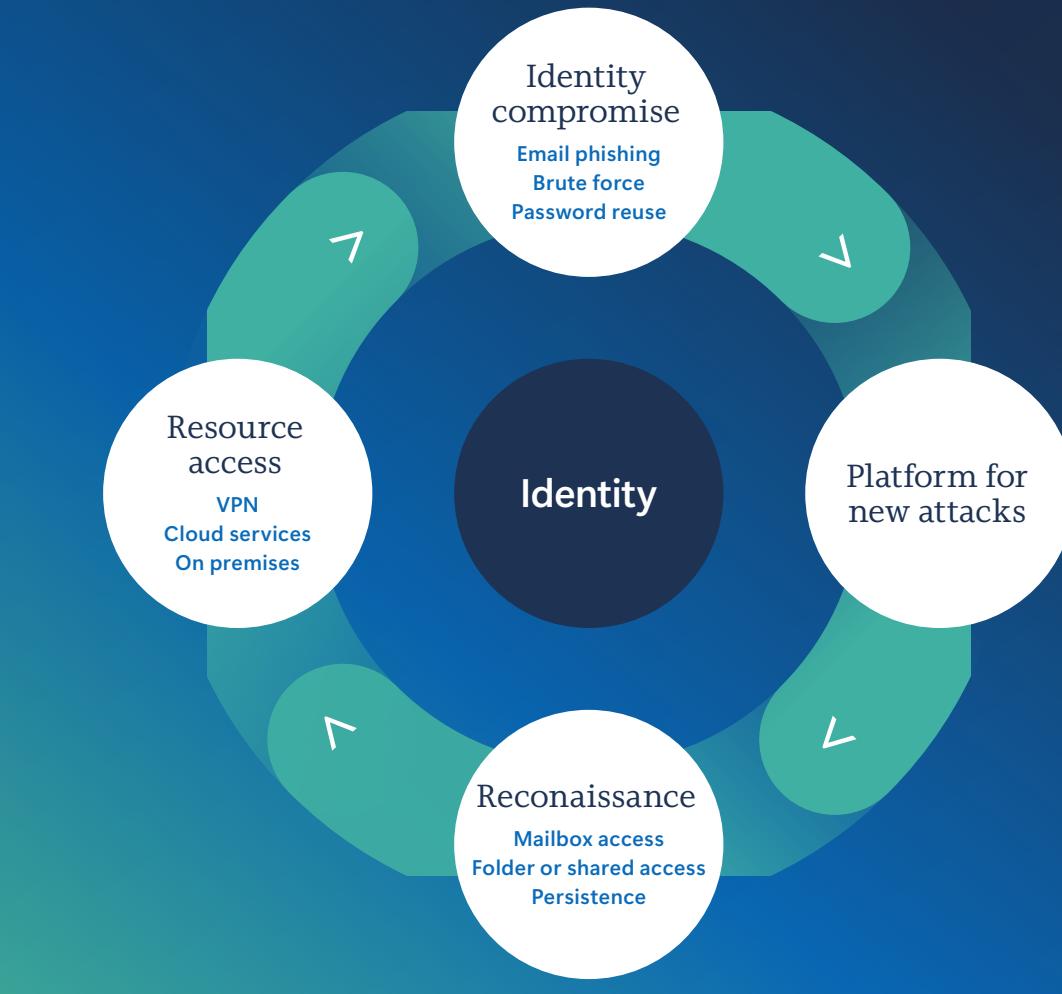
As threat intelligence and attribution capabilities improve, sophisticated threat actors are evolving their infrastructure strategies. Rather than relying on centralized command-and-control (C2) servers or conventional bulletproof hosting (which refers to hosting services that knowingly allow malicious activity to persist online), threat actors might shift toward peer-to-peer (P2P) covert networks built atop blockchain technologies or dark web overlays. These networks could be used to coordinate espionage, facilitate decentralized malware distribution, or obfuscate ownership and control of malicious assets. In particular, ransomware-as-a-service (RaaS) actors and nation-state actors are likely to create semi-autonomous affiliate networks that can survive takedowns and adapt quickly by redistributing workloads across participants, much like resilient botnets.

How threat actors are shaping the cyber risk environment continued

## 4 Increasing cloud identity abuse

Cloud identity systems are a primary target for attackers seeking persistent, covert access. Attackers are targeting these systems by deploying malicious OAuth apps, abusing legacy authentication, and evolving device code phishing and adversary-in-the-middle (AiTM) attacks. These methods bypass MFA and enable long-term access and data exfiltration without triggering alerts. To confront this threat, defenders must enforce app governance, conditional access policies, and continuous token monitoring.

## Lifecycle stages for a cloud abuse attack

Identity compromise
Email phishing
Brute force
Password reuse

Platform for new attacks

Identity

Resource access
VPN
Cloud services
On premises

Reconaissance
Mailbox access
Folder or shared access
Persistence

## 5 The growth of high-stakes commercial intrusion markets

Cyber mercenaries are private sector entities who offer their hacking skills and tools for hire and/or sale. As the commercial offensive cyber market continues to grow, so does the demand for high-precision, low-detection exploits. In the future, these markets could shift from surveillance to disruption. For example, a cyber mercenary might offer to sell a zero-click implant capable of disabling satellite uplinks or manipulating public financial data feeds to governments or corporate competitors. The commodification of such advanced capabilities introduces scenarios such as the outsourcing of sabotage or political interference campaigns, which would create layers of deniability and complicate attribution for defenders.

**The future threat environment is poised to become more adaptive, covert, and focused on using humans to achieve initial access. This shift will challenge existing security paradigms and demand more anticipatory, behavior-based defense models across the public and private sectors.**

# Identity, access, and the cybercrime economy

**Identity attacks in perspective**

## Modern multifactor authentication still reduces the risk of identity compromise by more than **99%**.

While attacks against identity infrastructure (such as Microsoft Entra, Okta, Identity Provider (IdP), and hybrid components) are still limited in volume and are rare relative to other attacks, their variety is increasing. Novel attacks are continually being discovered, often targeting on-premises to cloud vertical attack paths.

More than **97%** of identity attacks are password spray or brute force attacks

Less than **3%** of attacks are…

A Token theft by malware
**2.4042%**

B Infrastructure
**0.1692%**

C AiTM
**0.2375%**

D Attacks on MFA
**0.0033%**

E Consent phishing
**0.0005%**

Source: Microsoft Defender XDR and Entra ID Protection alerts (April-June 2025)

**Identity, access, and the cybercrime economy** continued

## From end users to workloads: The new horizon in identity threats

As phishing-resistant MFA and conditional access strengthen user defenses, attackers are pivoting to workload identities—apps, services, and scripts that access cloud resources. These non-human identities often hold elevated privileges but lack sufficient security controls, resulting in a growing blind spot that attackers are exploiting.

App consent phishing tricks users into granting malicious apps OAuth permissions, bypassing MFA and persisting beyond password resets. Key Vault pivoting involves compromising apps with access to secrets, enabling lateral movement and privilege escalation, often undetected. Microsoft has observed layered attacks that combine device code phishing and OAuth consent phishing, sometimes redirecting users to AiTM sites. Compromised identities are also used for internal phishing and lateral movement.

Identity protection must extend to every identity—including non-human identities—by verifying explicitly, enforcing least privilege, and assuming breach.

↗ **Learn more**

https://aka.ms/identity-attack-techniques

Configure cryptographic key auto-rotation in Azure Key Vault | Microsoft learn (May 2025)

---

**\***

In the first half of 2025, identity-based attacks rose by 32%. This escalation may reflect adversaries' increasing use of AI to craft highly convincing social engineering lures—posing new challenges for detection and response at scale.

---

## User impersonation tactics

### User impersonation

As organizations move to technologies like phishing-resistant MFA which make the hacking or phishing of passwords exponentially more difficult, adversaries are being forced to use more sophisticated methods to compromise user accounts. These include:

- **Token theft**. Stealing a user's token after they've authenticated, meaning no password compromise is necessary.
- **Slow password spray.** Trying multiple passwords over an extended period to avoid detection.
- **Location proximity emulation.** Mimicking a legitimate user's location to bypass policies with geographical restrictions.
- **One-time code (OTC) intercept.** Tricking a user into generating an OTC and then intercepting it to authenticate.

### Secret store compromise

A secret store is a secure, local vault that protects sensitive information—including API keys, passwords, tokens, and certificates—from unauthorized access, allowing only approved systems to retrieve them as necessary. While platforms like Microsoft Azure Key Vault, AWS Secret Manager, and HashiCorp Vault offer significant improvements over patchwork solutions of the 2010s, they've also become highly valuable targets.

### Application impersonation and malicious applications abuse

Attackers compromise applications and users with the same toolbox. Apps often have more permissions than they need—the exact elevated permissions that attackers seek. Another attack vector lures users into installing malicious apps and granting them broad permissions that the attacker can use until the user or the administrator explicitly revokes them. Application consent screens look legitimate and seem benign because they don't ask for credentials.

### Authentication system impersonation

The most catastrophic scenario in identity security is the theft of a signing key, which compromises the trust and integrity of entire identity systems. A signing key is the private half of a public-private cryptography key pair used to encrypt and decrypt data. It signs messages so that systems can verify their authenticity using the pair's public key. With a captured signing key, attackers can impersonate the authentication system itself, forging credentials to gain access to protected resources and high-value data.

## Strategic threats to the research and academia sector
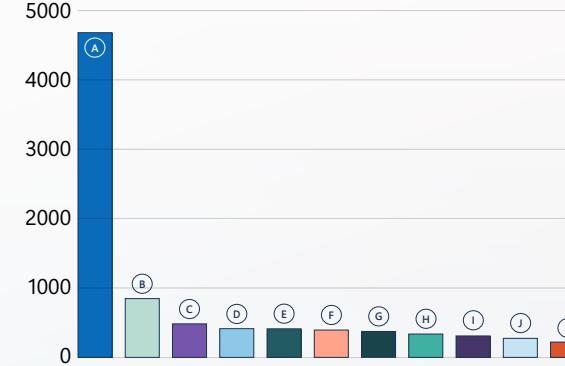
The research and academia sector continues to be a strategic incubator for adversarial cyber activity.[1] In 2025, it ranked among the top targets for threat actors due to its high-value IP, decentralized infrastructure, and expansive digital footprint. These conditions make it an ideal environment for adversaries to test and refine advanced attack techniques before deploying them against hardened targets such as government agencies and critical infrastructure.

Both nation-state actors and cybercriminal groups are leveraging the sector's open networks to pilot sophisticated identity-based attacks. Techniques such as AiTM, and AI-enhanced business email compromise (BEC) are increasingly prevalent. In the first half of 2025, identity-based attacks surged by 32%, with research and academia accounting for 39% of all identity compromise incidents observed by Microsoft. Environments across research and academia have some of the largest tenants and most complex identity systems of any sector, often making it difficult to detect and respond to advanced identity attacks.

Protecting the research and academia sector is both a community responsibility and a strategic necessity. Disrupting adversarial incubation here is critical to safeguarding downstream and upstream sectors.

### Count of unique organizations with identity compromise signals, by sector
(December 2024-May 2025)



| | | |
|---|---|---:|
| ■ | A. Research and academia | 4,647 |
| ■ | B. Services | 841 |
| ■ | C. Technology | 480 |
| ■ | D. Manufacturing | 411 |
| ■ | E. Miscellaneous | 409 |
| ■ | F. Travel | 391 |
| ■ | G. Retail | 371 |
| ■ | H. Energy | 334 |
| ■ | I. Logistics | 307 |
| ■ | J. Media | 272 |
| ■ | K. Healthcare | 219 |

Source: Microsoft Threat Intelligence, commercial cloud

### Inside the cybercrime marketplace: Brokers, mercenaries, and monetization

Cyber mercenaries can pose a serious threat to human rights, cybersecurity, and international stability as they enable governments that would otherwise lack the capability to conduct offensive cyber operations. While cyber mercenary products are often touted as enabling legitimate action against bad actors online, cyber mercenary intrusion capabilities have been widely used to target journalists, political dissidents, and other vulnerable groups. The cyber mercenary market is expanding rapidly, meeting a growing demand. According to the Atlantic Council, there are over 430 known entities operating in at least 42 countries.[2] This ecosystem includes intrusion experts, investors, intermediaries, and tech providers.

Although cyber mercenaries are frequently linked by the press to spyware, this gray market is much larger and poses even greater systemic risks—for example, the sale of zero-day vulnerabilities, which significantly destabilize the online environment and technology on which critical infrastructure relies by exposing a broad range of targets simultaneously through the breach of entire systems.

Because of the dangers associated with cyber mercenary activity, it's important for industry partners to work individually and together to combat the growing cyber mercenary market. Microsoft, for example, is committed to eradicating hack-for-hire services through its Digital Crimes Unit (DCU), which drives takedowns and enforcement actions against cyber criminals.

Microsoft is also a founding member of the Cybersecurity Tech Accord, which in 2023 laid out a set of principles on how to limit the activity of cyber mercenaries.

Governments, too, must do more to control this threat—for example, supporting the ongoing Pall Mall Process, which aims to create guardrails around the development, purchase, and use of commercially available cyber intrusion capabilities by supporting guiding principles for governments.

⊞ **Learn more on page 67**

*

A security researcher may earn $10,000 for responsibly disclosing a vulnerability to a bug bounty program, but may earn over $100,000 by selling the same exploit to a cyber mercenary.

↗ **Learn more**

Microsoft Corporate Responsibility | Cybersecurity

Identity, access, and the cybercrime economy continued
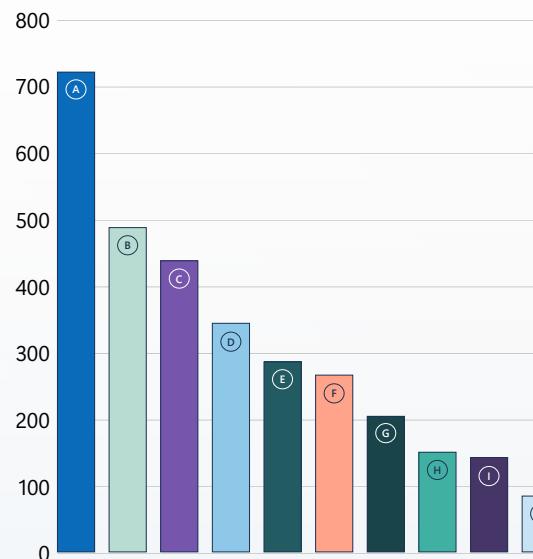
## Access brokers: The hidden gatekeepers of cybercrime

In the cybercrime economy's highly specialized and scalable ecosystem, access brokers play a pivotal role. These actors specialize in breaching enterprise environments and selling persistent access to other criminals, including ransomware operators, data extortion groups, and cyber mercenaries. Their services are foundational to the cybercrime-as-a-service (CaaS) model, enabling threat actors to outsource initial access and focus on monetization instead. These brokers often bundle access with reconnaissance data, making it even easier for buyers to deploy ransomware or exfiltrate data.

As part of a wider strategy to degrade infrastructure supporting large-scale cybercrime, Microsoft's DCU has intensified its focus on disrupting access brokers through a combination of legal, technical, and intelligence-driven actions.

In the last year, Intel 471 identified 368 access brokers, whose activities affected 68 industries across 131 countries and over 4,000 victims. These brokers primarily targeted victims in the United States (31%), the United Kingdom (6%), and Thailand (5%).[3]
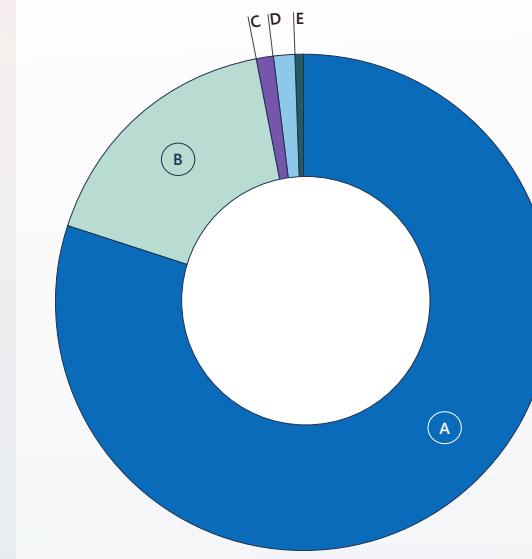
### Ten sectors most impacted by access broker activity



| | | |
|---|---|---|
| **A.** Public sector | | 722 |
| **B.** Consumer and industrial products | | 488 |
| **C.** Professional services and consulting | | 438 |
| **D.** Manufacturing | | 344 |
| **E.** Real estate | | 286 |
| **F.** Technology, media, and telecommunications | | 266 |
| **G.** Energy, resources, and agriculture | | 204 |
| **H.** Life science and health care | | 150 |
| **I.** Financial services | | 142 |
| **J.** Nonprofit sector | | 84 |

Source: Intel 471 data
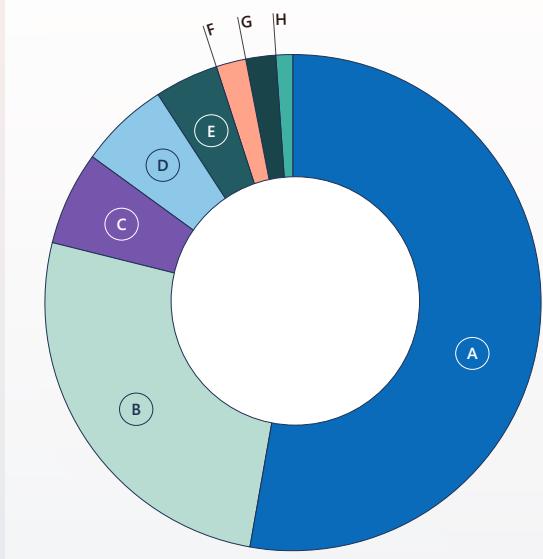
### Initial access vectors used by access brokers



| | | % |
|---|---|---|
| **A.** Credential-based attack | | 80 |
| **B.** Vulnerability exploitation | | 17 |
| **C.** Multiple | | 1.25 |
| **D.** Malware operation | | 1.25 |
| **E.** Insider access | | 0.5 |

Source: Intel 471 data

### Top access technologies offered for sale in the cybercrime economy



| | | % |
|---|---|---|
| **A.** RDP tools | | 53 |
| **B.** Corporate remote access portals | | 26 |
| **C.** Web server technologies | | 6 |
| **D.** Email platforms | | 6 |
| **E.** Victim-owned web infrastructure | | 4 |
| **F.** Government-owned web infrastructure | | 2 |
| **G.** Remote access protocol | | 2 |
| **H.** RMM tools | | 1 |

Source: Intel 471 data

**Identity, access, and the cybercrime economy** continued

## Exploiting vulnerabilities: The persistent threat of unpatched systems

Vulnerability exploitation remains one of the most reliable, scalable, and silent methods of initial access for threat actors. In the last year, Microsoft Defender Experts observed a surge in exploitation campaigns targeting known flaws in widely used enterprise systems and third-party IT tools. In most cases, exploitation achieves one of three outcomes:

- initial access into protected environments,
- privilege escalation from user to admin
- arbitrary code execution to enable lateral movement or persistence

This activity demonstrates that a strategic pivot toward infrastructure-level compromise is the new baseline for initial access.

What makes this threat vector especially dangerous is its lack of dependency on user interaction. From remote code execution (RCE) in infrastructure software to logic flaws in authentication mechanisms, attackers are increasingly skipping phishing and going straight for the code. Even misconfigurations in trusted platforms become high-value entry points. Most of these attacks start with a known Common Vulnerabilities and Exposures (CVE) exploit and end in compromise.

This year, key vulnerability exploitations that our Defender Experts observed included:

- SimpleHelp RCE chain (CVE-2024-57726/27/28)
- BeyondTrust Privileged Remote Access (PRA) and Remote Support (RS) Command Injection Vulnerability (CVE-2024-12356)
- Fortinet FortiClient EMS SQL Injection Vulnerability (CVE-2023-48788)
- Cleo Multiple Products Unrestricted File Upload Vulnerability (CVE-2024-50623)
- Apache Tomcat Path Equivalence Vulnerability (CVE-2025-24813)

Effective defense isn't just patching fast—it's expecting gaps and building layers of resilience through anomaly detection, behavior-based analytics, and hardening high-risk assets.

> " Vulnerability exploitation remains one of the most reliable, scalable, and silent methods of initial access for threat actors.

### Recommendations

**Patch fast, patch early**
Prioritize patching for high-impact CVEs, especially in internet-facing infrastructure and remote access tools.

**Isolate management interfaces.**
Where possible, restrict RMM tools and administrative consoles to management networks or VPN-only access.

**Employ exploit detection.**
Use behavior-based analytics to flag abnormal post-exploitation behavior (for example, Local Security Authority Subsystem Service (LSASS) access, registry dumping, and outbound tunnelling).

**Identity, access, and the cybercrime economy** continued

## Password spray: Anatomy of a high-volume attack

Despite their low per-attempt success rate, password spray attacks remain a persistent and high-volume threat.

These attacks rely on substantial infrastructure, allowing adversaries to distribute their activity across numerous IP addresses (IP).

Autonomous System Numbers (ASNs) are unique identifiers for collections of IP networks managed by single organizations. While over 50,000 ASNs carry authentication traffic daily, just 20 ASNs—only 0.04%—account for more than 80% of malicious password spray activity. This concentration underscores the importance of targeted threat intelligence and infrastructure-aware defenses.

Microsoft uses AI to analyze authentication data and detect subtle patterns of password spray activity hidden within legitimate traffic. When suspicious IPs are identified, authentication attempts can be temporarily blocked, disrupting attacker operations without affecting legitimate users. This approach enables real-time protection and adapts to evolving attacker tactics like automation and rapid IP rotation.

To avoid detection, attackers often employ a "low and slow" strategy, using a single IP address to target a small number of identities over extended periods. To reach a larger scale, they automate attacks across many IP addresses. Cloud-based infrastructure is particularly attractive to attackers, as it offers virtualization, orchestration, and access to a wide range of distinct IP addresses.
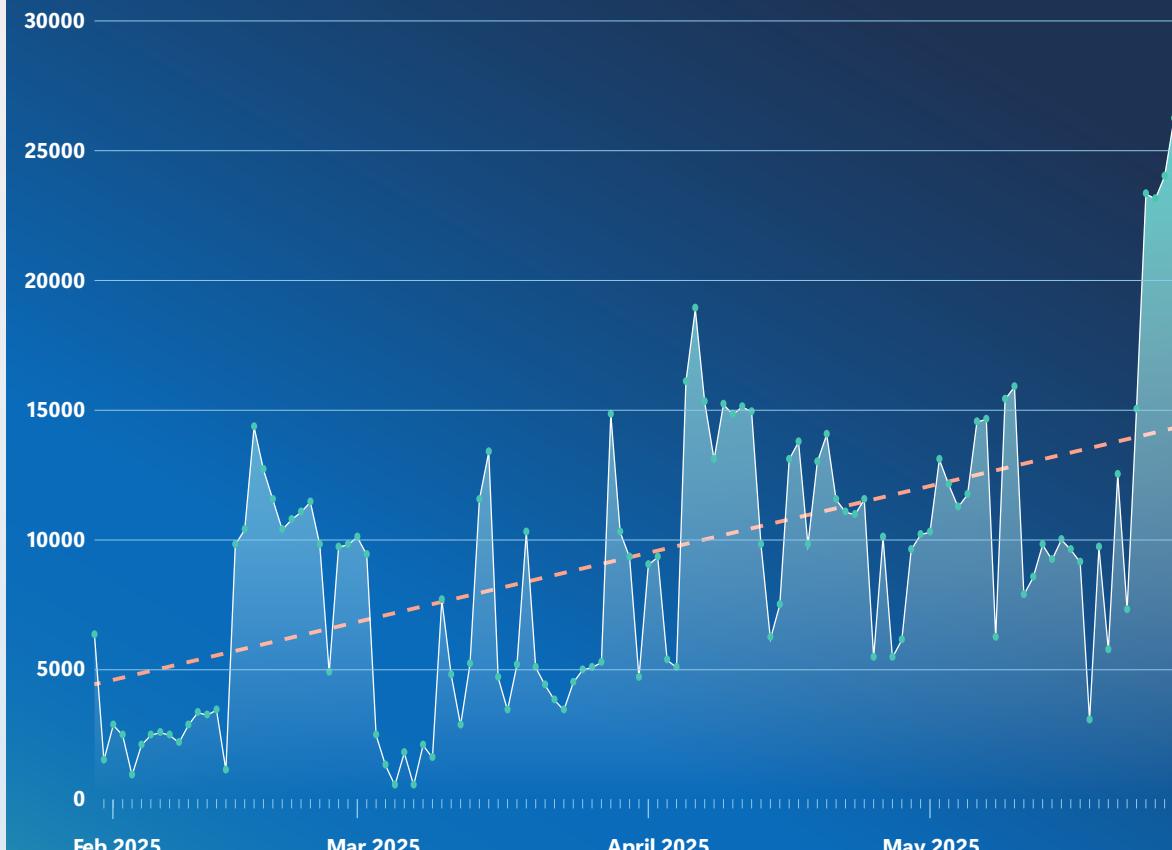
> " Just 20 ASNs—only 0.04%—account for more than 80% of malicious password spray activity.

**Count of IP addresses engaged in high volume password spray attacks by day (where count of targeted users is >50)**

This chart illustrates how attackers are using more IP addresses as a means to avoid detection. At the same time, advancements in AI are enabling defenders to identify more suspicious IP addresses. Together, this means more IP addresses are being detected that are involved in password spray attacks.
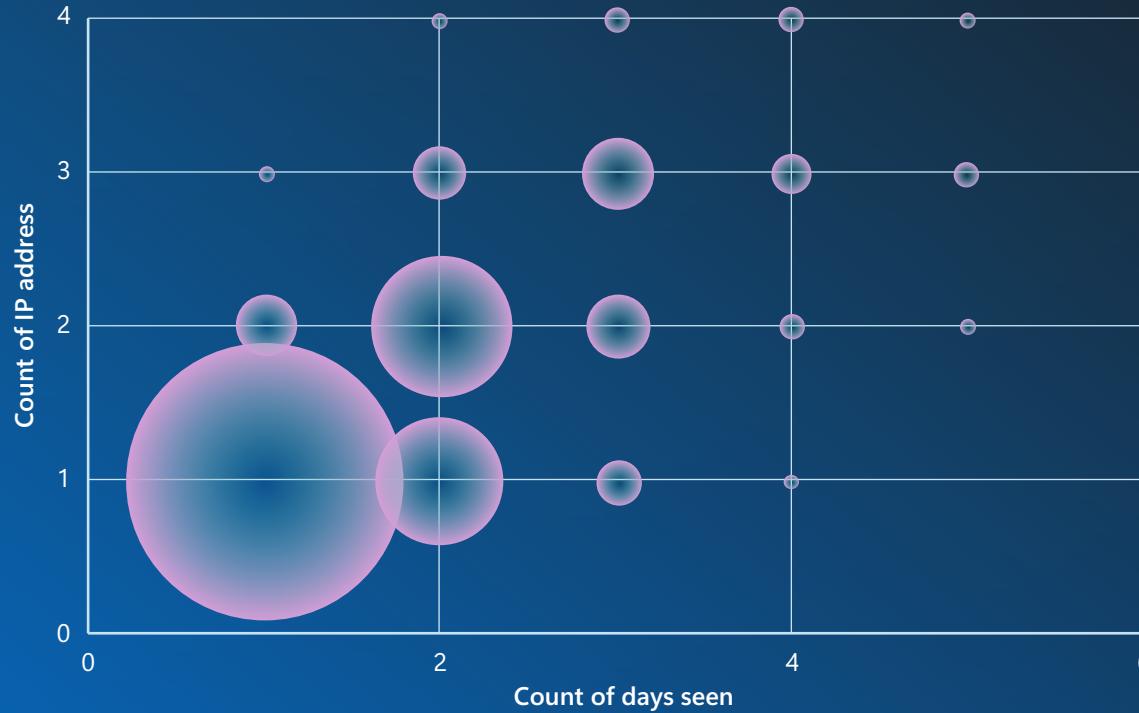


Source: Microsoft Digital Crimes Unit

**Identity, access, and the cybercrime economy** continued

## High volume password spray IP
How a password replay differs from a dictionary attack



Count of IP address (y-axis, 0 to 4)
Count of days seen (x-axis, 0 to 6)

Source: Microsoft Digital Crimes Unit

## Password spray IP addresses are transient



% of password spray IPs (y-axis, 0% to 40%)
Number of active days (x-axis, 0 to 15)

Source: Microsoft Digital Crimes Unit

## Credential abuse patterns
An analysis of 12.2 million accounts in a password spray attack reveals the following about cybercriminal login attempts prior to being blocked:

### Login attempts using correct username and password, but blocked by multifactor authentication: only 1.5%
This illustrates the limited MFA adoption in this scenario rather than its effectiveness. Given that modern MFA techniques are proven to prevent over 99% of identity-based attacks, expanding MFA usage across all accounts would dramatically reduce organizational risk.

### Incorrect passwords for valid usernames: "wrong password": 45%
This underscores the importance of avoiding password reuse, since usernames are commonly recycled.

Source: Microsoft Threat Intelligence

Most single username/password combination attempts are used for a single day, attacking from one IP address. This is generally seen in a replay attack, in which a threat actor replays a set of leaked usernames and passwords against Microsoft 365 accounts. When usernames are seen across multiple IP addresses and/or multiple days using multiple passwords, this generally represents a low and slow password spray attack. Multiple attempts at guessing a password are made in these attacks, often named a "dictionary attack." If the attacker were to try the same username with multiple passwords in close succession, the account would be temporarily locked out and easily detected.

**Identity, access, and the cybercrime economy** continued

## Target demographics and exposure

Research and academic environments remain disproportionately targeted in password spray attacks, accounting for 52% of observed spray attempts. Factors contributing to this include decentralized IT management, high user turnover, and inconsistent MFA enforcement—conditions also observed in other vulnerable sectors such as rural healthcare. A May 2025 comparative analysis with the Have I Been Pwned database revealed that 85% of usernames targeted in spray attacks appeared in known credential leaks. On average, each compromised username appeared in three separate logs, highlighting the magnitude of the global credential leak problem and the importance of users regularly changing passwords.

## Recommendations

**To reduce the risk and impact of password spray attacks, organizations should adopt a multi-layered identity protection strategy. This includes taking the following measures:**

### Enforce phishing-resistant MFA for all users
Phishing-resistant MFA remains the most effective control against unauthorized access using compromised credentials. Even when attackers possess valid usernames and passwords, MFA blocks access in over 99% of cases. Organizations should monitor for accounts with valid credentials but unenrolled MFA and enforce enrollment policies to close this gap. Organizations should also implement conditional access policies and use risk-based conditional access to block or challenge sign-ins from suspicious IP addresses, geographies, or device types.

### Monitor and block malicious IP addresses and ASNs
Continuously monitor authentication logs for error code 50053 and other indicators of spray activity. Block IP addresses and ASNs with repeated failed sign-in attempts or known malicious behavior.

### Audit and decommission stale accounts
Regularly review and disable inactive accounts, which are often targeted in spray attacks. Ensure that deprovisioned accounts are removed from all authentication systems.

### Educate users on credential hygiene
Promote the use of strong, unique passwords and discourage password reuse. Encourage users to check their credentials against breach databases such as Have I Been Pwned.[4]

### Deploy AI-based detection and response
Use AI-driven tools to detect anomalous sign-in patterns and flag potential spray attacks in real time.

> " On average, each compromised username appeared in three separate logs, highlighting the magnitude of the global credential leak problem.
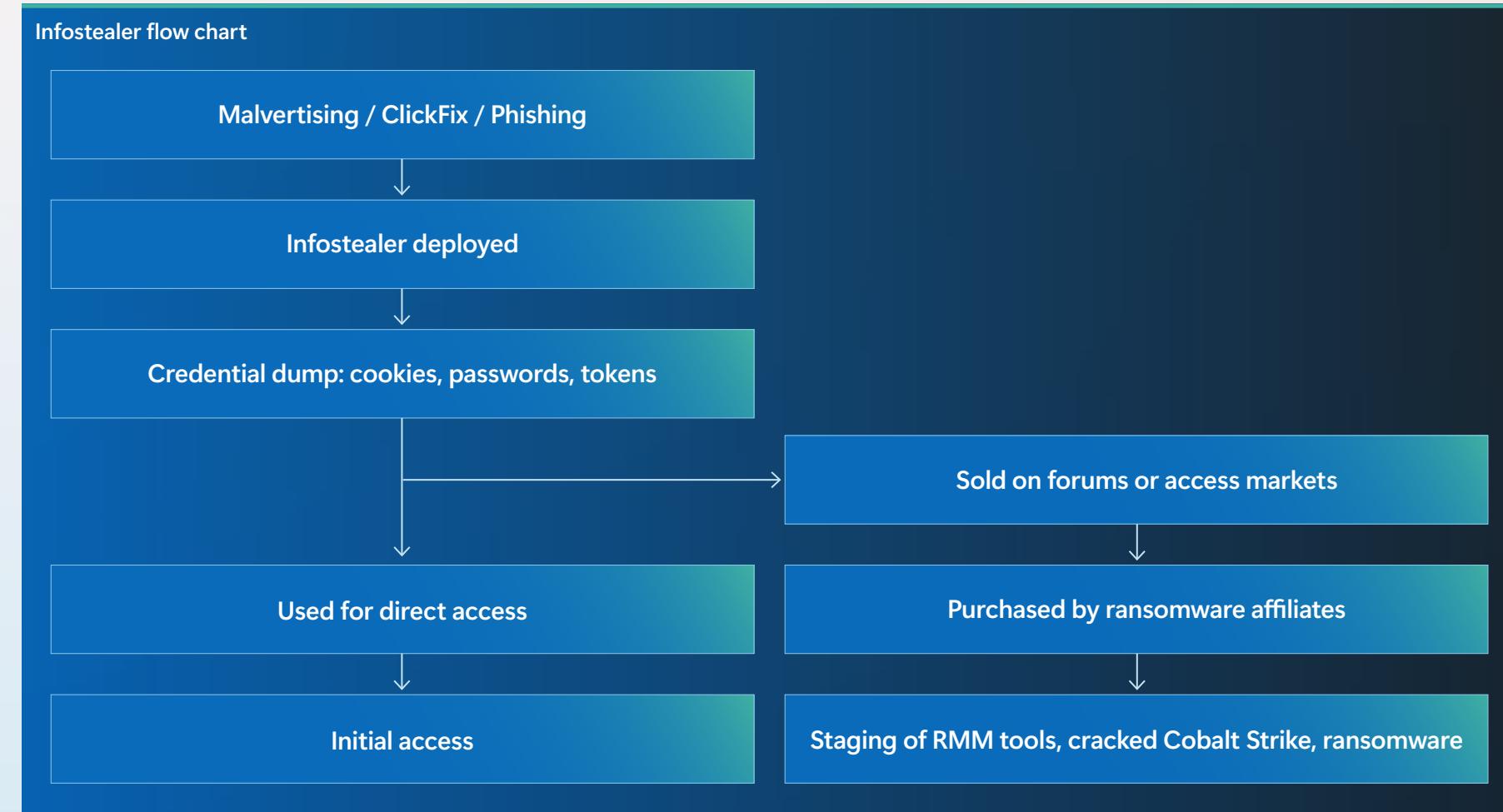
# Human-operated attacks and ransomware

**Human-operated intrusions: From infostealers to ransomware**

One of the most concerning trends this year is the rapid rise in the use of infostealers. Traditionally considered post-exploitation tools, malware families such as Lumma Stealer, RedLine, Vidar, Atomic Stealer, and Raccoon Stealer are now increasingly deployed as first-stage payloads.
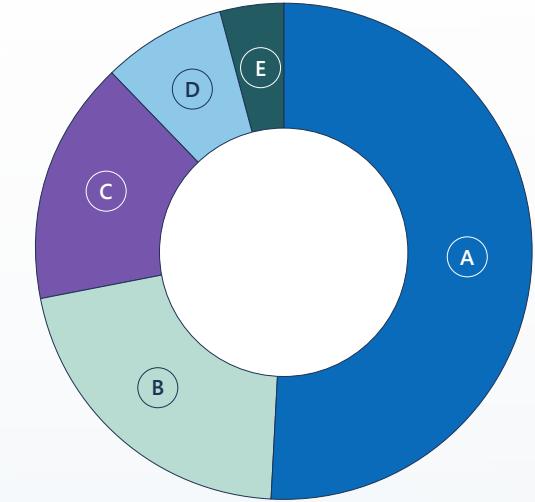
These tools, which are typically delivered through malvertising, search engine optimization (SEO) poisoning, cracked software, and deception techniques like ClickFix, are designed to collect credentials, browser session tokens, and system context data at scale.

This shift has elevated infostealers from isolated threats to foundational components of modern access campaigns. They enable a division of labor across the cybercriminal ecosystem: initial operators deploy the malware, access brokers monetize the stolen data, and users such as ransomware groups use it to gain footholds in enterprise environments. As a result, infostealer infections represent more than just local compromises—they pose a strategic risk of broader enterprise-wide intrusions.

**Infostealer flow chart**

Malvertising / ClickFix / Phishing
↓
Infostealer deployed
↓
Credential dump: cookies, passwords, tokens
↓ →  Sold on forums or access markets
↓         ↓
Used for direct access   Purchased by ransomware affiliates
↓         ↓
Initial access   Staging of RMM tools, cracked Cobalt Strike, ransomware

**Human-operated attacks and ransomware** continued

## Top five infostealers



|  |  | % |
|---|---|---|
| A. | Lumma Stealer | 51 |
| B. | Atomic | 21 |
| C. | Node.js based stealer | 16 |
| D. | Sys01 | 8 |
| E. | Rhadamanthys | 4 |
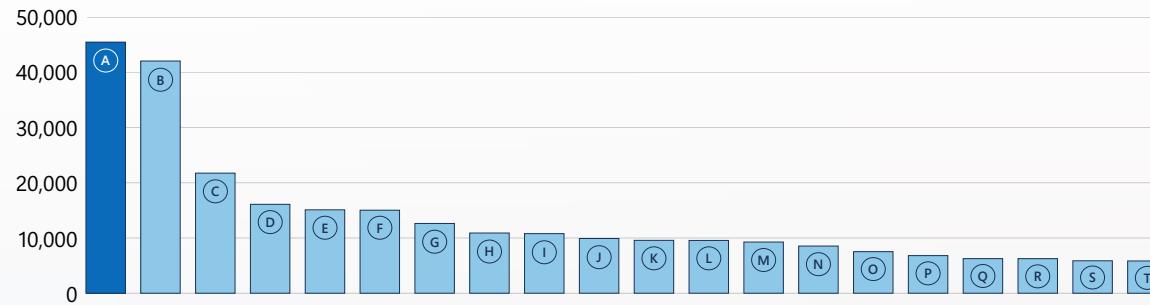
Source: Defender Threat Expert notifications

## Windows devices affected by Lumma (March 16-May 16, 2025)

Region: Worldwide (Top 20)



| A. India | 44,197 | K. Vietnam | 9,310 |
|---|---|---|---|
| B. Russia | 40,868 | L. Türkiye | 9,292 |
| C. Brazil | 21,137 | M. Philippines | 9,008 |
| D. United States | 15,647 | N. Colombia | 8,303 |
| E. Indonesia | 14,681 | O. France | 7,314 |
| F. Pakistan | 14,616 | P. Peru | 6,618 |
| G. Egypt | 12,277 | Q. China | 6,086 |
| H. Spain | 10,598 | R. Bangladesh | 6,083 |
| I. Argentina | 10,486 | S. Poland | 5,712 |
| J. Mexico | 9,634 | T. Germany | 5,680 |

Source: Microsoft Threat Intelligence

Lumma Stealer (also known as LummaC2 or LummaC) was the most prevalent infostealer observed in the last year. As a malware-as-a-service (MaaS) platform, Lumma Stealer is inexpensive, feature-rich, and constantly evolving. Its capabilities include real-time updates, credential theft, session hijacking, and crypto wallet draining. In early 2025, Microsoft observed a sharp increase in Lumma Stealer activity, with campaigns growing in both frequency and sophistication.

The scale and impact of Lumma Stealer's operations made it a priority target for disruption for Microsoft—leading to a landmark global intervention by the DCU in May 2025.

## Recommendations

**Defenders must treat infostealer infections as precursors to wider compromise, not isolated malware events.**

**We recommend:**

Hunting for loader activity (especially HijackLoader or Legion) that precedes payloads like Lumma Stealer

Blocking clipboard-to-shell behavior, especially PowerShell scripts from suspicious download paths

Monitoring for abnormal downloads from GitHub or Content Delivery Networks (CDN) mimicking popular software

Limiting password storage and autofill features on unmanaged or shared endpoints

Educating users about deceptive downloads, fake update pages, and cracked tools

**Human-operated attacks and ransomware** continued

## Lumma Stealer in Latin America

Countries such as Brazil, Argentina, Mexico are frequently targeted by cybercriminals, with credential theft, phishing, and ransomware the most common threats across the Latin America region. Credential theft has become the leading concern due to increased data breaches and frequent infostealer malware infections. Between March and May 2025, Brazil was the third most impacted country in the world by Lumma Stealer. More broadly, the Latin America region has been significantly affected by this infostealer.

To address these threats, Microsoft has strengthened partnerships with local law enforcement agencies, Computer Emergency Response Teams (CERT), and regional security teams. These collaborations facilitate intelligence sharing, victim notification, and affirmative disruption actions against malicious botnets such as Necurs and Trickbot and cybercriminal tools such as "cracked" versions of Cobalt Strike, which have been linked to over 68 ransomware attacks across 19 countries, including attacks on Latin America's healthcare sector.

**Windows devices affected by Lumma (March 16-May 16, 2025)**

Region: Latin America (Top 10)



| | | | |
|---|---|---|---|
| **A.** Brazil | 21,137 | **F.** Chile | 5,606 |
| **B.** Argentina | 10,486 | **G.** Venezuela | 2,617 |
| **C.** Mexico | 9,634 | **H.** Ecuador | 2,394 |
| **D.** Colombia | 8,303 | **I.** Dominican Republic | 1,918 |
| **E.** Peru | 6,618 | **J.** Bolivia | 1,415 |

Source: Microsoft Threat Intelligence

"

Credential theft has become a leading concern in Latin America, due to increased data breaches and frequent infostealer malware infections.

**Human-operated attacks and ransomware** continued
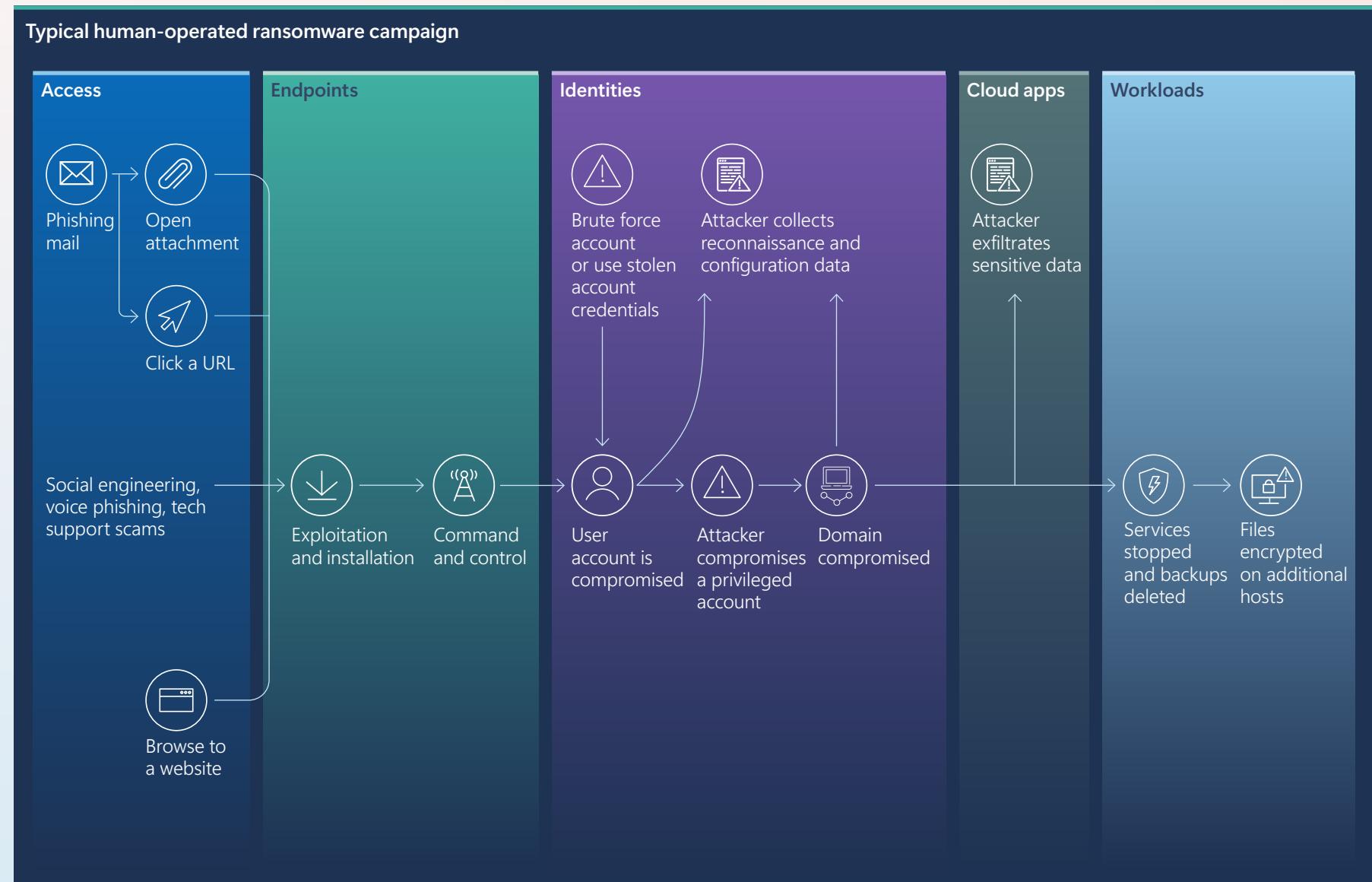
## Ransomware's shifting tactics

The overall ransomware picture this year did not change significantly from last year, with organizations worldwide continuing to face a persistent threat of attack from a small army of ransomware actors leveraging both commodity and custom ransomware strains.

According to Intel 471's review of ransomware leak sites, 120 ransomware variants were used against 71 industries.[3] Slightly over half (53%) of the victims were based in the US, while Canada (6%) and the United Kingdom (4%) were the next most impacted. Almost half (48%) of organizations whose size is known had an annual revenue of USD 50 million or less.
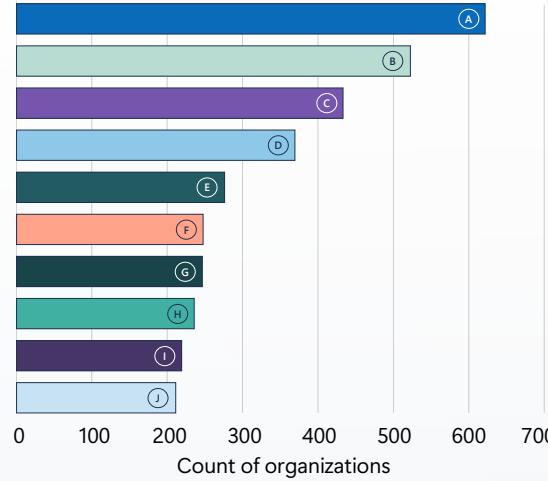
In a continuing shift away from phishing as the primary means of initial access, ransomware operators are increasingly leveraging social engineering to obtain or reset credentials, particularly through vishing or tech support scams. For example, this year multiple actors conducted help desk-themed social engineering, using messaging platforms such as Teams to communicate with targets and the Windows utility Quick Assist for remote access.

**Typical human-operated ransomware campaign**

**Access**
- Phishing mail
- Open attachment
- Click a URL
- Social engineering, voice phishing, tech support scams
- Browse to a website

**Endpoints**
- Exploitation and installation
- Command and control

**Identities**
- Brute force account or use stolen account credentials
- Attacker collects reconnaissance and configuration data
- User account is compromised
- Attacker compromises a privileged account
- Domain compromised

**Cloud apps**
- Attacker exfiltrates sensitive data

**Workloads**
- Services stopped and backups deleted
- Files encrypted on additional hosts

**Human-operated attacks and ransomware** continued

## Top ten industries impacted by ransomware



Count of organizations

| | | |
|---|---|---|
| A. | Industrial products and services | 633 |
| B. | Engineering and construction | 532 |
| C. | Retail, wholesale, and distribution | 441 |
| D. | Health care providers and services | 376 |
| E. | Technology | 281 |
| F. | IT or technology consulting | 252 |
| G. | Education | 251 |
| H. | Law services and consulting | 240 |
| I. | Transportation | 223 |
| J. | Financial and investment consulting | 215 |

Source: Intel 471 data

## Ransomed organizations by organization size in revenue (USD)



Revenue (USD)

| | | |
|---|---|---|
| A. | Over 1 billion | 239 |
| B. | 500 million – 1 billion | 124 |
| C. | 100-500 million | 464 |
| D. | 50-100 million | 405 |
| E. | 10-50 million | 1,739 |
| F. | 5-10 million | 1,013 |
| G. | 1-5 million | 4 |
| H. | Revenue not available | 1,722 |

## Five most prolific ransomware families (percentage of total)

| | |
|---|---|
| Akira | 22% |
| RansomHub | 11% |
| Fog | 11% |
| Qilin | 7% |
| Play | 5% |

Microsoft also observed several threat actors using fake software updates or ClickFix techniques to convince targets to download malicious software or run commands locally on their device.

Exploitation of public-facing applications also remains a key entry vector. For example, Storm-1175, known for deploying Medusa ransomware, has been observed exploiting vulnerabilities in several platforms. These exploits are often chained with credential theft and lateral movement to establish deeper access.

Meanwhile, Octo Tempest,the most sophisticated ransomware actor, uses advanced social engineering, SIM swapping, and identity compromise to access privileged accounts.[5] Known for its lateral movement techniques in cloud, this year the threat actor used Dragon Force, RansomHub, and Qilin, showing how easy it is for threat actors to move between RaaS affiliations. Octo Tempest continues to focus on targeting VMWare ESXi servers, often resulting in high-impact encryption events, particularly in hybrid environments.

Overall, targeting hybrid environments is becoming more prevalent, with ransomware operators leveraging compromised identities and tools like AADInternals to move laterally from on-premises into cloud environments. These techniques allow them to maintain persistent access, compromise multiple cloud applications, delete virtual machines (VM) and backup systems, exfiltrate data from cloud storage, and encrypt cloud resources.

Over 40% of ransomware attacks today involve hybrid components. Two years ago, less than 5% did.

As in years past, ransomware actors continue to use RMM tools for persistence and further intrusion. Approximately 79% of ransomware cases Microsoft observed this year involved at least one RMM tool.

Last year, we highlighted that ransomware actors were tampering with security solutions post-compromise. This year, we saw a focus on exploiting antivirus (AV) exclusions to avoid detection. AV exclusions are typically used by IT or security teams to stop AV software from wasting resources scanning trusted files or directories. Attackers seek out misconfigurations such as overly broad exclusions, which they could use to disable or sidestep defenses during hands-on-keyboard intrusions. This year, attackers used exclusions to bypass AV defenses in 30% of observed human-operated ransomware incidents.

Despite these evolving threats, attacks reaching the encryption stage have slowed and are now increasing at a rate of only 7% in 2024-2025 compared to 102% in 2023-2024, per our incident tracking. EDR solutions have proven highly effective at limiting the impact of attacks. Improved defense means attackers are now focused more on data exfiltration—in 82% of observed ransomware incidents, we saw large-scale data exfiltration.

## Data exfiltration and impact: Are you prepared?

At the outset of an incident response engagement, responders generally have to answer two primary questions: "How did the threat actors get in?" and "What data was stolen?" While proving exfiltration can be challenging, it remains a significant concern for customers, regulatory bodies, and downstream organizations.

In cases of stolen data, there is clear evidence that data has been extracted. In cases of data exposure, there is evidence that threat actors accessed sensitive data, but the process of exfiltration may not be visible or may not have occurred. Organizations and responders should adhere to zero trust and the 'always assume breach' principles when seeking evidence of access. In the past year, the Microsoft Detection and Response Team (DART) observed exfiltration in 51% of reactive engagements, while data collection—which includes data access and staging—was noted in 80% of engagements.

To address exfiltration effectively, it's important to remember that the absence of evidence indicating data exfiltration does not necessarily mean there's no impact. Understanding the motivations of the threat actor also provides crucial context. Financially motivated threat actors, for example, tend to be opportunistic, seeking large volumes of data for extortion or sale. Nation-state affiliated threat actors, on the other hand, focus on specific information such as intellectual property (IP) or state secrets. In either case, organizations should keep in mind the serious consequences that stolen data can pose, like legal risks, impact to industry accreditation, and reputational damage.

" Data collection—which includes data access and staging—was noted in 80% of reactive incident response engagements.

**Data exposure and exfiltration preparedness**

| Step | | Purpose |
|---|---|---|
| 1 | **Classify and inventory data** | **Identify and label data based on sensitivity, particularly crown jewels (most valuable data).** |
| 2 | **Protect critical data** | **Evaluate current protection mechanisms to ensure robust safeguards for sensitive information.** |
| 3 | **Establish response procedures** | **Understand obligations following data exposure/ exfiltration for compliance with legal and regulatory requirements. Understand obligations following data exposure/ exfiltration for compliance with legal and regulatory requirements. Establish a business resilience plan to ensure continuity of operations.** |
| 4 | **Maintain visibility and detection capabilities** | **Maintain oversight across all environments and implement rapid response to unusual data access.** |

**Human-operated attacks and ransomware** continued

# A study in time: What happens when you hesitate?

Time is of the utmost essence in cybersecurity. The ability of security professionals to respond swiftly, effectively, and efficiently to early signs of a potential breach determines whether an organization regains control or falls behind. In some cases, delaying a response even by one day could have a significant impact on an organization's ability to fully evict a threat actor and rebuild an environment successfully.

The length of threat actor activity is the number of days between the earliest identified evidence of threat actor activity and the latest. Among attacks investigated by DART, almost half (39%) lasted between zero and seven days from earliest to latest identified threat actor activity, and another 17% lasted between seven and fourteen days. Threat actors are moving faster than ever, making it even more important that organizations have the right mechanisms in place to match that speed.

⊞ Learn more on page 68

## Average time to engage by threat actor type

**Average length of threat actor activity**

## 58 Days
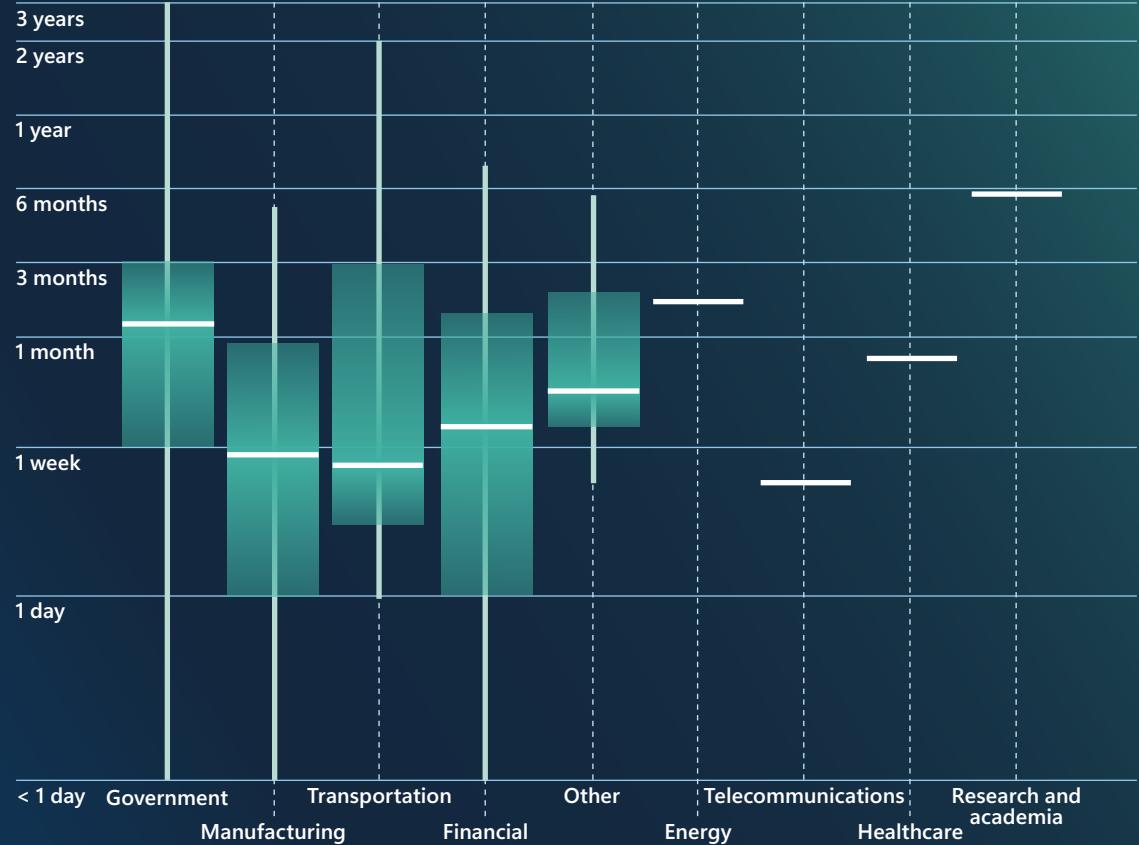
**Average dwell time**

## 12 Days

**Average time to engage**

## 9 Days

## Boxplot of length of threat actor activity by industry



This chart compares the length of threat actor activity across customer industries over the past year. The horizontal lines mark median duration in days, and the rectangular boxes indicate the range, where applicable. The research and academia sector recorded the longest average duration of activity, while the telecommunications sector experienced the shortest. These differences likely reflect the risk profiles (overall maturity) inherent to each industry and the threat actor goals.

Source: Microsoft Incident Response, Detection and Response Team (DART)

**Human-operated attacks and ransomware** continued

As threat actors move faster, they're using an increasingly aggressive attack chain. As a result, early detection is crucial. For 46% of our reactive engagements, the customer detected the threat actor's presence in their environment within 48 hours. Most attacks (59%) have short dwell times of 7 days or less.

Attacks with short activity are largely conducted by financially motivated actors. Threat actors prioritized evading detection and maintaining access primarily when attacking government entities.

When it comes to responding, 54% of customers engaged DART within three days of detecting a compromise, and nearly 70% did so within a week. Building an effective incident response plan allows organizations to quickly identify workstream leads, establish effective communication, set expectations with stakeholders, and call in experts. All of this can mean the difference between millions of dollars of impact.

## Recommendations for evaluating your incident response posture

Does your security budget support your organization's ability to rapidly respond to an inevitable cyber incident?

Do you have clearly defined roles and responsibilities in the case of a security incident?

Are you supported by a detection or a security operations center (SOC) team?
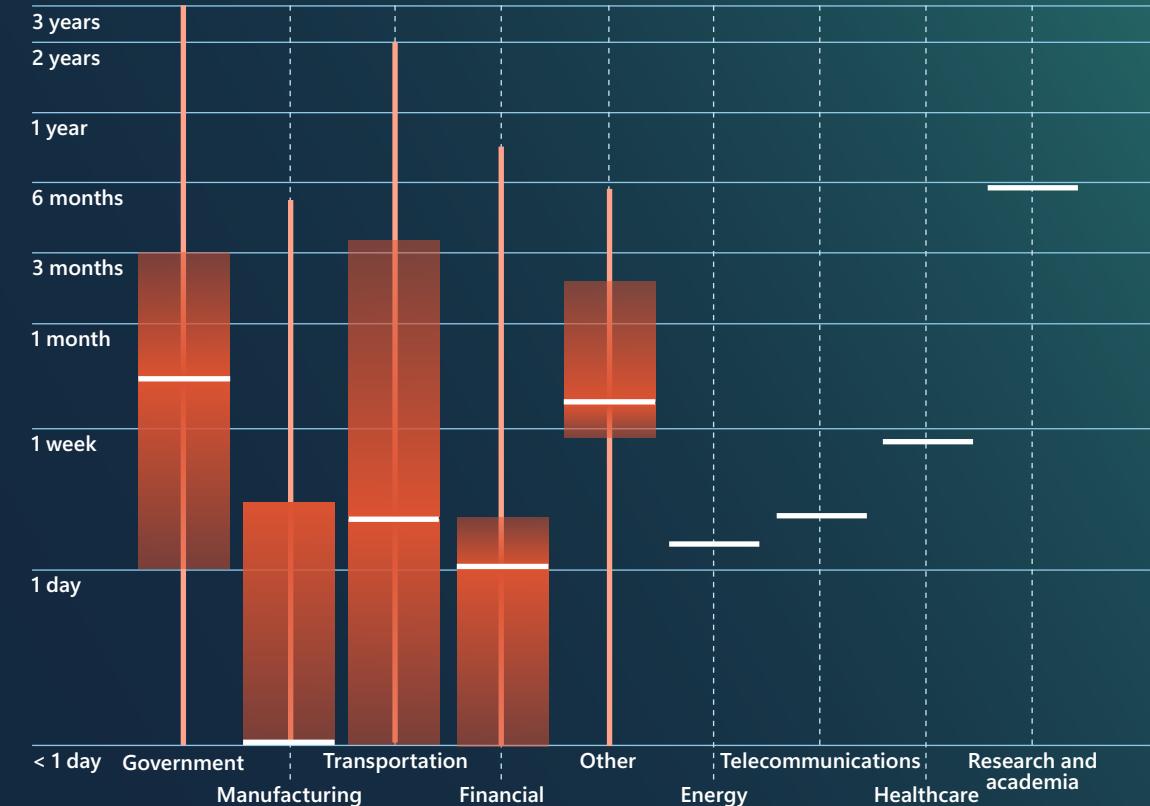
Do you conduct proactive threat hunting supported by threat intelligence?

### ⬈ Learn more

"Navigating the maze of incident response" is an evergreen product published by DART to provide a tactical guide and starting point for organizations building out their incident response processes who aren't sure where to start.

Navigating the Maze of Incident Response | Microsoft Security Blog

## Boxplot of dwell time in days by industry



This chart compares the average dwell time across customer industries in the past year. The research and academia sector had the longest average dwell time, indicating that threat actors remained undetected for a significant period. Conversely, the financial sector had the shortest average dwell time, suggesting quicker detection and response to threats. Dwell time is a function of attacker motivation in addition to being influenced by attack complexity and an organization's threat detection and hunting capabilities.

Source: Microsoft Incident Response, Detection and Response Team (DART)

# Fraud and social engineering

## The new age of super-charged fraud and what to do about it

Fraud is as old as commerce itself, rooted in the exploitation of trust and information gaps. Throughout history, fraudsters have adapted, leveraging new technologies and systems to deceive individuals, businesses, and governments.

Today, we face a pivotal shift: AI is amplifying the scale, speed, and sophistication of fraud and social engineering. While the core tactics remain unchanged—manipulating trust and exploiting human psychology—the risks are now global, immediate, and increasingly targeted.

### Malicious bot protection

As ecommerce and digital platforms continue to scale, so too does the sophistication of fraudsters who exploit automation to bypass traditional defenses. This is done through bots.

Bots themselves are neither good nor bad. They can be used to automate repetitive tasks, provide instant access to information, and enhance user experiences through personalized, real-time support. Bots can improve efficiency, reduce human error, and free up valuable time for more strategic work. Their ability to operate 24/7 and scale effortlessly makes them a powerful tool across industries.

At the same time, bot-assisted attacks are a rapidly evolving threat in the digital fraud landscape. More than 90% of the 15.9 billion Microsoft account creation requests in the first half of 2025 were from bad bots. Across the entire year, Microsoft's anti-fraud systems blocked approximately 1.6 million bot-driven or fake account signup attempts per hour across its services—an astounding volume indicating how attackers are abusing automation and false identities at scale.

Bots are increasingly used to execute high-speed credential stuffing, inventory hoarding, fake account creation, and card testing—often at a scale and frequency beyond human capability.

## Recommendations

Microsoft recommends organizations implement the following strategies to help identify malicious bots and detect bot-assisted cyber fraud:

### Residential proxy detection
Integrate third-party proxy intelligence databases. Build an internal proxy reputation system based on labeled "bot-assisted" transactions.

### Customer input pattern analysis
Detect automation by analyzing patterns in user-submitted data (e.g., names and addresses).

### Behavioral biometrics
Monitor mouse movement, click timing, and keystroke dynamics to distinguish bots from humans.

### Retrospective remediation
Deactivate accounts or subscriptions identified through offline detection.

### Advanced machine learning (ML) approaches
Use AI and ML to make sense of complex data—turning things like email addresses or product descriptions into comparable data points and analyzing user actions over time to spot unusual patterns or behaviors.

**Fraud and social engineering** continued

## The rise of deepfakes and synthetic identities: How AI is fueling identity fraud at scale

Using AI, scammers can quickly generate entire fake websites, profiles, and customer service chats to impersonate real businesses or use deepfake voices and videos to appear as trusted individuals, all at minimal cost.

In the past year, Microsoft thwarted USD 4 billion worth of fraudulent transactions and scams, many likely aided by AI content, and rejected 49,000 bogus enrollment attempts in its partner programs, stopping threat actors who were using fake or stolen identities to pretend to be legitimate partners.

Deepfakes involve using AI to create highly realistic audio and visual content, which can be used for malicious purposes such as impersonation, fraud, and misinformation. A deepfake impersonation can lead to business email compromise (BEC) or result in leaked information or the resetting of a password or two-factor authentication (2FA) for an important account.

Another area where AI deepfakes can be used is in tech support scams, where fraudsters impersonate a tech support agent to trick users (often seniors) into paying for fake support or installing malware. Traditionally, these scams used phone calls, emails, and pop-up ads; now threat actors are leveraging AI-modified voices when impersonating support agents for phone or video calls. These customer-facing deepfake tech scams directly impact not only the victims, but the impersonated company's reputation and customer trust.

**Microsoft fraud attempts thwarted**
Value of fraud schemes (many AI-enabled) blocked by Microsoft in one year (Apr 2024–Apr 2025)

# USD 4B

**Automated bot sign-ups blocked**
Fake account creation attempts (bots/synthetic) blocked on Microsoft services per hour

# 1.6M
per hour

On platforms like LinkedIn, there may exist fake profiles that use AI-generated portrait photos.[6] These fake LinkedIn personas might carry out data scraping or other abuses like social engineering (for example, posing as recruiters or vendors). This not only threatens LinkedIn's integrity but also can spill over into direct attacks on Microsoft employees or partners who might connect with a convincing fake profile.[7]

Synthetic identities are also a rising risk. In the digital services realm, verifying user identity is a cornerstone of security. Deepfakes and AI-generated documents threaten to weaken those verification checkpoints. For example, attackers often try to register new Microsoft accounts using fake or stolen identities. Their goal might be to obtain free trial resources for spam/scams or establish throwaway tenant accounts to launch attacks. Many of these sign-up attempts use bots—and probably synthetic information (such as random names and AI-generated email addresses)—to get past basic filters. The scale of this activity indicates a systematic attempt to create fake identities at volume.

AI-generated IDs are now often more convincing than real forgeries, growing by 195% globally in usage.[8] In situations where organizations use selfie checks or document uploads to verify new users, deepfake techniques can even defeat liveness tests (for example, a deepfake video can simulate a person blinking and turning their head).

> "
> AI-generated IDs are now often more convincing than real forgeries, growing by 195% globally in usage.

↗ **Learn more**

AI-powered Deception: Emerging Fraud Threats and Countermeasures | Cyber Signals Issue 9

Fraud and social engineering continued

## Virtual credit cards and the shifting fraud landscape

Sitting at the crossroads of convenience, privacy, and security, virtual credit cards (VCCs) are reshaping online payments while simultaneously moving the fraud battleground for merchants.

The global virtual-card market reached USD 19 billion in 2024 and is projected to expand at a robust 21% Compound Annual Growth Rate (CAGR) through 2030 to USD 60 billion.[9] This surge is driven by consumer demand for secure digital payments, the subscription economy, and strong adoption among younger demographics. Generated via apps, VCCs feature unique details and configurable rules (for example, limits, merchant category, and lifespan) designed to reduce card-not-present (CNP) fraud. Businesses, especially in business-to-business (B2B) settings, are also turning to VCCs to improve payment efficiency. However, the swift adoption of VCCs creates new fraud vulnerabilities and operational complexities for merchants, necessitating strategic adaptation.

VCCs often appear like standard cards, making traditional fraud detection rules less effective. Single-use cards also disrupt recurring billing, causing authorization failures.

Subscription abuse and refund fraud are rising as bad actors exploit VCCs' ease of generation and anonymity. The widespread use of synthetic identities adds further complexity, evading common blocklist approaches and creating a whack-a-mole effect for fraud teams.

### Growth in single use credit card sales



354.7

Source: Internal Microsoft commerce telemetry; values indexed to Nov 2023 = 100

### Recommendations

VCCs require a distinct approach to risk management. Their unique qualities demand that merchants treat them as a separate payment type—one that calls for agility, collaboration, and customer-centric design. Microsoft recommends the following strategies to help organizations strengthen defenses while maintaining a smooth customer experience:

Adapt billing models and strengthen payment verification by introducing small validation charges or prepaid options for high-risk customers, or by requiring backup payment methods for larger transactions.

Enhance fraud detection using velocity monitoring and behavioral analytics that go beyond static card data.

Advocate for industry collaboration by encouraging consistent VCC flagging and transparency across card networks.

Monitor for VCC-specific fraud signals such as unusually short card lifespans, single-use patterns, or merchant mismatches

Engage customers proactively with clear messaging about VCC limitations for recurring payments.

## Domain impersonation in the age of AI: Defending against scale and speed

Domain impersonation, or cyber-squatting, involves registering or using domain names with malicious intent to exploit trademarks or deceive users.

Domain impersonation has become one of the fastest-growing online threats due to large-scale, AI-driven attacks. Common motives include extortion, affiliate abuse, phishing, malware distribution, and cyber-smearing.

Fraudsters use the following techniques to create deceptive domain names:

- **Typo-squatting:** Minor spelling errors (for example, "micorsoft.com")
- **Homograph-squatting:** Using visually similar characters (such as "rn" for "m")
- **Combo- and level-squatting:** Adding extra words or subdomains to appear legitimate

Beyond these, cybercriminals use AI-driven adversarial domain generation, such as generative adversarial networks (GAN), to bypass traditional detection in targeted attacks. The GAN's generator learns from real domain datasets, like popular brand URLs, and produces convincing lookalike domains. Meanwhile, the discriminator evaluates their authenticity, refining the output until the fake domains are nearly indistinguishable from real ones. AI automation allows for the rapid creation of thousands of impersonation domains in untargeted attacks, enabling large-scale phishing and scam campaigns in the space of minutes.

Organizations can reduce domain impersonation risk by registering their main domain and common variations and secure their brand presence by verifying official social media accounts and monitoring fake profiles or fraudulent ads. It is also important to educate employees and customers to recognize fake URLs, urgent payment requests, and spoofed emails, and share examples of recent impersonation attempts to raise awareness. Having a rapid response plan that includes takedown procedures with registrars and hosting providers, as well as playbooks for isolating suspicious emails and domains quickly, will also aid in the event of an identified incident.
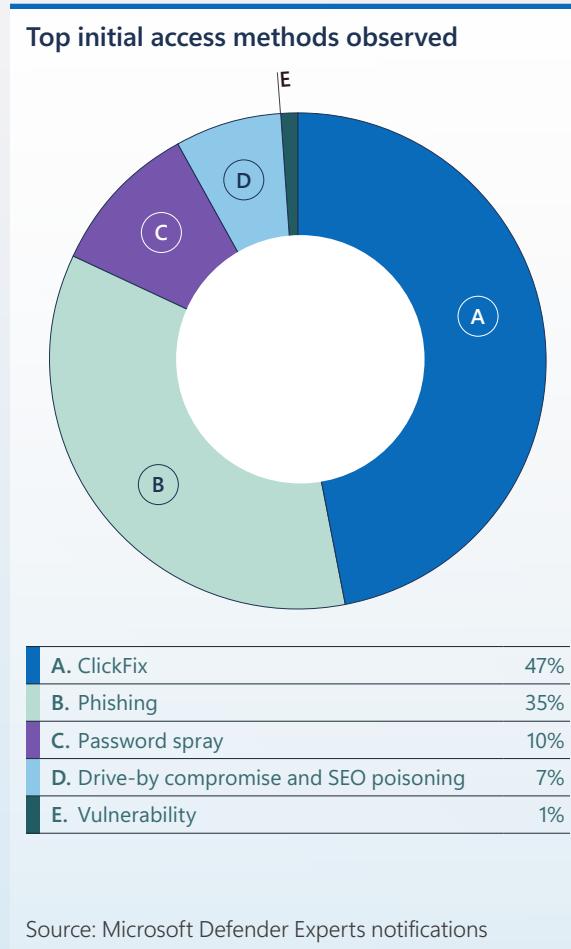
# Social engineering exploits

## The rise of ClickFix

A particularly notable trend beginning in November 2024 was the rapid surge in the use of ClickFix. ClickFix tricks users into copying a command—often embedded in a fake pop-up, job application, or support message—and pasting it into the Windows Run dialog (Win + R) or a terminal, which then executes PowerShell or mshta.exe. These commands pull malicious payloads directly into memory—a clean, fileless process that is often invisible to traditional security tools.

ClickFix was the most common initial access method that Microsoft Defender Experts observed in Defender Expert notifications in the last year, accounting for 47% of attacks. ClickFix has been used by both cybercriminal and nation-state actors to deliver malware, including infostealers, remote access trojans (RATs), and worms. Successful campaigns have led to credential theft, malware staging, and persistent access using just a few keystrokes from the user.

### Top initial access methods observed



| | | |
|---|---|---|
| A. ClickFix | | 47% |
| B. Phishing | | 35% |
| C. Password spray | | 10% |
| D. Drive-by compromise and SEO poisoning | | 7% |
| E. Vulnerability | | 1% |

Source: Microsoft Defender Experts notifications

### Recommendations

Because traditional phishing protections won't catch ClickFix, detection must move beyond static indicators of compromise and focus on behavioral signals. Microsoft recommends implementing the following:

**Awareness training.** Teach users that pasting commands from unknown sources is as risky as clicking suspicious links.

**Script block logging.** Enable PowerShell logging and use Constrained Language Mode to limit abuse.

**Clipboard-to-terminal monitoring.** Watch for unusual clipboard activity followed by shell launches (cmd.exe, powershell.exe).

**Browser hardening.** Disable clipboard access and scripting in untrusted zones.

**Contextual detections.** Correlate clipboard usage with downstream execution patterns to catch suspicious flows.

### ⬈ Learn more

Phishing campaign impersonates Booking.com, delivers a suite of credential-stealing malware | Microsoft Security Blog

Think before you Click(Fix): Analyzing the ClickFix social engineering technique | Microsoft Security Blog

## Phishing landscape

The most significant change in phishing over the last year is the increase in the scale and efficiency of attacks.

AI-automated phishing emails achieved 54% click-through rates compared to 12% for standard attempts—a 4.5x increase. AI enables more targeted phishing and better phishing lures. More concerning, AI automation has the potential to increase phishing profitability by up to 50 times by scaling highly targeted attacks to thousands of targets at minimal cost.[10] This massive return on investment will incentivize cyber threat actors who aren't yet using AI to add it to their toolbox in the future.

### Email bombing as a precursor to social engineering attacks

In 2025, one of the most effective social engineering tactics was email bombing (also called spam bombing or subscription bombing). In email bombing, attackers enroll a target's email account in thousands of newsletters, online services, and so on to flood the target's inbox with hundreds or thousands of subscription emails. This is done to hide critical alerts—for example, MFA prompts, password resets, fraud alerts, or transaction notifications— or to create urgency and confusion.

This year, email bombing evolved from being used as a smokescreen to being used as a first-stage attack vector in a broader malware delivery chain. Email bombing is now often used as a precursor to vishing or Teams-based impersonation, where the attacker contacts the target posing as IT support and offering to resolve the issue. Once trust is established, targets are guided into installing remote access tools, enabling attackers to gain hands-on-keyboard control, deploy malware, and maintain persistence.

### Recommendations

**Filter inbox floods**
Use rules or heuristics to detect mass sign-up emails and alert users or security teams.

**Control Teams exposure**
Restrict external tenant communication and monitor impersonation attempts.

**Educate users**
Make employees aware of fake IT support scams, especially those asking them to run Quick Assist.

**Limit RMM use**
Approve and monitor all remote access tools; block or alert on unauthorized ones through Windows Defender Application Control (WDAC) or AppLocker.

**Correlate behavior**
Flag sequences like inbox flood → Quick Assist → PowerShell/MSHTA execution.

## BEC: A high-impact threat driven by identity compromise

While business email compromise (BEC) represented just 2% of total threats observed over the past year, its impact is disproportionately high, particularly when linked to compromised user accounts. In fact, BEC was a more frequent outcome in attacks (21%) than ransomware (16%), underscoring the need for organizations to defend against both threat types.

BEC attacks are typically initiated through identity compromise. Attackers gain initial access through phishing or password spraying, then pivot to BEC-specific activities such as inbox rule manipulation, unauthorized SharePoint access, internal phishing, email thread hijacking, new MFA authentication method registration, or MFA tampering.
These techniques are used to gain trust, escalate privileges, and ultimately exfiltrate sensitive data or execute financial fraud.

### Business email compromise by sector (January-June 2025)

| Rank | Sector | Percentage |
|---|---|---|
| 1 | Research and academia | 49% |
| 2 | Telecommunications | 11% |
| 3 | Financial services | 7% |
| 4 | Logistics | 6% |
| 5 | Professional services | 5% |
| 6 | Retail and consumer goods | 5% |
| 7 | Power and utilities | 4% |
| 8 | Discrete manufacturing | 3% |
| 9 | Healthcare and public health | 3% |
| 10 | Hospitality and travel | 2% |
| 11 | Insurance | 2% |
| 12 | Nonprofit | 2% |
| 13 | Government | 1% |
| 14 | Manufacturing | 1% |
| 15 | Agriculture, forestry, and fishing | 0.3% |
| 16 | Public safety | 0.2% |

**Social engineering exploits** continued

## Active BEC threat groups

### 1

### Storm-0259

**Operating country:** Türkiye and TRNC

**Nationality:** Nigerian (likely using student visas)

**Active:** 2020 to present

**Tactics:** Use of PhaaS for ATO, Email exfiltration, NameCheap domains, RedVDS for RDP

**Victims:** US, Canadian, UK small and medium businesses

### 2

### Storm-2502

**Operating country:** Nigeria

**Nationality:** Nigerian

**Active:** 2021 to present

**Tactics:** International money laundering, illicit cryptocurrency usage, and US based mule herding

**Victims:** Under assessment

### 3

### Storm-2227

**Operating country:** United Arab Emirates

**Nationality:** Nigerian

**Active:** 2021 to present

**Tactics:** ATO, Email exfiltration, NameSilo/Hostinger domains, Azure/RedVDS for RDP

**Victims:** US construction and architecture

### 4

### Storm-2126
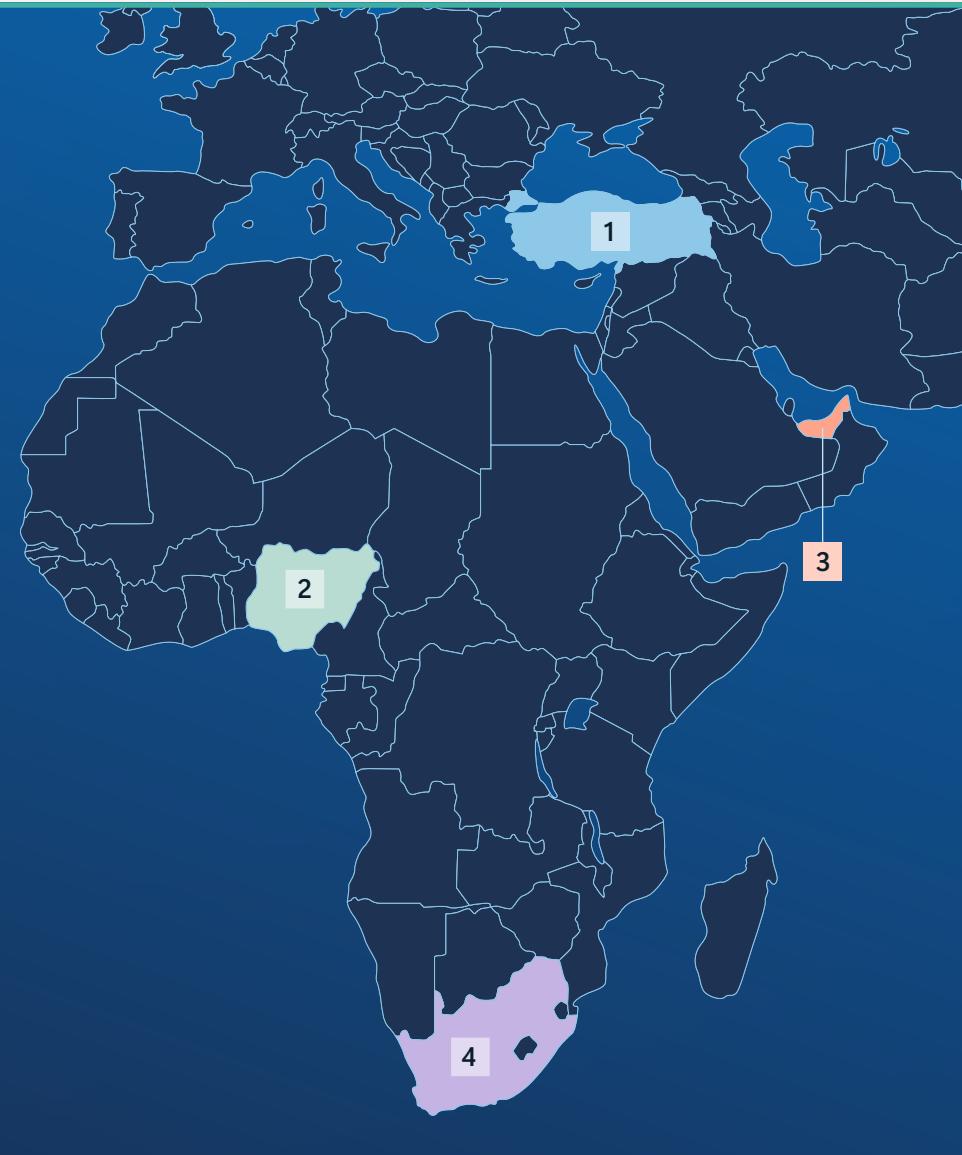
**Operating country:** South Africa

**Nationality:** Nigerian

**Active:** 2017 to present

**Tactics:** Use of ads for phishing, consumer email targeting, GoDaddy domains

**Victims:** US real estate, tile companies and law firms

## Global BEC hotspots

BEC activity is not evenly distributed across the globe. Microsoft telemetry and law enforcement collaboration have identified regional hotspots where BEC operations are particularly active. These areas often serve as hubs for infrastructure setup, money mule recruitment, and laundering operations.

### Recommendations

Correlate identity-related alerts with suspicious mail flow rules, external forwarding, and MFA changes.

Monitor for unusual sending patterns, especially those involving financial requests.

Audit mailbox access and MFA device registrations regularly.

Educate users on how identity compromise fuels BEC—not just phishing.

### ↗ Learn more

## Device code phishing: The next generation of credential theft

This year, Microsoft observed a notable uptick in threat actors conducting device code phishing campaigns worldwide.

In device code phishing, attackers exploit the device code authentication flow to capture access and refresh tokens, which could then be used to access target accounts, data, and other services linked to the compromised accounts. This technique could also enable persistent access or lateral movement as long as the token remains valid.

Threat actors exploit the device code authentication flow by tricking users into entering a device code on seemingly legitimate authentication portals that the actor provided in phishing emails or other communications. Most threat actors first contact victims using third-party messaging applications, at times posing as trusted contacts such as an administrator or program organizer. Once the user enters the code into the portal, the actor is granted access and can capture the access and refresh tokens that are generated.

Threat actors have been particularly successful combining targeted social engineering with out-of-band communications, which allow these actors to circumvent antivirus or other detection systems that would typically identify such activity as spam or phishing.

Device code phishing poses a high risk of data theft and exfiltration, since it grants threat actors access to data where the compromised user has permissions, such as email or cloud storage, without needing a password. In a recent and concerning development, Microsoft observed a threat actor prompting a victim to enter the device code into a Teams invitation, making it harder for users to identify fraudulent activity.

Device code phishing poses a considerable threat to organizations in all sectors worldwide. Microsoft has observed nation-state actors from Russia, Iran, and China as well as cybercriminal groups like Octo Tempest using this technique to access targets in the IT sector, NGOs, government agencies, and private businesses. Ninety-three percent (93%) of the device code phishing events that Microsoft observed in the last twelve months occurred in the second half of the year, indicating the rapid adoption of this technique.

\*

While device code phishing is not new, most users have not been taught to look for attacks that target the device code flow, and because the attacker authentication is through legitimate codes and tokens, traditional phishing detection tools often miss it, making it a particularly dangerous phishing evolution.

↗ **Learn more**

Storm-2372 conducts device code phishing campaign | Microsoft Security Blog

# Cloud threat trends

As organizations accelerate their cloud technology adoption, attackers are increasingly targeting cloud environments, leveraging new tactics and exploiting emerging technologies to compromise assets, disrupt operations, and exfiltrate sensitive data. Understanding these trends is essential for defenders to prioritize protections and respond effectively.

## Cloud under fire: Escalating attacks in cloud environments

Recent telemetry from Microsoft Defender for Cloud highlights a significant escalation in the volume and sophistication of attacks targeting Azure cloud environments. When comparing the first 100 days of 2025 to the second, trends include:
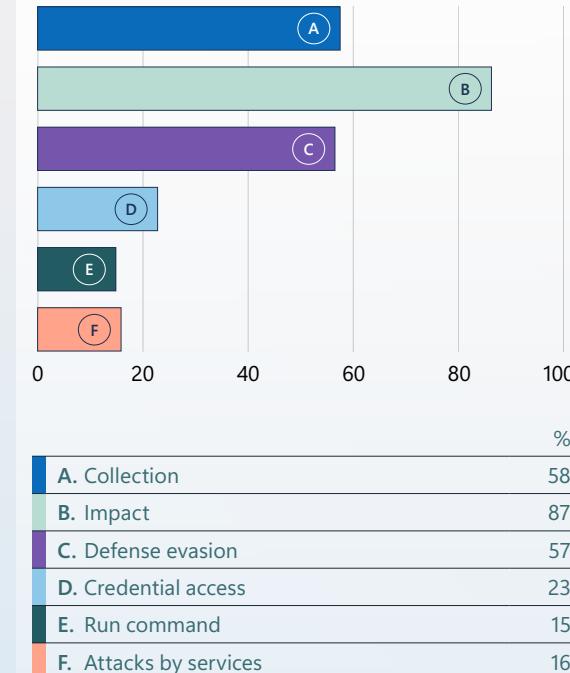
- **A sharp increase in attack volume.** In the number of observed incidents against Azure-based environments, the second 100-day period saw a 26% increase in incidents compared to the first 100 days.

- **A rise in disruptive attack campaigns.** There has been an 87% increase in campaigns aimed at disrupting customer environments through ransomware, mass deletion, or other destructive actions.
- **An escalation in credential theft and data exfiltration attempts.** Credential and access key theft attempts are up 23%. Attempts to extract sensitive data from storage accounts and databases increased by 58%.
- **Improved attacker evasion techniques.** Threat actors demonstrate a growing awareness of cloud defenses and are increasingly employing evasion tactics to bypass detection and mitigation.

Attacks that originate with compromised Entra ID identities and escalate into cloud-based activity within Azure are becoming increasingly prevalent. At the same time, the use of service principals for cloud compromise has remained stable or slightly decreased, potentially reflecting improved hardening efforts in this area. Of note, there is a marked rise in the use of cloud-native mechanisms—such as Azure Run Command—for remote code execution (RCE) within compromised environments.

**Percent increase in some alert notifications this year** (Second 100 days in 2025 compared to first 100 days)



| | % |
|---|---|
| **A.** Collection | 58 |
| **B.** Impact | 87 |
| **C.** Defense evasion | 57 |
| **D.** Credential access | 23 |
| **E.** Run command | 15 |
| **F.** Attacks by services | 16 |

Source: Microsoft Defender for Cloud

**\***

Identity is a primary entry point for cloud attacks, making its protection critical. Enforce MFA and Conditional Access to block unauthorized sign-ins, and use Privileged Identity Management (PIM) with least-privilege principles to tightly control access to sensitive roles.

↗ **Learn more**

Defending against evolving attack techniques | Microsoft Security Blog

**Cloud threat trends** continued

## Container security in focus

A container is a lightweight, standalone, executable package of software that includes everything necessary to run an application.

Containers can be created and taken down quickly, but they introduce unique security challenges in cloud-native environments. Microsoft Defender for Cloud telemetry reveals that container compromise often occurs shortly after deployment.

**Analysis of container runtime and alert timing over 100 days in January-April 2025, surfaced the following conclusions:**

**Most compromised containers are attacked within the first 48 hours of deployment.** This emphasizes the critical need for immediate runtime protection.

**Cryptomining dominates the attack landscape.** Cryptojacking is the most prevalent threat in Kubernetes environments, exhibiting the fastest median time to compromise—less than two days post-deployment.
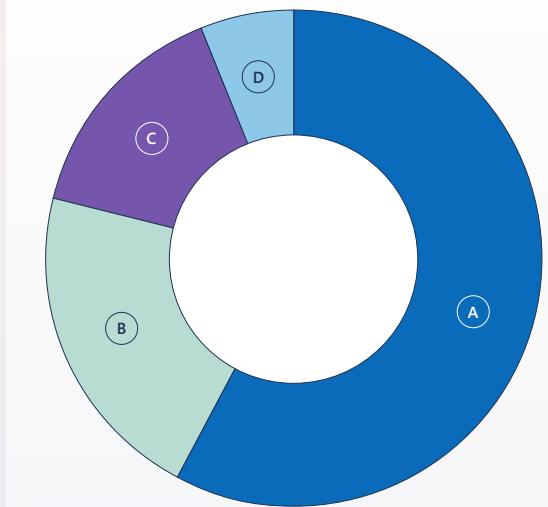
**Credential theft attacks take longer to manifest.** These attacks, the second most common type observed, had the highest median infection time, occurring approximately 3.5 days after container creation.

**Long-tail attacks are a risk.** While most attacks occur early, outliers with significantly delayed infection highlight the importance of sustained monitoring beyond initial deployment.

### Cloud threat infection types
100 days January-April 2025
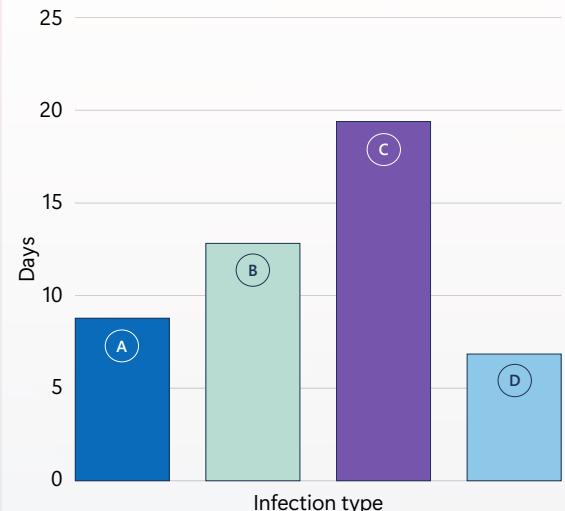


| | % |
|---|---|
| **A.** Crypto miner | 58 |
| **B.** Credential theft | 21 |
| **C.** Known attack tools | 15 |
| **D.** Web shells | 6 |

Source: Microsoft Defender for Cloud

### Median infection time by infection type
100 days in January-April 2025



| | Days |
|---|---|
| **A.** Crypto miner | 8.7 |
| **B.** Credential theft | 12.7 |
| **C.** Known attack tools | 19.3 |
| **D.** Web shells | 6.8 |

Source: Microsoft Defender for Cloud

# Nation-state adversary threats
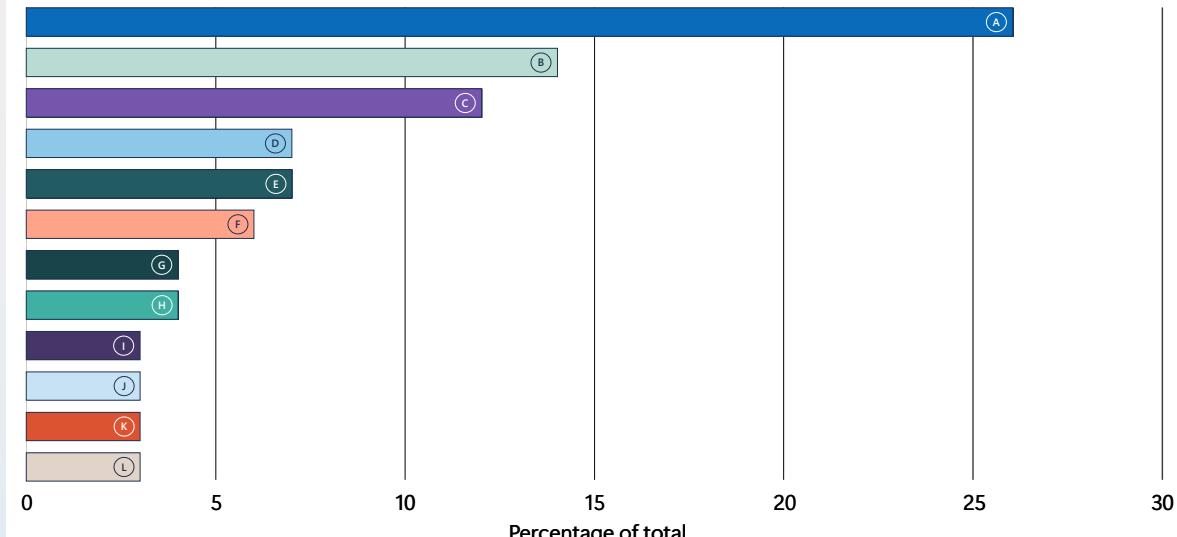
Nation-state cyber activity this year prioritized espionage against traditional intelligence targets—IT, research and academia, government, and think tanks/ NGOs.

A minority of attacks, for example against the Defense Industrial Base, sought to steal proprietary information for economic advantage. An even smaller number of attacks had other goals, including sabotage and ransom.

A major threat that emerged this year was the discovery of the magnitude of North Korea's program to stealthily embed remote workers at organizations around the world. As will be discussed later, this growing threat has multiple facets, including the risk of sanctions violation, espionage, extortion, and sabotage.

In line with geopolitical hotspots and longstanding intelligence priorities, the primary geographical targets of nation-state activity this year were in Israel, the United States, and the United Arab Emirates. Predictably, Ukraine was also an extreme focus for Russian actors.

**Most-targeted sectors by nation-state actors**



Percentage of total

| | % of total | | | % of total |
|---|---|---|---|---|
| **A.** IT | 26 | | **G.** Transportation | 4 |
| **B.** Research and academia | 14 | | **H.** Communications | 4 |
| **C.** Government | 12 | | **I.** Finance | 3 |
| **D.** Think tanks/NGOs | 7 | | **J.** Health | 3 |
| **E.** Consumer retail | 7 | | **K.** Defense | 3 |
| **F.** Manufacturing | 6 | | **L.** Energy | 3 |

Source: Microsoft Threat Intelligence nation-state notification data

# Regional sample of nation-state activity levels observed

**Observed event activity count per country**
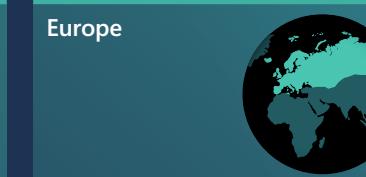
Over 200 events

100—200

50—100

0—50

## Americas

### Top activity levels

| | |
|---|---|
| United States | 623 |
| Canada | 51 |
| Brazil | 24 |
| Peru | 16 |
| Argentina | 11 |
| Colombia | 10 |
| Mexico | 9 |
| Dominican Republic | 5 |
| Chile | 4 |
| Costa Rica | 3 |

## Asia & Pacific

### Top activity levels

| | |
|---|---|
| Taiwan | 143 |
| Korea | 126 |
| India | 100 |
| Hong Kong SAR | 95 |
| China | 49 |
| Australia | 47 |
| Thailand | 39 |
| Japan | 38 |
| Singapore | 33 |
| Indonesia | 32 |

## Europe

### Top activity levels

| | |
|---|---|
| Ukraine | 277 |
| United Kingdom | 144 |
| Poland | 97 |
| Germany | 74 |
| France | 72 |
| Spain | 61 |
| Russia | 60 |
| Italy | 51 |
| Azerbaijan | 35 |
| Belgium | 30 |

## Middle East & Africa

### Top activity levels

| | | | |
|---|---|---|---|
| Israel | 603 | Kenya | 9 |
| United Arab Emirates | 166 | Nigeria | 8 |
| Saudi Arabia | 70 | Tanzania | 5 |
| Türkiye | 70 | Mali | 4 |
| Iraq | 67 | Namibia | 4 |
| Jordan | 44 | Botswana | 2 |
| Lebanon | 39 | | |
| Egypt | 32 | | |
| Iran | 27 | | |
| Morocco | 26 | | |
| South Africa | 31 | | |
| Ethiopia | 20 | | |
| Angola | 9 | | |

**Nation-state adversary threats** continued

# China
## Global espionage at scale

The breadth and scale of Chinese targeting operations continue to stand out from other nation-state actors.

In line with their emphasis on espionage and the collection of proprietary information, Chinese actors have primarily targeted organizations in the IT sector, internet service providers (ISPs) and telecommunications, government agencies, military and defense, and NGOs. Chinese threat actors' targets mostly reside in the United States, Asia, North Africa, and Latin America.

China uses espionage operations as a key method of pursuing economic competitive advantage. While state-sponsored actors continue to conduct operations based on this primary objective, these tactics and operations often rely on unexpected partnerships.
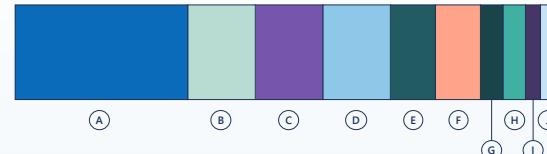
Chinese state actors increasingly rely on partnerships with public/non-government organizations to conduct vulnerability research, create custom malware, or provide covert networks to obfuscate operations. This behavior reflects China's longstanding focus on operational

security, customization of tradecraft, and obfuscation of their espionage operations.

Chinese threat actors regularly refine their techniques as they adapt to advancements in security and defensive measures. For example, they are increasingly using covert networks to avoid detection and are focused on targeting vulnerable internet-facing devices. Because these devices are often less protected and integrated within an organization's security programs, they offer both an entry point and an additional layer of obfuscation for further attacks.In recent years, Chinese actors have become faster at operationalizing newly disclosed vulnerabilities, a threat compounded by the growing complexity of digital supply chains, which introduces more components for exploitation.

Throughout 2024, a year with a record number of elections worldwide, Chinese actors spent significant effort collecting intelligence or attempting to influence their outcomes. Through coordinated influence operations campaigns, and cyber intrusions, China seeks to undermine democratic institutions, sow discord among allies, and promote narratives that legitimize its governance model. This push reflects and long-term ambition to reshape the international order, elevate China's geopolitical standing, and counter Western influence in key regions.

## Ten sectors most targeted by Chinese threat actors
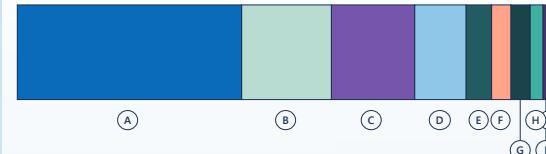


| | % of total |
|---|---|
| **A.** IT | 23% |
| **B.** Government | 10% |
| **C.** Think tanks/NGOs | 9% |
| **D.** Manufacturing | 9% |
| **E.** Research and academia | 6% |
| **F.** Consumer retail | 6% |
| **G.** Communications | 3% |
| **H.** Finance | 3% |
| **I.** Transportation | 2% |
| **J.** Health | 2% |

Chinese nation-state threat actors focused on IT, government, and think tanks or NGOs to support China's goal of reshaping the international order, elevating China's geopolitical standing, and countering Western influence in key regions.

Source: Microsoft Threat Intelligence nation-state notification data

## Ten regions most targeted by Chinese threat actors



| | % of total |
|---|---|
| **A.** United States | 35% |
| **B.** Thailand | 14% |
| **C.** Taiwan | 12% |
| **D.** Korea | 8% |
| **E.** Japan | 4% |
| **F.** Philippines | 3% |
| **G.** United Kingdom | 3% |
| **H.** India | 2% |
| **I.** Germany | 1% |
| **J.** Hong Kong SAR | 1% |

While Chinese actors have persistently targeted the United States, this year they demonstrated an elevated focus on Thailand, reflecting a strategic expansion of influence efforts in Southeast Asia.

Source: Microsoft Threat Intelligence nation-state notification data
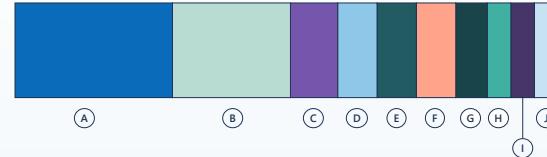
**Nation-state adversary threats** continued

# Iran
## Persistent and adaptive

In a year where Iran was impacted by conflict, Iranian state actors continued to direct broad campaigns against historic adversaries, targeting organizations and individuals across the Middle East, Europe, and North America.

The volume of Iranian state-linked cyber activity remains consistently high, with persistent campaigns observed across diverse industries. In late 2024, some national security agencies warned about a surge in Iranian nation-state credential harvesting attacks, targeting the healthcare, government, IT, energy, and engineering sectors, reflecting a continuation of Iran's historically consistent focus on critical infrastructure. At the same time, Iran's intelligence services continue to focus heavily on regional adversaries, conducting long-term espionage against critical infrastructure.

Microsoft has observed increased overlap in tactics, techniques, and procedures (TTPs) among certain Iranian state actors, suggesting possible formal or informal collaboration, including shared resources or personnel. This convergence could also reflect centralized strategic direction, shared development pipelines, third-party contractor support, or deliberate efforts to obscure attribution. In the last year, three Iranian actors targeted shipping and logistics operations across Europe and the Persian Gulf in sophisticated campaigns. These compromises indicate an intent to gain long-term access to operational systems and sensitive commercial data. Access to maritime companies and data raises concerns given it could potentially enable espionage or interference with commercial shipping operations.

A growing and significant trend across a few Iranian threat actors is the abuse of cloud infrastructure, particularly Microsoft Azure, for command and control, persistence, email exfiltration, and other malicious activities, often using fraudulently created or compromised subscriptions. By abusing subscription models such as Azure for Students and trial accounts within compromised tenants, threat actors create low-cost, disposable infrastructure that is difficult to detect and trace.

## Ten sectors most targeted by Iranian threat actors



| | % of total |
|---|---|
| **A. IT** | 21% |
| **B. Research and academia** | 15% |
| **C. Government** | 8% |
| **D. Transportation** | 6% |
| **E. Consumer retail** | 5% |
| **F. Communications** | 5% |
| **G. Commercial facilities** | 3% |
| **H. Manufacturing** | 3% |
| **I. Think tanks / NGOs** | 3% |
| **J. Defense industry** | 2% |

Iran's focus on the IT sector stemmed from the sector's utility in espionage, influence, and disruption. By compromising IT providers, Iranian nation-state threat actors gained access to sensitive data, trusted communications, and a pathway into multiple downstream sectors simultaneously.

Source: Microsoft Threat Intelligence nation-state notification data

## Ten regions most targeted by Iranian threat actors



| | % of total |
|---|---|
| **A. Israel** | 64% |
| **B. United States** | 6% |
| **C. United Arab Emirates** | 5% |
| **D. India** | 2% |
| **E. Greece** | 2% |
| **F. Azerbaijan** | 2% |
| **G. Saudi Arabia** | 2% |
| **H. United Kingdom** | 1% |
| **I. Türkiye** | 1% |
| **J. Iraq** | 1% |

Iran views Israel as its top regional rival. Targeting Israel through cyber operations enabled Iran to gather intelligence, disrupt critical services, retaliate below the level of open war, and project ideological resistance for domestic and regional audiences.

Source: Microsoft Threat Intelligence nation-state notification data

# Russia

## Expanding its target set but still focused on Ukraine

Russian state actors expanded the scope of their targeting this year to infiltrate networks and devices primarily in Ukraine and North Atlantic Treaty Organization (NATO) member states.

This shift to a broader target set—while maintaining the same geographical focus—has put more organizations at risk of compromise, although outside of Ukraine that risk is almost exclusively for cyber espionage.

For example, we have observed a modest increase in Russian actors targeting smaller businesses in countries supporting Ukraine. This is an expansion of these actors' scope, which previously had been mostly limited to conventional political targets like government agencies. Russian state actors might also view these smaller targets of opportunity as less resource-intensive pivot points they can use to access larger organizations.

On the technical front, Russian state actors are pursuing different approaches to achieve their goals. This year, we observed nation-state actors outsourcing pre- or post-compromise operations

and continuing to co-opt cybercriminal or other nation-state infrastructure.
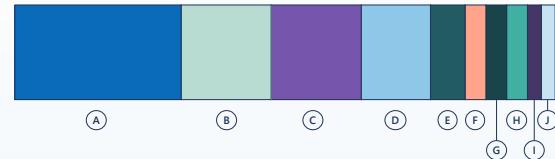
These actors appear to have reduced their efforts to develop bespoke operations in favor of leveraging the cybercriminal ecosystem. This growing reliance on less sophisticated methods and commodity tools is likely a response to exposure by government agencies and cybersecurity firms of their tools and techniques. This shift in TTPs could make it more difficult for network defenders to attribute simple operations to sophisticated threat actors and recognize the implications of a breach. At the same time, it highlights the need to defend against known Russian TTPs.

Microsoft separately tracks notifications related to Ukraine. This data shows Ukraine accounted for 25% of Russia's cyber operations, making it the primary target. These actors appear to have reduced their efforts to develop bespoke operations in favor of leveraging the cybercriminal ecosystem.

> " These actors appear to have reduced their efforts to develop bespoke operations in favor of leveraging the cybercriminal ecosystem.

## Ten sectors most targeted by Russian threat actors

| | % of total |
|---|---|
| **A.** Government | 25% |
| **B.** Research and academia | 13% |
| **C.** Think tanks/NGOs | 13% |
| **D.** IT | 10% |
| **E.** Energy | 5% |
| **F.** Defense industry | 3% |
| **G.** Manufacturing | 3% |
| **H.** Transportation | 3% |
| **I.** Finance | 2% |
| **J.** Inter-governmental organizations | 2% |

Russian nation-state actors focused on government organizations and think tanks or NGOs in Europe and North America, reflecting their intelligence value to Russia amid the ongoing war.

Source: Microsoft Threat Intelligence nation-state notification data

## Ten regions most targeted by Russian threat actors

| | % of total |
|---|---|
| **A.** United States | 20% |
| **B.** United Kingdom | 12% |
| **C.** Ukraine | 11% |
| **D.** Germany | 6% |
| **E.** Belgium | 5% |
| **F.** Italy | 3% |
| **G.** Estonia | 3% |
| **H.** France | 3% |
| **I.** Netherlands | 3% |
| **J.** Poland | 3% |

Outside of Ukraine, the top ten countries most affected by Russian cyber activity all belong to NATO—a 25% increase compared to last year. Although Ukraine appears in third place in our nation-state notification system, Microsoft's dedicated tracking for Ukraine reveals it was the primary focus of Russian state actors.

Source: Microsoft Threat Intelligence nation-state notification data

# North Korea

### Revenue generation and remote workers

North Korean state actors remain a persistent threat to a narrow target set, with a few exceptions pursuing the same sectors and geographies using the same TTPs year over year.
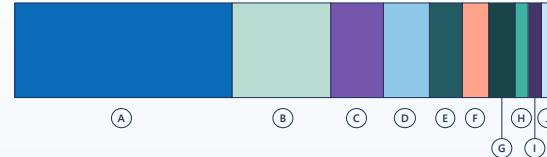
A major issue this year was the North Korean IT worker problem. For over a decade, North Korea has remotely stealthily embedded tens of thousands of workers at organizations around the world in a trend that is quicky accelerating. As discussed below, this growing army of workers remits hundreds of millions of dollars a year to North Korea. When discovered, some of these workers have turned to extortion, another approach to bringing in money for the regime. They could also use their emplacement for the delivery of malware like ransomware. We discuss this issue more in-depth in the insider threats section of this report.

Globally, North Korean threat actors are largely focused on the IT sector, any organization or asset associated with banking or blockchain technology, defense, and manufacturing. In addition, any entity that has a nexus with East Asian policy, from NGOs and universities to ministries of foreign affairs, is a priority target. In the Asia-Pacific (APAC) region specifically, North Korean threat actors are interested in heavy manufacturing and a broad spectrum of organizations in South Korea.

This year, as North Korean state actors pursued an even more aggressive approach to revenue generation, they doubled down on traditional avenues such as cryptocurrency theft and ransomware. Microsoft Threat Intelligence observed a North Korean actor participating as a RaaS affiliate for the first time. A pivot to RaaS participation could lead to more ransomware attacks as North Korea outsources parts of the ransomware cycle, freeing up resources to focus on compromising targets. Microsoft also observed an increase in phishing operations to collect IP associated with weapons systems.

This year, Microsoft observed at least a few using cloud infrastructure to conceal their C2 infrastructure, an increase in their sophistication that will make it harder for defenders to detect and block attacks. While this is still a nascent trend, it may be an indicator of North Korean state actors exploring new ways to evade defenders.

## Ten sectors most targeted by North Korean threat actors
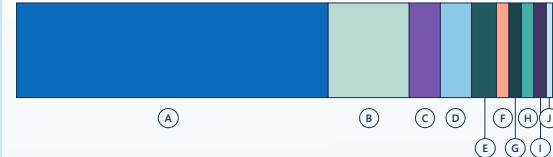
|  | % of total |
|---|---|
| **A.** IT | 33% |
| **B.** Research and academia | 15% |
| **C.** Think tanks/NGOs | 8% |
| **D.** Consumer retail | 7% |
| **E.** Finance | 5% |
| **F.** Manufacturing | 4% |
| **G.** Health | 4% |
| **H.** Communications | 2% |
| **I.** Defense industry | 2% |
| **J.** Commercial facilities | 2% |

North Korean nation-state threat actors focused primarily on organizations with access to blockchain technology or cryptocurrency and sources of East Asian policy, reflecting these actors' mandates for revenue generation and intelligence collection.

Source: Microsoft Threat Intelligence nation-state notification data

## Ten regions most targeted by North Korean threat actors

|  | % of total |
|---|---|
| **A.** United States | 50% |
| **B.** Italy | 13% |
| **C.** Australia | 5% |
| **D.** United Kingdom | 4% |
| **E.** Switzerland | 2% |
| **F.** India | 2% |
| **G.** Germany | 2% |
| **H.** United Arab Emirates | 2% |
| **I.** France | 1% |
| **J.** South Korea | 1% |

The United States ranks first in our nation-state notification system for North Korea due to the high volume of remote IT worker activity targeting US-based companies. These workers primarily pursue roles at US companies because they often offer the highest salaries.

Source: Microsoft Threat Intelligence nation-state notification data

## Nation-state abuse of AI in influence operations: Emerging tactics and strategic implications

Nation-state actors continue to evolve their cyber and influence operations with the rapid adoption of AI, employing more advanced, targeted, and scalable tactics. This year, the Microsoft Threat Analysis Center observed several new trends shaping the landscape of AI-enabled operations:

- **AI twinning:** the creation of digital replicas of trusted news anchors that deliver state-backed narratives with a veneer of credibility.
- **Training data poisoning:** the attempt to deliberately insert biased, misleading, or manipulative content into the datasets that inform AI models, with the aim of influencing model behavior and output.
- **Voice cloning and masking:** the use of generative AI audio and visual tools to impersonate individuals in ways that skirt legal thresholds but challenge ethical norms.

The objectives remain consistent: to manipulate public perception and shape conflict narratives. The integration of AI tools with conventional cyber techniques—such as phishing, credential harvesting, and insider recruitment—has made these operations easier to scale, more effective, and harder to trace. Attribution will become increasingly challenging as AI blurs the line between state-linked and opportunistic influence campaigns.

### Strategic implications

A critical change, however, is the emergence of AI-first actors—entities that prioritize AI-generated content and tools over traditional methods and manipulations. These actors are shifting from spectacle to saturation, flooding the information space with synthetic media to desensitize audiences and exhaust detection systems. In some cases, they appear to operate semi-independently, drawing from state-aligned narratives while relying on AI to maintain volume, speed, and plausible deniability.

The shift carries strategic implications. The convergence of AI and cyber operations enables persistent, low-cost, and scalable influence campaigns. Policymakers and defenders must adapt accordingly—rethinking attribution models, updating content authentication standards, and preparing for influence operations where AI is not just a tool, but the core strategy.

✴
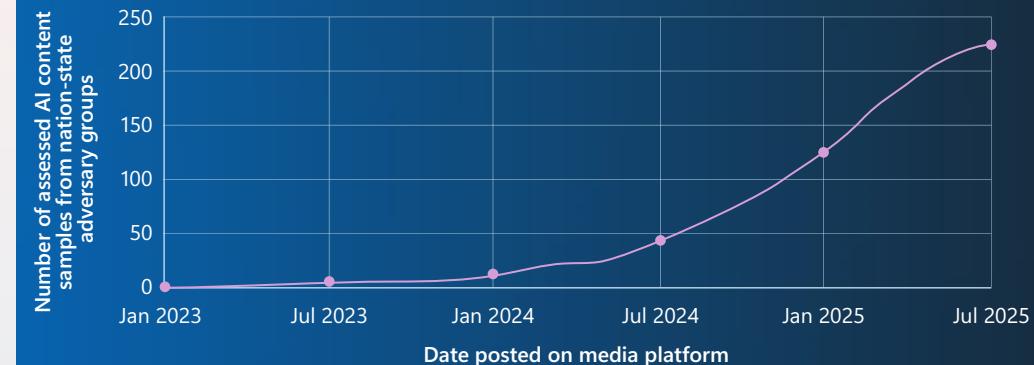
In the last six months, AI in influence operations has picked up aggressively

↗ **Learn more**
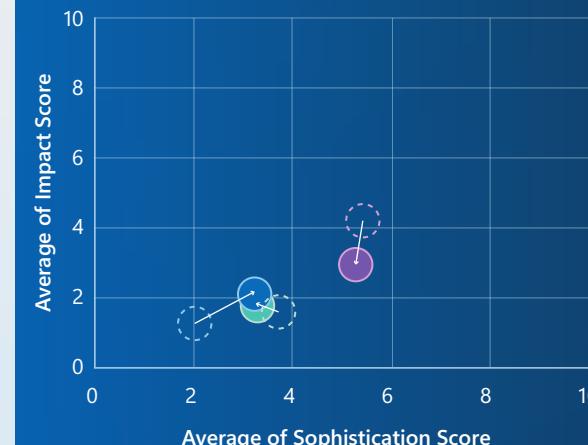
5 things you need to know about tracking today's nation-state threats



**Rapid growth in assessed AI content samples attributed to nation-state adversaries**

Source: Microsoft Threat Intelligence



**Overall sophistication impact by country**

| July-Dec 2024 | Jan-June 2025 | |
|---|---|---|
| China | China | 🟢 |
| Iran | Iran | 🔵 |
| Russia | Russia | 🟣 |

Microsoft evaluates AI-generated content from nation-state adversaries using a structured impact framework—assessing the potential stakes of the content, its reach, and persistence across media platforms and audiences.

Source: Microsoft Threat Intelligence

## Insider risk in the age of strategic geopolitical competition

### Insider threats: Emerging dimensions and mitigations

In an era of increasing geopolitical tensions and blurring of public and private sector interests, nation-state actors have increased their use of insiders to gain access to intelligence. These efforts are often long-term operations that are more difficult to detect than traditional hacks. Nation-states increasingly use non-state actors—including cyber mercenaries, criminal syndicates, and front organizations—to conduct insider threat operations that target private sector entities.[11] These proxies obscure attribution while enabling scalable, persistent campaigns.

China and Russia have both cultivated ecosystems to infiltrate corporate environments, often using academic or professional affiliations to identify and exploit vulnerable insiders.[12] The sectors most at risk—AI, quantum technologies, biotechnology, and defense—have both economic and military value. Insider espionage can cause immediate financial loss and long-term competitive harm, erasing years of innovation and market advantage through stolen research and development.

Most organizations' cybersecurity frameworks were not initially built with an insider threat in mind. Compliance standards and cybersecurity best practices traditionally assume that the attacker is an outsider trying to break in, but when the threat actor is an insider with valid access, many of those measures could be bypassed by default.

Additionally, many internal cybersecurity tools are not designed to detect trusted insiders working covertly with sophisticated external actors. For example, data loss prevention (DLP) tools that would flag large, suspicious file transfers often miss the slow, stealthy exfiltration of an espionage-minded insider. While zero trust network architecture adds protection against unauthorized devices and external connections, it requires consistent operationalization on the comprehensive zero trust principles and security strategy to prevent unauthorized use of a legitimate user account.

According to DTEX Systems and the Ponemon Institute, companies take 81 days on average to contain an identified insider incident.[13] This long dwell time gives nation-state actors a persistent foothold to expand their access, cover their tracks, and even establish back doors for future use.

Layoffs and workforce reductions across government and private industry add another dimension to the insider landscape. Such workforce adjustments can inadvertently exacerbate insider threat risks through disgruntled employees or weakened security oversight due to budget cuts and staff reductions. Malicious insiders can leak sensitive data or redirect corporate assets to corporate adversaries. Third-party suppliers with privileged access might unknowingly introduce vulnerabilities, making rigorous vetting and alignment with internal security policies essential to mitigating insider-driven exposure. Facing this threat requires an intentional strategy. For businesses, the issue of insider risk should be elevated to the boardroom and C-suite. Executives should include insider risk in regular risk assessments and incorporate insider risk programs information when business decisions may impact the workforce. Key recommendations for enterprise leaders include:

- **Identify your crown jewels.** Pinpoint the data or technologies that would be most devastating to lose (for example, trade secrets, source code, formulas, merger and acquisition plans) and implement extra safeguards around these assets such as strict need-to-know access, encryption, and monitoring of access logs in real time.
- **Implement continuous identity verification.** Move beyond one-time sign-ins and use adaptive authentication and behavioral biometrics (like typing patterns or mouse movements) to continuously verify that the person behind an account is the genuine user. If an account starts behaving oddly—for example, a finance employee begins downloading large engineering design files—require immediate re-authentication or manager approval.
- **Divide and limit access.** Architect your systems on the assumption that an insider might turn malicious. No single individual should be able to access all critical data. Use segregation of duties and data fragmentation so that even if one account is compromised, an attacker can't sweep up everything.
- **Foster a vigilant culture.** Employees are often the first to notice unusual behavior in a peer. Create a culture where reporting a concern is encouraged and rewarded.
- **Conduct exit interviews and post-employment monitoring.** Exit interviews are an effective safeguard against insider risk. They provide a final opportunity to detect warning signs, reinforce confidentiality and data protection obligations, and ensure access to sensitive information is revoked. These conversations also reduce the risk of disgruntled retaliation, highlight potential process weaknesses, and remind departing staff of their continuing obligations at a time when adversarial entities may seek to recruit them (this is more specific to those with security clearances). Documenting the exchange creates an audit trail, demonstrating that the organization has taken prudent steps to protect its assets, reputation, and people during workforce transitions.
- **Engage in holistic insider risk management.** Effective insider risk management requires a blend of technology, culture, and collaboration. Deploy behavioral analytics and DLP solutions to detect unusual data transfers or privilege escalations, particularly among highly privileged users. Intelligence sharing between government, private industry, and recruiting platforms also helps expose fake companies and protect organizations from risky potential hires.

Additionally, companies can use dedicated insider threat monitoring tools to reduce the overarching risk profile. As an example, companies utilizing Communications Compliance can be notified of potential talent recruitment outreach.[14]

## Detecting North Korean IT workers

North Korea has quietly built a large remote employment staffing apparatus that has emplaced thousands of workers at unwitting companies globally. These state sponsored workers, who are physically located either in North Korea or abroad, submit tens of thousands of job applications a month for software, web development, and other technology/IT positions. This year, we also saw these workers branching into other job types, such as structural engineering. Because these workers opportunistically apply to remote job postings, they represent a threat to organizations anywhere in the world, in any sector.

To help organizations identify potential North Korean state sponsored remote workers, we recommend the following employment vetting recommendations. For a more extensive discussion, see our blog on Jasper Sleet.[15]

**During the pre-hire stage:**
- Check resumes for consistency of names, addresses, educational history, and job titles. Consider contacting references by phone or video teleconference.
- Confirm that the applicant doesn't have multiple social media accounts under different names.
- Scrutinize staffing company employees, since this is a primary avenue for North Korean state sponsored workers to land jobs.
- Ensure that the applicant is seen on camera during multiple video telecommunication sessions.
- Confirm that the applicant's contact information includes a real phone number, not Voice over Internet Protocol (VoIP), and a residential address.
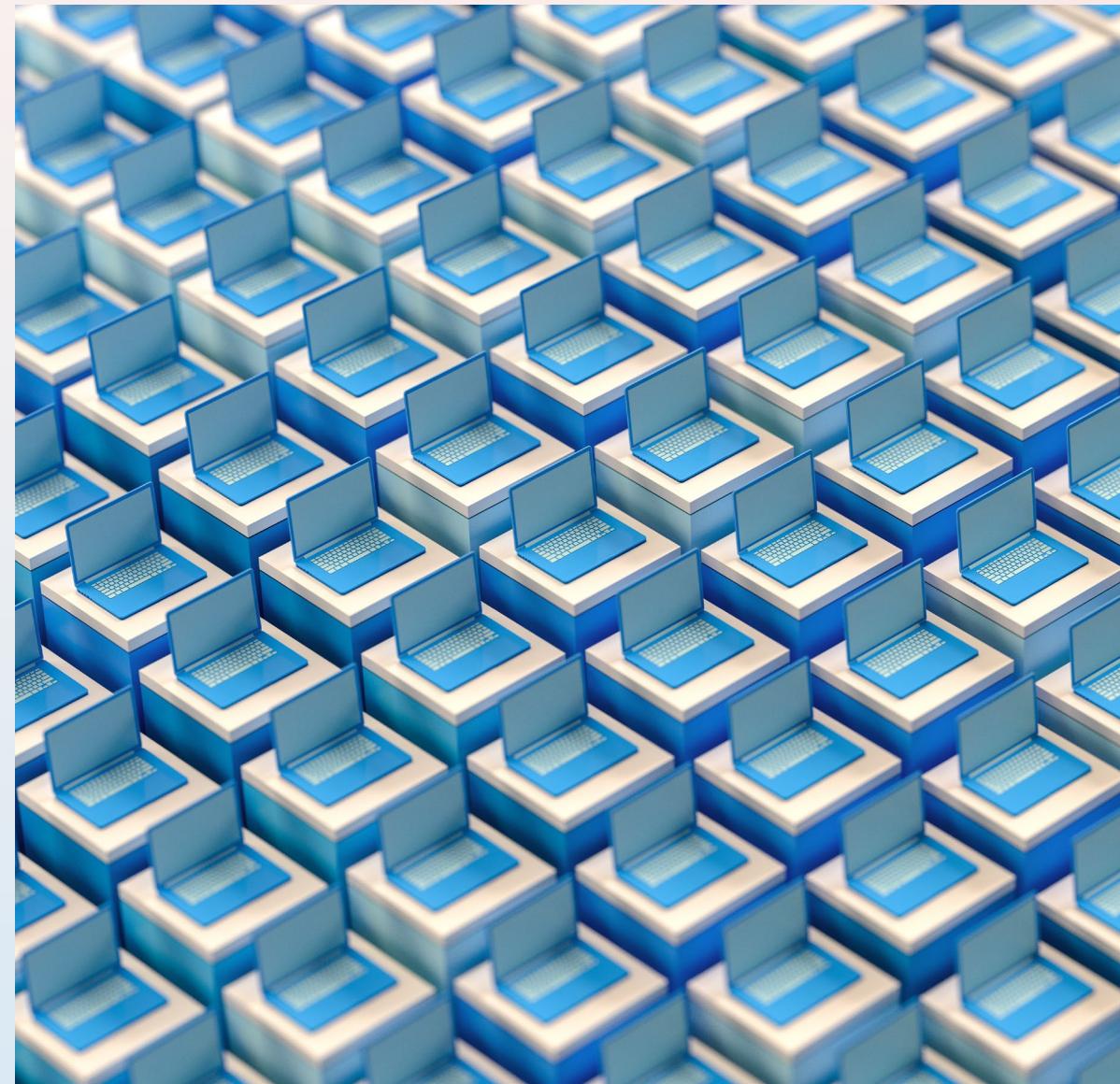
**Once hired, employees should be monitored for the following:**
- Installation of unauthorized software such as RMM tools and virtual private networks (VPN), especially Astrill VPN.
- Geographical irregularities— for example, a supposedly United States-based employee signs in from an IP address associated with China, or the employee device engages in impossible travel, in which the IP address location changes faster than it would be possible for the employee to travel between those locations.
- Camera avoidance—the employee creates excuses for why they are never seen on camera.

In addition to technical monitoring, organizations can also use simple, non-technical identity verification techniques such as asking employees to turn on their camera periodically and comparing the person on camera with the one that took delivery of the corporate laptop.

# AI's double-edged influence:
## Defending and disrupting the digital landscape

The AI threat landscape is diverse and rapidly evolving. The distinctive nature of AI-related threats demands that organizations develop new strategies and adaptive approaches to effectively manage emerging risks.

For example, as AI adoption accelerates, so does AI's access to sensitive data. Whether through user-supplied inputs, credentialed access to existing content, or the creation of custom fine-tuned models built on proprietary data, the volume and sensitivity of data involved continue to grow—which means risks associated with the compromise of or unauthorized access to that data are also growing.
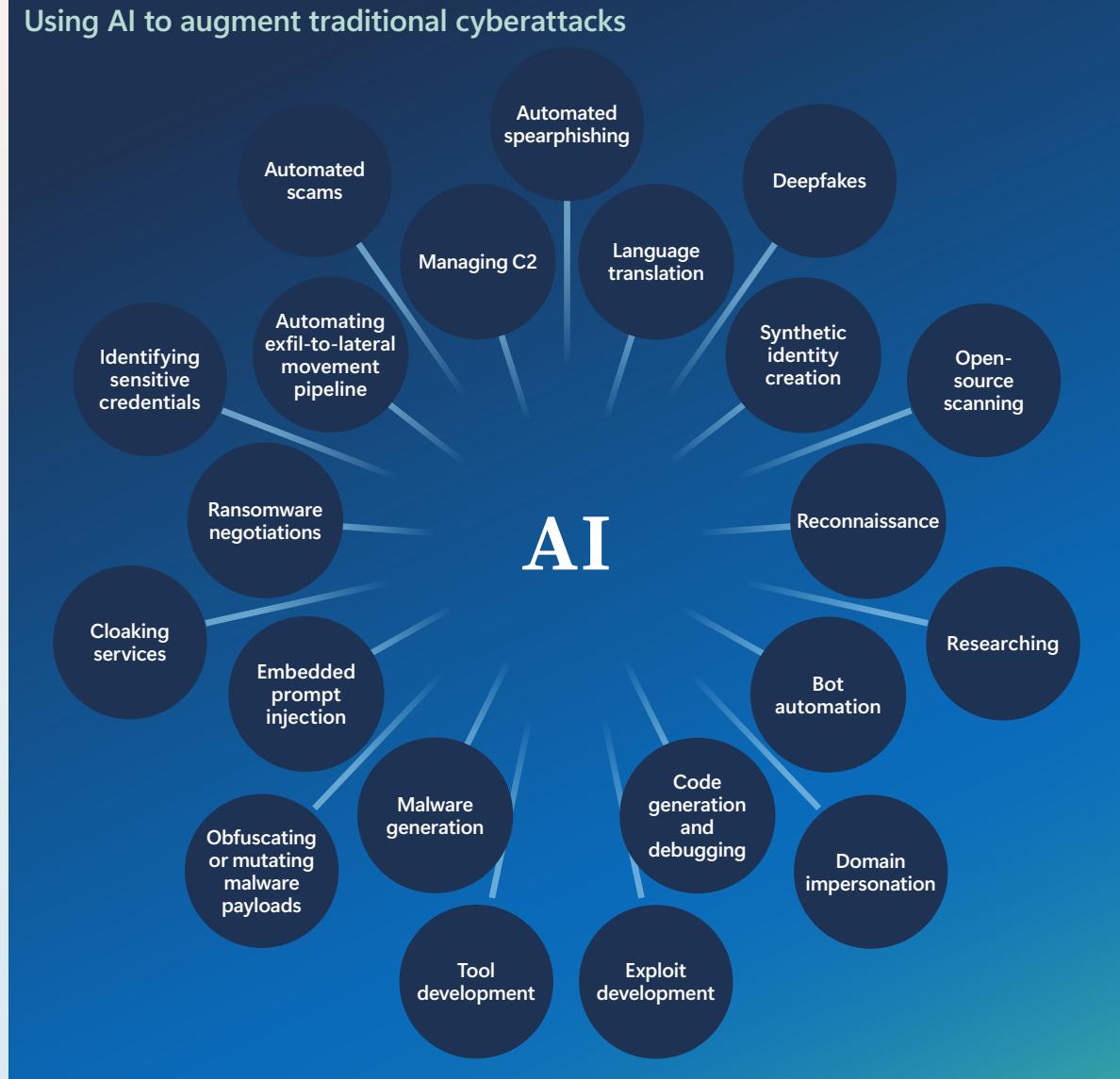
AI-associated challenges include both threats to AI and its users and threats enabled by AI. AI-associated threats can be divided into five major categories: traditional cybersecurity, malfunction, dangerous capabilities, operational issues, and emerging threats.

### Traditional cybersecurity

This category encompasses both cyberattacks that are amplified using AI and direct attacks on AI systems. These threats target underlying infrastructure and exploit human vulnerabilities. Actors conducting these attacks range from less-skilled individuals to sophisticated state-sponsored groups.

**Cyberattack augmentation** refers to the use of AI to enhance traditional cyberattacks. The chart on the right highlights the primary areas of augmentation, most of which are based in the automation of previously time-intensive activities.

Defenders must counter AI augmentation by fostering a strong cybersecurity culture, training users to recognize manipulation tactics, and implementing authenticated communication channels. AI-driven detection systems that flag anomalies in communication patterns or identify deepfake content in real time can also serve as critical safeguards, while AI can detect vulnerabilities, automate patching, and improve threat intelligence.



Using AI to augment traditional cyberattacks

AI

- Automated spearphishing
- Automated scams
- Managing C2
- Language translation
- Deepfakes
- Synthetic identity creation
- Open-source scanning
- Automating exfil-to-lateral movement pipeline
- Identifying sensitive credentials
- Reconnaissance
- Ransomware negotiations
- Researching
- Cloaking services
- Bot automation
- Embedded prompt injection
- Code generation and debugging
- Domain impersonation
- Obfuscating or mutating malware payloads
- Malware generation
- Tool development
- Exploit development

**AI's double-edged influence: Defending and disrupting the digital landscape** continued

## Generative AI threat map

**AI usage security**

### User interaction with generative AI-based apps

**Sensitive information disclosure**

**Shadow IT/harmful third-party LLM-based app or plugin**

### Generative AI-based extended risks

**AI insider risk, excessive agency and overreliance**

**AI application security**

### Generative AI-based app lifecycle

**Prompt injection UPIA/XPIA**

**Data leak, exfiltration**

**Insecure plugin design**

**AI platform security**

### Fundamental model and training data

**Training data poisoning**

**Model theft and model poisoning**

Use of generative AI has introduced new layers of complexity to the threat landscape. The diagram on the left illustrates how risks span across usage, application, and platform levels—highlighting issues such as sensitive data exposure, prompt injection, insecure plugin design, and foundational threats like model theft and training data poisoning.

**Indirect prompt injection** attacks are particularly concerning for developers and organizations that rely on large language models (LLMs) to process untrusted or user-generated content. In these attacks, malicious instructions are embedded in seemingly benign data—such as a resume containing hidden text that instructs the AI to favor a candidate. If the AI is trusted to act autonomously, it might execute these hidden commands, leading to biased decisions, unauthorized outputs, or even system compromise. Defending against these attacks requires both technical tools—like filters that detect hidden or malicious text—and strong coordination across teams. Developers, security experts, and decision-makers must work together to ensure protections are built, tested, and enforced consistently.

**Model theft** involves the unauthorized replication of an AI system's architecture, behavior, or training data. This can be a result of corporate or nation-state espionage, especially when the stolen model is used to develop competing technologies. Mitigation strategies include access controls, encryption, threat monitoring, secure development practices, and coordinated response plans—shared responsibilities among developers, hosts, and regulators.

**Cyberattack automation** refers to the use of AI to enhance traditional cyberattacks. Threat actors can now automate vulnerability discovery, malware generation, and data analysis. In response, defenders are also leveraging AI to detect vulnerabilities, automate patching, and improve threat intelligence.

In the realm of **social engineering**, AI can automate phishing campaigns, generate deepfakes, and craft highly convincing fraudulent messages. Defenders must counter this AI augmentation by fostering a strong cybersecurity culture, training users to recognize manipulation tactics, and implementing authenticated communication channels. AI-driven detection systems that flag anomalies in communication patterns or identify deepfake content in real time can also serve as critical safeguards.

**AI's double-edged influence: Defending and disrupting the digital landscape** continued

## Adversarial exploitation of inherent risks
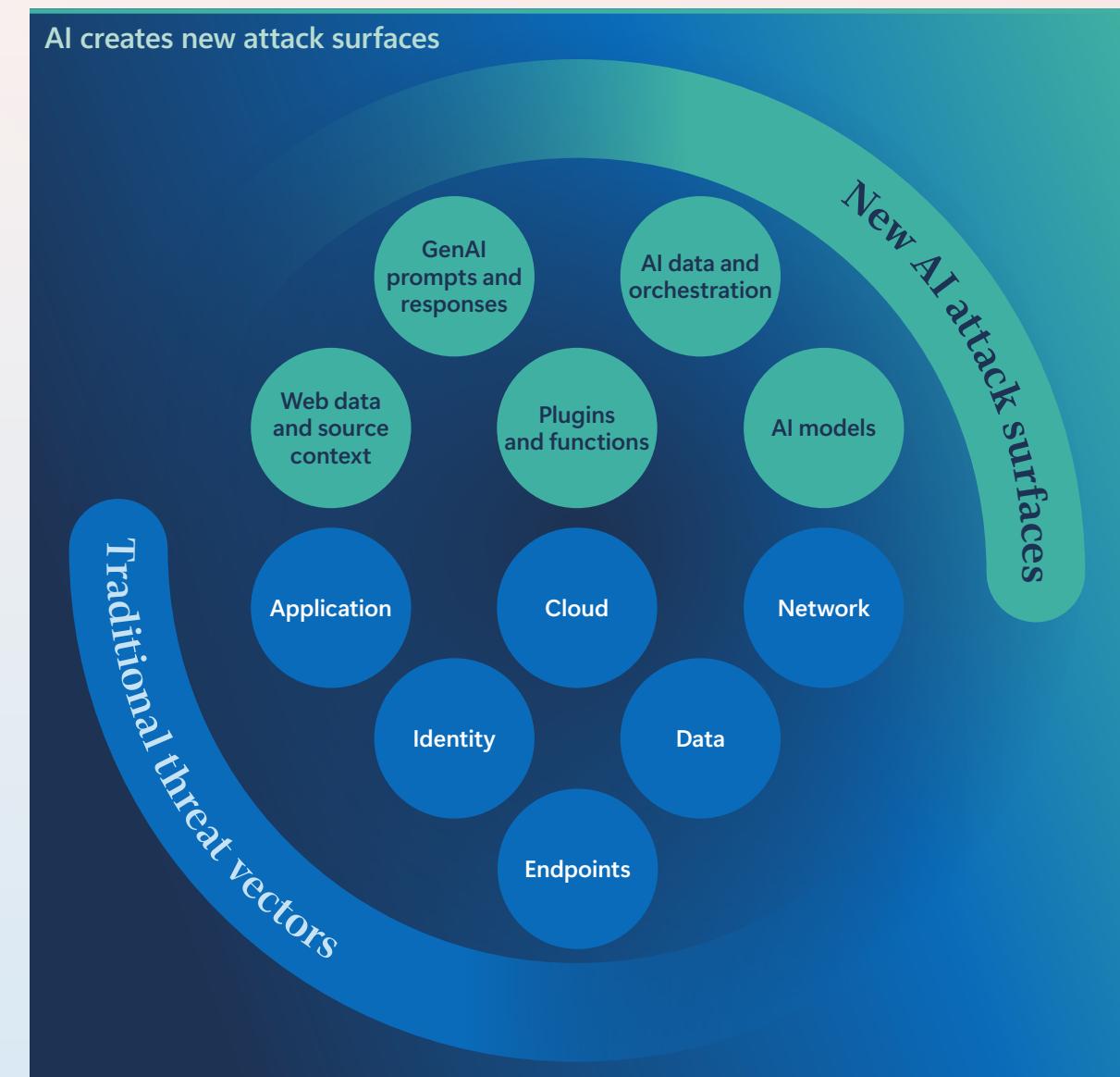
AI malfunctions can overlap with adversarial threats. Overreliance on AI, information leakage, and agency problems are key concerns that affect all users of AI.

**Overreliance** on AI can be exploited by attackers who feed manipulated data or craft deceptive scenarios, seeding false information into systems or even triggering disruptions to operations. One way to reduce this risk is to treat the AI like a new hire whose work will benefit from review and feedback, not an infallible expert. On the development side, this includes designing systems to be secure by design and default, with human oversight. On the deployment side, this means ensuring AI is just one component of a larger process that you secure. Implementing periodic audits of AI outputs, establishing review protocols, and fostering a culture of questioning AI recommendations are essential strategies for users of AI to mitigate potential overreliance risks. It is critical to treat AI as an augmenting tool rather than an infallible decision-maker.

**Information leakage** can be exploited by attackers during runtime or from training data, exposing sensitive organizational details. AI systems handling customer interactions or proprietary data are prime targets—threat actors can extract confidential information through prompts or vulnerabilities in datasets. Strong defenses require strict data governance: enforce data labeling and permission expiration, encrypt sensitive data, and implement policies to prevent over-permissioning. You can also use AI preemptively to detect information leakage by regularly asking your own AI tools to research confidential subjects within your organization which they shouldn't be able to access, and securing any leaks they inadvertently discover and exploit. These measures reduce the risk of adversaries weaponizing AI to compromise critical information.

**Agency risks** can be exploited by attackers manipulating AI objectives to favor their interests over stakeholders'. For example, adversaries might inject biased data or influence reward signals so an AI agent prioritizes advertisers or malicious actors over users, eroding trust and security. Organizations must counter this with transparency, explainability, and strong governance. Scenario testing for conflicts and embedding ethical safeguards into design are critical to prevent adversarial exploitation of AI goals, especially in autonomous agentic systems.

**AI creates new attack surfaces**

New AI attack surfaces:
- GenAI prompts and responses
- AI data and orchestration
- Web data and source context
- Plugins and functions
- AI models

Traditional threat vectors:
- Application
- Cloud
- Network
- Identity
- Data
- Endpoints

## Dangerous capabilities

AI's powerful capabilities extend to producing sensitive materials and enhancing skills in ways that, if misused, pose significant security risks. As a result, it is essential for developers of AI and policymakers to establish clear guidelines to ensure appropriate use while minimizing the risk of misuse.

**Production of sensitive materials** refers to the generation of content such as manipulated imagery or videos that could be used unethically, such as child sexual abuse material (CSAM). As previously mentioned, deepfakes also pose serious risks in areas like financial fraud, corporate espionage, and spreading false information during crises, which can cause confusion and hinder emergency responses. Deepfakes can also facilitate identity theft and nonconsensual intimate imagery (NCII). NCII is frequently used to facilitate harassment and extortion, especially of minors.

**Skill uplift** through AI can empower individuals to acquire new knowledge, but it does require oversight to ensure that those skills are not used for malicious purposes. For example, bad actors could use AI to learn how to develop chemical weapons or plan mass-casualty attacks. AI should be designed with strict filters and intent detections to block requests for harmful knowledge, with suspicious queries reviewed by humans.

## Operational issues

Addressing operational issues in AI systems requires robust strategies that balance technical precision with business demands. Issues that might arise during the use of AI include logging and monitoring, ensuring model integrity, and equipping product teams with the skills to manage AI-associated risks effectively.

**Logging and monitoring** of AI use are foundational for incident detection, response, and compliance, but they can also expose sensitive data, create security risks, and overwhelm teams with unfiltered or biased information. When done correctly, the process of logging and monitoring involves systematically recording user inputs, system outputs, and internal behaviors of AI systems to ensure transparency and accountability. At the same time, logging conversations might raise privacy concerns. While the volume and sensitivity of this data can pose challenges, solutions such as advanced analytics and automated auditing tools can streamline the process. For example, implementing systems that track anomalies in real time can help detect fraudulent activities or unusual system behaviors before they escalate.

**Model integrity** ensures that the AI systems operate reliably and as intended over time. However, AI models are subject to the same supply-chain risks as other software. "Time bomb" attacks, in particular, modify the model during training to cause it to produce attacker-prescribed outputs when specific inputs appear, such as when the model is used by a particular company, past a particular date, or when an image includes a certain embedded visual trigger.

The relative opacity of model files makes compromise of a model very difficult to detect after the fact, making securing of the build process especially important. Common model-building tools should produce dependable artifacts, such as a comprehensive software bill of materials (SBOM) that can be used to verify the authenticity and functionality of deployed models. For instance, frequent integrity checks can ensure that no unauthorized alterations have been made to the system, safeguarding against potential breaches.

## Emerging threats

New threats related to AI and its use are continually emerging as AI technology evolves. Current research to mitigate these threats includes securing long-running agents and managing risks from read-write memory.

**Securing long-running agents** involves ensuring agents remain aligned to their goals and managing errors and confusion from hostile data (which might come from external manipulation). This focus area is particularly relevant for industries relying on automation and AI-based decision-making or for companies that use AI agents to automate customer service. Corrupted data or adversarial attacks can disrupt operational efficiency or lead to a reputational loss. Enterprise users of AI can implement strategies like continuous goal verification protocols, anomaly-detection systems, and adaptive learning algorithms which are essential to maintain reliability and enhance trust in agents.

**Risks from read-write memory** include issues such as data corruption, latent poisoning attacks, and positive-feedback loops. These can erode a system's reliability, particularly when it relies on dynamic memory updates, and are a pressing concern for developers and security professionals who manage AI-driven systems. AI developers and platform providers should implement strict data validation, use immutable data structures, and employ advanced monitoring tools to help mitigate these risks.

> ↗ **Learn more**
>
> Researchers find—and help fix—a hidden biosecurity threat | Microsoft Signal Blog

## Storm-2139: How Microsoft disrupted an AI exploitation and abuse ring

Microsoft, together with generative AI technology providers worldwide, is navigating the challenge of driving AI innovation while staying true to our core principles. Our Digital Crime Unit's action against a group we track as Storm-2139 exemplifies how we can proactively shape the future of responsible AI.

In July 2024, Microsoft uncovered a global network exploiting stolen API keys to bypass AI risk and governance measures of various popular AI services, including Azure OpenAI. The developers were using and selling their nefarious tools, which were used to create thousands of abusive AI-generated images, including celebrity deepfakes, sexually explicit imagery, and misogynistic, violent, or hateful synthetic content. By using content provenance tools and open-source intelligence, the Digital Crimes Unit (DCU) was able to trace the origins of this malicious behavior. The network we uncovered included the software developers, providers who customized and distributed the software, and end users who deployed these tools to create synthetic content.

### A global network of developers, providers, and end users

Microsoft's amended complaint in February 2025 named the key developers and providers behind the nefarious tools used to create abusive AI-generated images.



Key

🔴 **D** Developers

🟢 **P** Providers

🔵 **U** Users

To disrupt the network, the DCU implemented a two-phase approach. In December 2024, the DCU filed a civil complaint to seize and sinkhole the primary domain used by Storm-2139 to communicate and collaborate. This action allowed the DCU to uncover additional evidence, leading to an amended complaint in February 2025 that named the key developers and providers behind the tools.

The response from the cybercriminal community was swift and revealing. Some users went silent, while others lashed out—posting warnings, blaming each other, and even doxing attorneys and investigators. Whistleblowers emerged, naming key figures and helping the DCU advance its investigation. In March 2025, Microsoft provided extensive criminal referrals to the Department of Justice (DOJ), Federal Bureau of Investigation (FBI), United Kingdom's National Crime Agency (NCA), and Europol's European Crime Center (EC3).

### Recommendations for defense

Regularly check and update access codes to help prevent unauthorized use, and set up alerts to notify you of unusual activity.

Adopt modern authentication methods like OAuth-based systems and enforce MFA for critical accounts.

Implement advanced monitoring and logging tools to detect irregular patterns and conduct periodic security audits.

Any evidence of violative images and prompts should be reported to national authorities.

### ↗ Learn more

Disrupting a global cybercrime network abusing generative AI | Microsoft On the Issues

Taking legal action to protect the public from abusive AI-generated content | Microsoft On the Issues

Microsoft files lawsuit against LLMjacking gang that bypassed AI safeguards | CSO Online

How Microsoft is taking down AI hackers who create harmful images of celebrities and others

Responsible AI Principles and Approach | Microsoft AI

# Quantum technologies:
## Strategic priority in a new era of competition

**Quantum technologies—computing, communications, and sensing—are foundational to future economic and national security.**
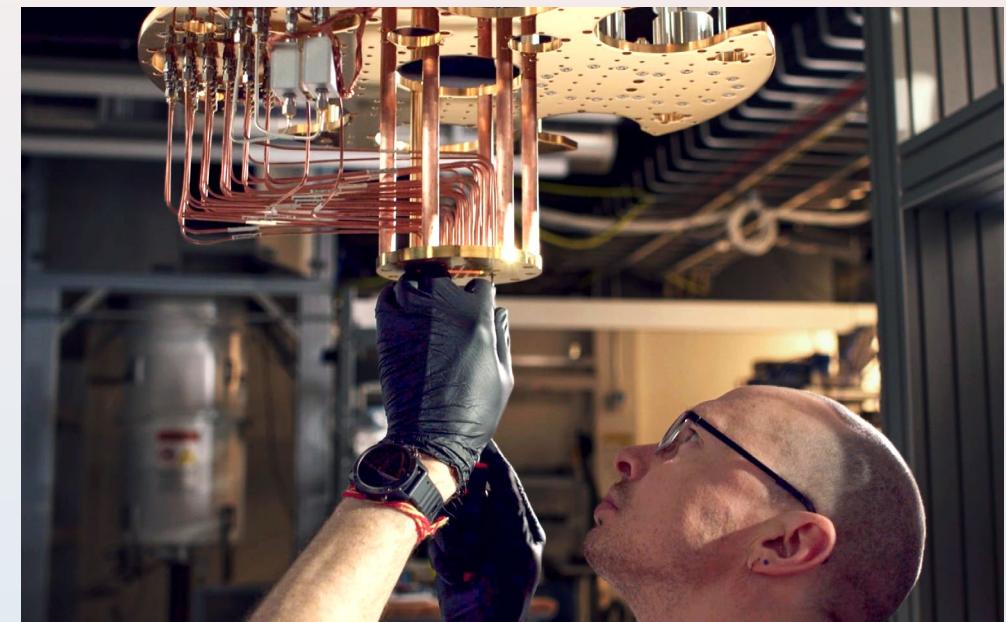
Quantum technologies' potential to accelerate scientific discovery, enable breakthroughs in secure communications, and disrupt encryption have made this technology a high-priority area. Indeed, governments have identified quantum technology as a national imperative. Allies and adversaries alike are pursuing quantum capabilities through new national research and development (R&D) programs, as well as investments to cultivate their own academic and private sector ecosystems. Certain adversaries may also leveraging additional capabilities to strengthen their position through espionage.

Commercial companies are driving a significant amount of current quantum R&D and private enterprise now sits at the epicenter of the global race to develop quantum technologies. Certain adversaries may also leverage additional capabilities to strengthen their position through espionage, including the possible targeting of Corporate R&D programs, startups, and academic spin-offs.[16] It is therefore imperative to establish robust safeguards and strategic preparedness now, before quantum technology becomes widely operational. The stakes are existential: leadership in quantum could determine not just competitive advantage but the future integrity of secure communications and the global digital economy.

The implications of the race to quantum advantage are sweeping:

- Industrial scientific leadership: Quantum technologies could drive a new wave of innovation across chemistry and material science discoveries.
- Impact to cryptography: A sufficiently powerful quantum computer could break widely used public-key algorithms, undermining the security of digital communications and data.
- Sensor superiority: Quantum sensors could detect stealth air or naval assets, eroding strategic deterrence.

**For a quantum future that is secure, prosperous, and inclusive, governments and industries must do three things:**

1. Prioritize security while simultaneously embracing innovation.

2. Reshape sectors of the economy to be first movers and capitalize on the quantum future.

3. Work globally to ensure that all humanity benefits through the responsible and ethical use of transformative technology.

# Part II
# The defense landscape

# Key takeaways

## Insights and actions for cyber defense

### 1. Cyber risk is business risk

As intrusion attempts become the norm, it is essential that governing boards and C-suites recognize that cyber risks are a form of business risk and treat them accordingly. Solutions to help mitigate this risk include conducting security exercises, implementing key performance indicators tied to cyber hygiene, and cross-training teams to build resilience.

### 2. AI-powered defense is essential

As adversaries begin to move at the speed of AI, so must defenders. Microsoft uses AI to conduct threat analytics, identify detection gaps, validate detections, identify phishing campaigns, automate remediation, and shield vulnerable users.

### 3. AI agents can help in threat mitigation and incident response

AI agents can help organizations automatically respond to threats, including by suspending suspicious accounts and initiating a password reset, containing a breach before an attacker can conduct further malicious activities. Agents can also enforce policies, monitor credentials and app permissions, and control employee accesses.

### 4. Organizations should implement a security framework for AI use

When using AI, it's important to mitigate risks such as data leaks or data oversharing, as well as risks to the AI itself such as prompt injections and insecure extensions. This means organizations require a strong security framework that helps them: prepare for AI adoption; discover how AI is being used within the organization; protect sensitive data, AI agents, applications and models; and govern AI operations.

### 5. Deterring cyberattacks requires political solutions

Individual defensive activities aren't enough to turn the tide of cyber threats from nation states. To protect cyber infrastructure, governments must build frameworks that signal credible and proportionate consequences for malicious behavior. This includes regularizing public attributions, signaling red lines, and imposing consequences.

### 6. Cooperation across borders is crucial to mitigate cyber risks

Whether addressing threats like ransomware and cyber mercenaries or managing emerging technologies like AI, cooperation between the public and private sectors and academia is essential. This includes formulating policy frameworks, establishing protocols, working on shared initiatives, intelligence sharing, and engaging in dialogue.

### 7. Resilience must be woven in by design

Given the persistence of cyber threats, it is important that systems are designed to anticipate, withstand, recover from, and adapt to disruptions. Resilience must be embedded into the very DNA of an organization's infrastructure.

### 8. Public-private collaboration is key to disrupting cybercrime ecosystems

Successful operations like the Lumma Stealer takedown demonstrate the power of coordinated legal, technical, and operational strategies across sectors to disrupt malicious infrastructure and protect critical assets.

### 9. Governments are moving away from voluntary compliance toward cyber requirements

Across the globe, governments are accelerating efforts to manage cyber risk through new laws and regulations. In particular, they are moving from voluntary guidelines to enforceable standards that emphasize accountability, risk management, and timely incident reporting. At the same time, to maximize their effectiveness, governments must pursue harmonized, risk-based approaches that promote interoperability and reduce duplication across borders.

### 10. Organizations must prepare for quantum computing

Quantum computing poses a serious threat to current cryptographic systems. As a result, organizations should inventory their cryptography (keys, certificates, and protocols) and establish a roadmap to replace vulnerable algorithms with PQC standards as they become available. Microsoft has established the Quantum Safe Program to achieve "quantum readiness" by systematically integrating post-quantum cryptographic algorithms into our services.
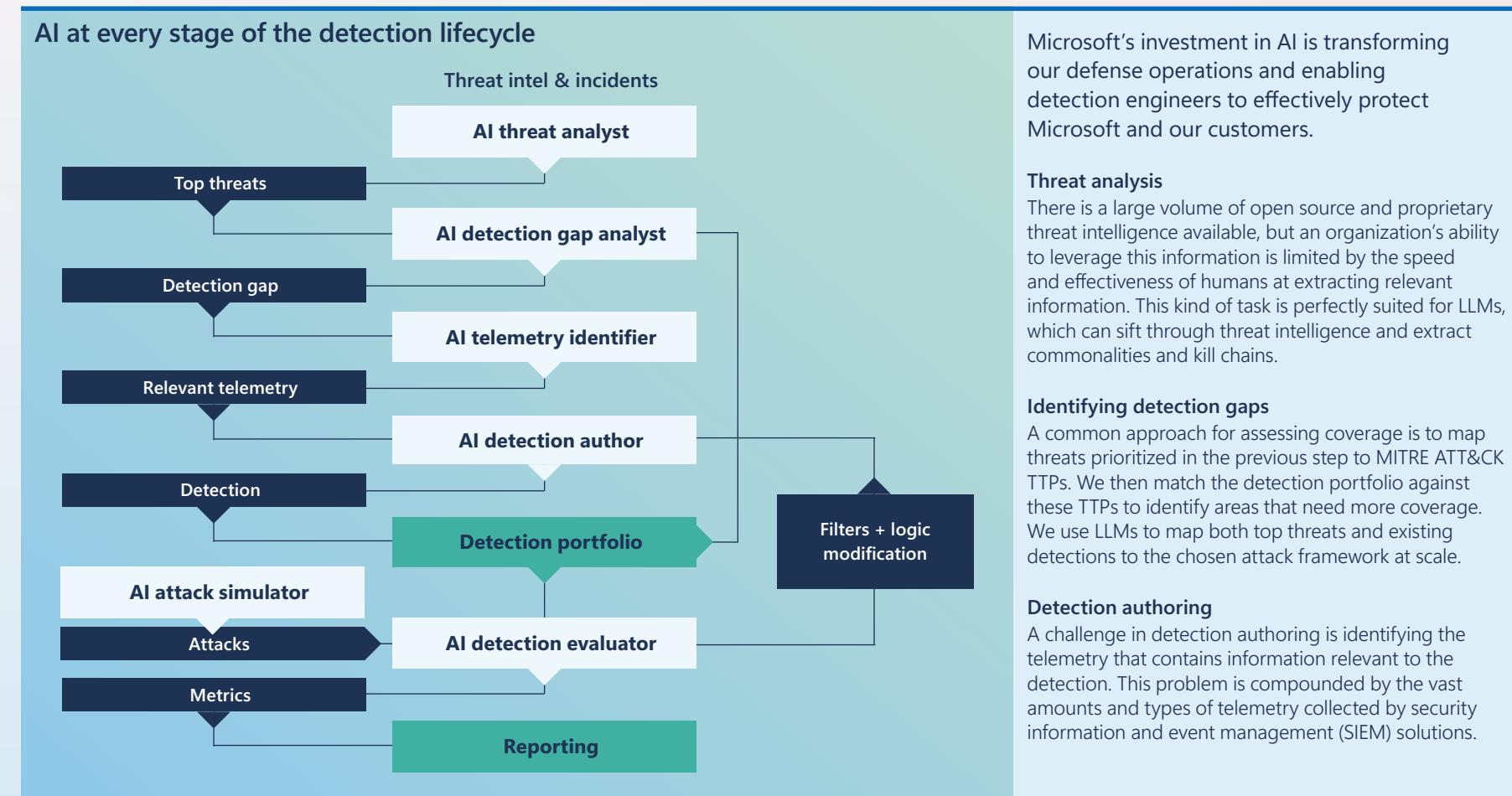
# AI and advanced defense

## AI-powered defense: Transforming threat detection

Detection Engineering (DE) is a fast-growing discipline in mature cybersecurity organizations. Given the increase in the volume and sophistication of cyberattacks, there is a pressing need for dedicated detection teams. While incident responders can write detections during an incident, these detections are typically incident-specific and narrow. The result is a detection portfolio that is always one step behind the attacker. DE teams focus on strategic, scalable prioritization and development of a dynamic, forward-looking detection portfolio.

At Microsoft, we developed a variety of AI solutions to help DE teams effectively manage detections throughout their lifecycle. On the right, we give examples of these AI solutions and how they can transform every stage of a detection's lifecycle.

### AI at every stage of the detection lifecycle

Threat intel & incidents

AI threat analyst

Top threats

AI detection gap analyst

Detection gap

AI telemetry identifier

Relevant telemetry

AI detection author

Detection

Detection portfolio

Filters + logic modification

AI attack simulator

Attacks

AI detection evaluator

Metrics

Reporting

Microsoft's investment in AI is transforming our defense operations and enabling detection engineers to effectively protect Microsoft and our customers.

**Threat analysis**
There is a large volume of open source and proprietary threat intelligence available, but an organization's ability to leverage this information is limited by the speed and effectiveness of humans at extracting relevant information. This kind of task is perfectly suited for LLMs, which can sift through threat intelligence and extract commonalities and kill chains.

**Identifying detection gaps**
A common approach for assessing coverage is to map threats prioritized in the previous step to MITRE ATT&CK TTPs. We then match the detection portfolio against these TTPs to identify areas that need more coverage. We use LLMs to map both top threats and existing detections to the chosen attack framework at scale.

**Detection authoring**
A challenge in detection authoring is identifying the telemetry that contains information relevant to the detection. This problem is compounded by the vast amounts and types of telemetry collected by security information and event management (SIEM) solutions.

We are developing AI solutions that create detections at different levels of sophistication. For basic tests, for example, AI can generate rule-based logic that checks for a small set of specific events. For correlation detections, AI leverages telemetry metadata to identify candidates, then generates the logic that correlates them with malicious activity. For behavioral detections, we use machine learning to establish baseline behaviors and identify anomalies that can signal malicious activity.

Rapid advances in the development of code-generating LLMs means that whatever the level of sophistication of the detection logic, we can automate its implementation of it in our chosen coding language.

## Detection validation

Detection Engineering must test the artifacts they generate. A simple test injects or executes the events or behaviors the detection is intended to catch. This is the DE equivalent of unit tests.

However, attackers use sophisticated multi-stage approaches and decide what to do in later stages based on information gained in earlier stages. While unit tests validate individual detections, we need end-to-end testing to ensure the detection portfolio as a whole is effective. Microsoft has developed agentic red teaming approaches where autonomous AI agents simulate complex, adaptive multi-stage attacks, enabling effective validation at scale.

## Securing identity in the age of AI: Proactive and automated protection

AI and machine learning are revolutionizing how we detect identity threats by finding subtle patterns humans miss. Modern AI-driven identity protection systems continuously analyze billions of sign-ins and user signals, learning what normal behavior looks like for each user and entity so they can spot the early signs of an attack. For example, AI can detect a slow password spray attack by recognizing a coordinated pattern of sign-in attempts spaced out over a long duration, a pattern that would slip past traditional rate-limit rules. Similarly, AI models evaluate each sign-in against dozens of risk factors (impossible travel, unfamiliar devices, abnormal time of access, etc.) to assign a risk score in milliseconds. With this information, advanced anomaly detection algorithms can instantly flag a threat actor using a stolen token from an unusual location or attempting to mimic a user's typical location.

While AI is still new, its impact is already significant: thanks to AI-based protections, providers report automatically neutralizing the vast majority of identity attacks. With the assistance of AI, security teams can remediate threats before they cause damage, with minimal false alarms or missed detections, making defenses both faster and smarter.

### AI agents for response and remediation

Beyond using AI to detect identity threats, organizations are increasingly using AI agents to respond automatically to threats. These agents can act in the identity environment with minimal human guidance—sometimes even on their own. When they either confirm or strongly suspect a threat, they can act within seconds—far faster than a human can respond manually. For example, if multiple high-risk signals indicate an account compromise, an AI agent can immediately suspend the account, initiate a password reset, and notify administrators, containing the breach before an attacker can escalate their access privileges.

AI agents also tackle preventative maintenance, working continuously to reduce the attack surface and fix security gaps that attackers might exploit.

- **Policy enforcement agents** review identity configurations and policies like MFA enrollment or conditional access rules and automatically reinforce any weak spots. After the agent flags users not covered by MFA, it can help enroll them or adjust the policy scope. This ensures security policies cover every user and scenario as intended.

- **Credential hygiene agents** monitor secrets and credentials. If an API key or client secret not used in months still sits in an app configuration, the agent might recommend rotating or removing it to prevent potential abuse. Similarly, this agent can monitor for leaked credentials or known compromised passwords and trigger immediate remediation.

- **Application risk detection agents** keep an eye on app permissions and behaviors. Should an app request higher privileges or exhibit anomalous behavior with user-granted access, the agent will alert security and preemptively revoke or quarantine the app. This will swiftly reverse unauthorized access, nullifying the malicious consent threat vector.

- **User lifecycle agents** govern access by automatically assigning the right permissions based on attributes such as role, department, group, and certifications. For instance, when an employee leaves, the agent revokes all their sessions and removes all access, closing a common gap that turns lingering accounts into backdoors. If an employee changes roles, the agent could suggest removing permissions no longer needed, preventing accumulation of privileges.

AI-driven agents operate under strict policies, only taking well-defined actions and requiring human-in-the-loop.

**AI and advanced defense** continued

## Cloud-scale AI defense: Guardian agents

As organizations accelerate their adoption of AI, threat actors target that AI. This demands a new class of defense: AI systems purpose-built to protect other AI systems.

One of the most pressing challenges is prompt manipulation attacks, including direct and indirect prompt injections, and exploitation through protocols such as Model Context Protocol (MCP) and Agent2Agent (A2A). The heart of these attacks is usually to inject a malicious payload into the AI's processing stream which hijacks its behavior and causes it to run attacker-controlled instructions. These attacks involve reconnaissance phases, where attackers systematically probe the model to identify vulnerabilities before launching targeted operations. Malicious content can be linguistically obfuscated or embedded in seemingly benign files, which defeat simple keyword and regex filters. Depending on the system affected, these attacks may execute read or write commands, exfiltrate data, or subtly modify the system's behavior to suit attacker objectives, such as by changing the outcome of analyses.

So, defenders deploy intelligent "guardian agents"—dedicated security agents with transparent access to the protected model. This visibility into the model's internal reasoning, tool usage, and decision chains enables real-time detection of malicious behavior that would otherwise remain hidden.
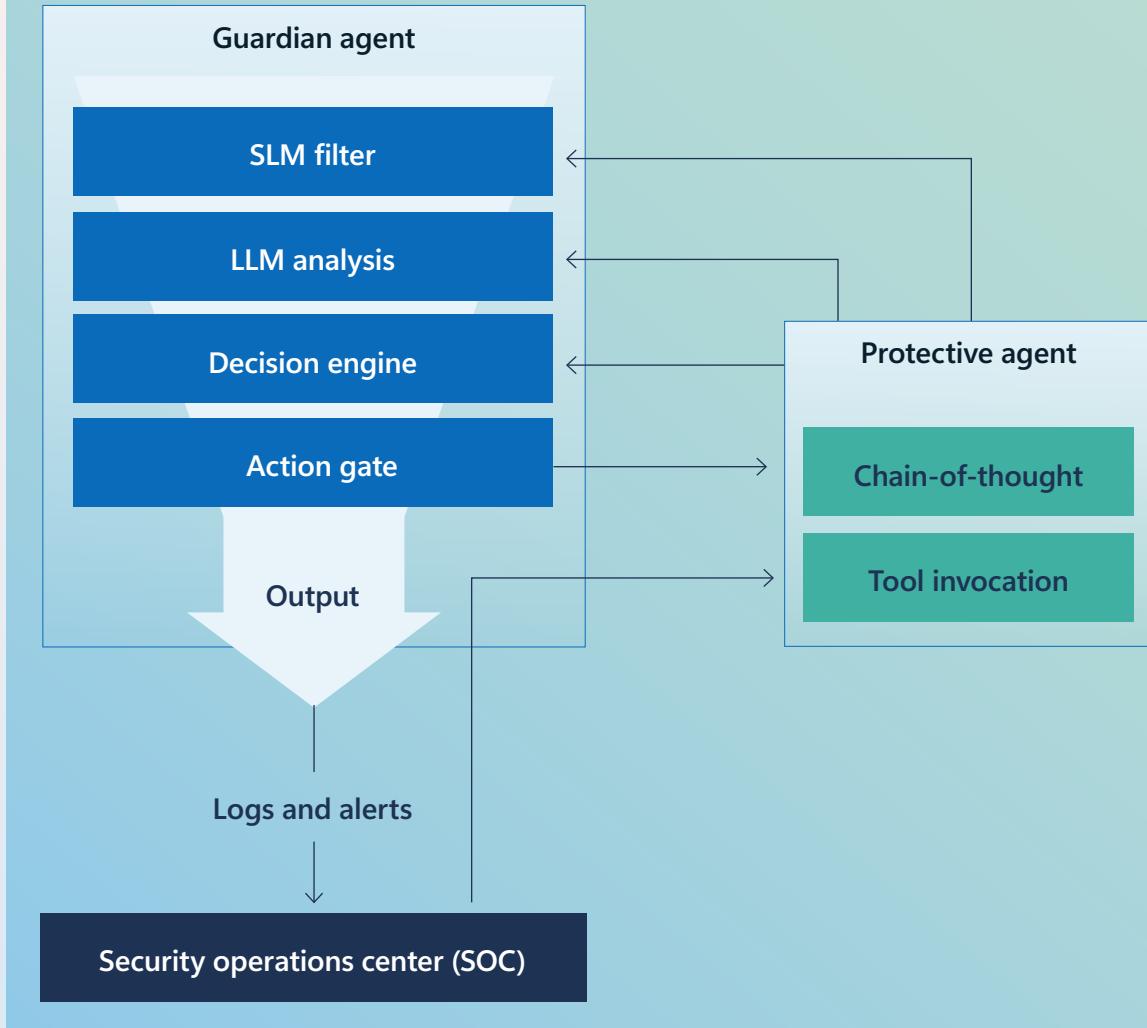
A layered defense strategy is essential. At the surface, Small Language Models (SLMs) provide lightweight, highspeed screening of prompts and responses, flagging suspicious patterns at scale. Deeper in the funnel, suspicious signals flow from the SLMs into advanced LLMs, combined with context from the agent's internal processes, such as tool invocations, reasoning traces, and state changes. The LLM correlates these signals, reconstructs the likely attack scenario, and issues a verdict: allow, rewrite, or block—and raises an alert. This funnel approach balances efficiency with depth, ensuring both broad coverage and precise decisions.

Beyond the model itself, telemetry from orchestration frameworks, APIs, and cloud services play a critical role. AI-driven engines baseline normal behavior across these systems and raise alerts when deviations occur, such as an agent invoking an unexpected function or accessing an untrusted domain. These signals are then correlated across identities, endpoints, SaaS applications, and additional cloud workloads including containers, serverless functions, virtual machines, Kubernetes pods, and managed platform services, to reveal coordinated attack patterns.

In this AI-first era, defending AI with AI is not just a necessity, it's a strategic advantage. By embedding intelligent, adaptive, and context-aware defense mechanisms directly into AI systems, organizations can stay ahead of adversaries and ensure the integrity of their AI assets.

**Defensive AI systems protecting AI agents**

Guardian agent
- SLM filter
- LLM analysis
- Decision engine
- Action gate

Output

Logs and alerts

Security operations center (SOC)

Protective agent
- Chain-of-thought
- Tool invocation

**AI and advanced defense** continued

## Securing AI systems: Safeguarding the enterprise and its innovations

The adoption of generative AI by enterprises introduces two security imperatives: securing the enterprise from risks associated with the enterprise AI and securing the enterprise AI itself.

The former focuses on mitigating risks posed by how generative AI is used across the workforce— for example, data leaks, data oversharing, misuse of third-party tools, or unintentional sensitive information disclosure. The latter addresses risks within the AI systems themselves, including prompt injections, training data poisoning, and insecure extensions. Per findings from our report, Secure Employee Access in the Age of AI, 57% of organizations have experienced an increase in security incidents linked to AI usage.[17] Yet despite growing awareness of the need for AI controls, many organizations may have yet to implement any. This creates a gap between adoption and protection in the enterprise.

As generative AI apps and agents become deeply embedded in business workflows, security teams need end-to-end visibility and control. A strong security framework helps organizations: prepare for AI adoption; discover how AI is being used within the organization; protect sensitive data, AI agents, applications, and models; and govern AI operations with clear policies and safeguards for compliance and new AI regulations.

### This framework should help organizations

#### Prepare

Anticipate AI adoption by establishing policies, training, and secure foundations before deploying AI, including data classification and security, access controls, and zero trust.

#### Discover

Gain visibility into how AI is used in the organization. Monitor AI applications and agents, detect unsanctioned shadow AI tools, identify what data is going into and coming out of AI systems, and discover risks and vulnerabilities in AI apps, agents, and models.

#### Protect

Safeguard sensitive data and AI systems. This includes preventing data, defending against prompt injection attacks, and securing AI apps and agents.

#### Govern

Enforce policies and oversight for AI use. Retain and audit AI interactions, ensure compliance with evolving regulations, and set clear guidelines for AI behavior.

**\***

Using innovative AI-driven tools, the DCU is accelerating its impact in the fight against cybercrime.

## AI vs. cybercrime: How automation is shifting the balance

The DCU is leveraging AI to confront the rapidly evolving threat landscape and the increasing sophistication of cybercrime. At the heart of the DCU's strategy is a suite of specialized AI tools that enhance its ability to monitor, investigate, and disrupt malicious activity. For example, the DCU has developed a machine learning system that analyzes password spray attacks to distinguish between normal and targeted behavior. This enables the team to identify high-risk users and proactively protect vulnerable populations—such as rural hospitals and political candidates—before harm occurs. Another powerful tool in the DCU's arsenal is its domain impersonation monitoring system. By using AI to detect and track impersonation (or homoglyph) domains, the DCU can anticipate and block phishing campaigns and other malicious activity that rely on these deceptive URLs.
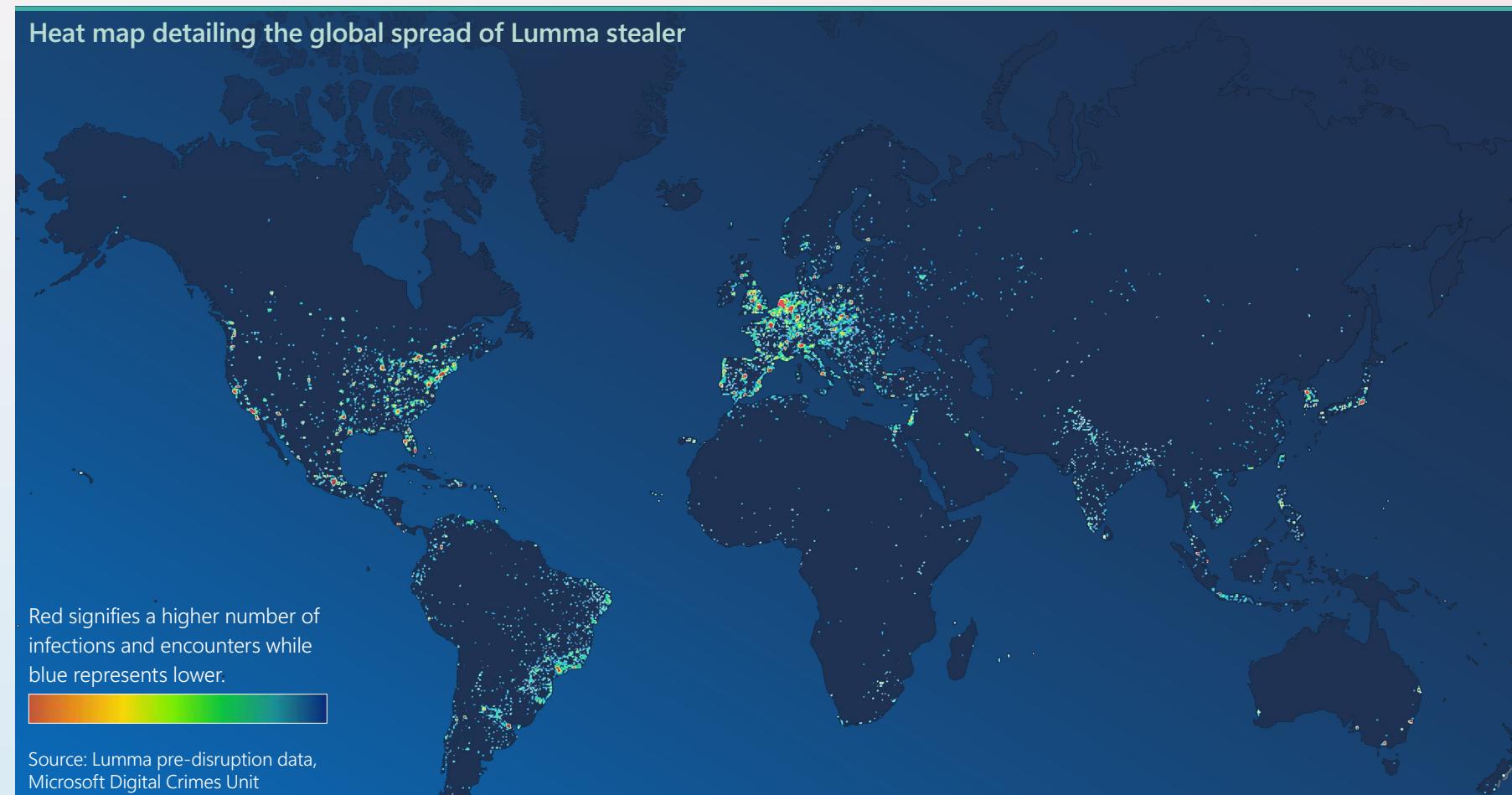
AI also plays a critical role in investigations. The DCU uses AI-powered agents to sift through massive datasets, extract key indicators of compromise (IOCs), and share them across Microsoft's security ecosystem. A reverse engineering plugin powered by AI further accelerates the analysis of malicious code, automating tasks that once took hours or days.

# Countering nation-state and emerging threats

**Disrupting cybercrime ecosystems: Lessons from the Lumma Stealer takedown**

Given Lumma Stealer's prominence in the infostealer ecosystem and its role in enabling broader cybercriminal operations, it became a high-priority target for disruption this year. In May 2025, the DCU, in collaboration with global law enforcement and cybersecurity partners, successfully disrupted the Lumma Stealer infrastructure in a joint operation exemplifying the power of public-private collaboration in proactive cyber defense.

Through a US court order and coordinated actions with the US Department of Justice, Europol, Japan's Cybercrime Control Center (JC3), and private sector partners like ESET, Bitsight, Lumen, CleanDNS, and GMO Registry, over 2,300 malicious domains were seized or blocked. These domains formed Lumma Stealer's infrastructure backbone.



Heat map detailing the global spread of Lumma stealer

Red signifies a higher number of infections and encounters while blue represents lower.

Source: Lumma pre-disruption data, Microsoft Digital Crimes Unit
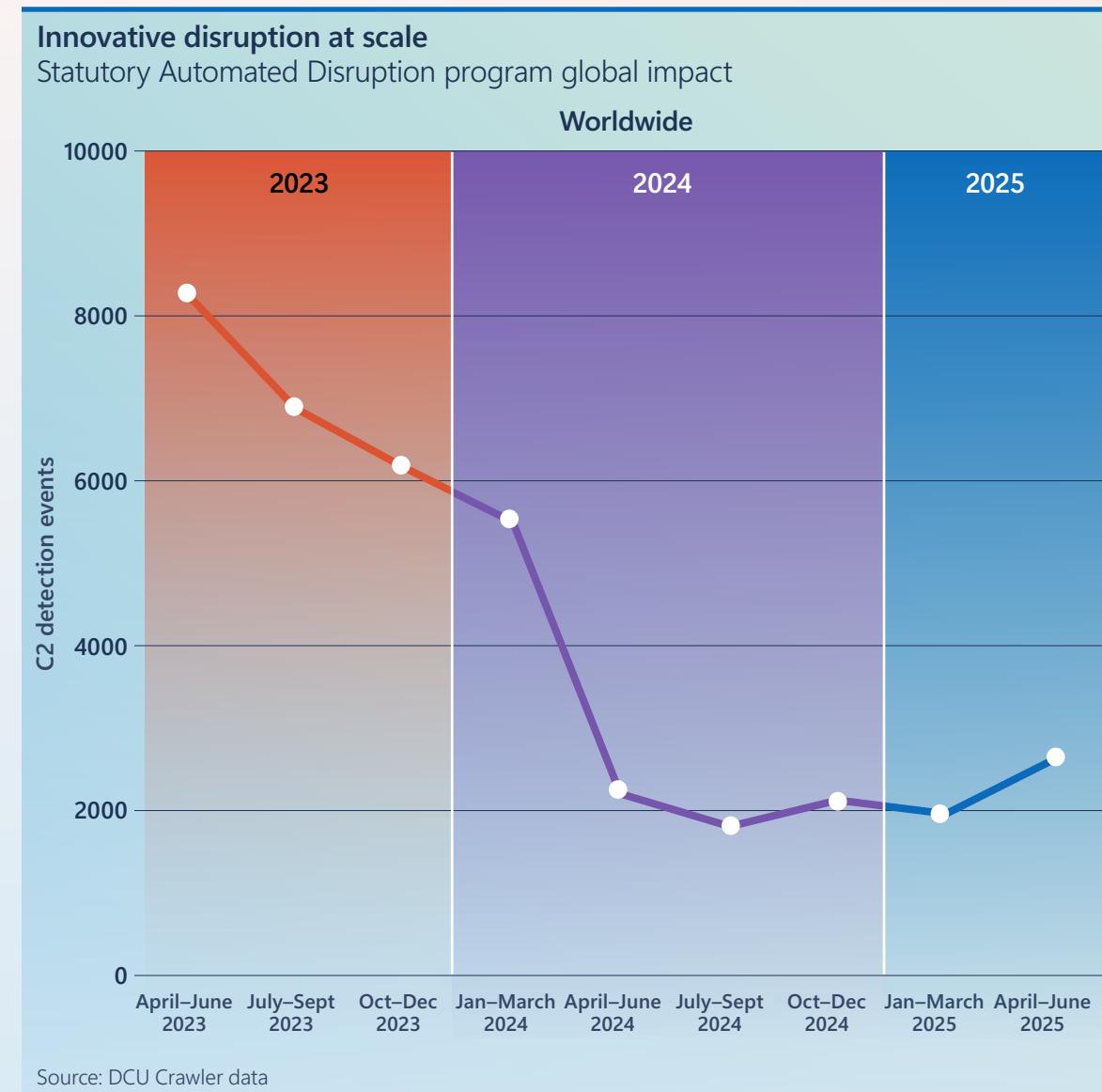
Countering nation-state and emerging threats continued

The disruption not only severed communication between infected devices and Lumma Stealer's command centers, but also redirected domain traffic to Microsoft-controlled sinkholes. This provided the DCU with enhanced threat intelligence, which it monitors, enriches, and shares with external partners through the Cyber Threat Intelligence Program (CTIP). The operation highlights how coordinated legal, technical, and operational strategies across sectors can significantly disrupt cybercriminal ecosystems and protect critical infrastructure.

These disruption actions were not one-time events, but part of a sustained strategy to limit threat actors' ability to rebuild. By employing innovative techniques—such as court-appointed monitors and DCU's Statutory Automated Disruption (SAD) program—the DCU continues to identify and dismantle new Lumma Stealer infrastructure. Although the medium term effect of this operation has yet to play out, the potential impact of our proactive approach to degrading malicious infrastructure is demonstrated by the DCU's 2023 disruption of cracked Cobalt Strike, a tool widely used in ransomware attacks, including those targeting hospitals.

After the initial domain seizures, the DCU issued over 238K abuse and takedown notices to hosting providers globally, resulting in a 68% reduction in the average number of command and control (C2) servers and shrinking their average lifespan from 49 days to just 18 days.

## Innovative disruption at scale
Statutory Automated Disruption program global impact



**Worldwide**

Source: DCU Crawler data

Beginning in March 2025 and continuing through July, DCU telemetry detected a rise in cracked Cobalt Strike C2 infrastructure, with a pronounced spike in China. This pattern aligns with recent cybersecurity reports of coordinated malware campaigns originating from China that leverage cracked instances of Cobalt Strike. This activity underscores the importance of persistent, scalable, and cross-jurisdictional takedowns of malicious infrastructure. Further, Microsoft's collaboration with Fortra— the cybersecurity software company behind Cobalt Strike—is central to this effort, as Fortra regularly provides DCU with updated signatures that enhance detection systems protecting against emerging C2 infrastructure. DCU and Fortra continue to add new sources of intelligence to support the effort.

↗ **Learn more**

Disrupting Lumma Stealer: Microsoft leads global action against favored cybercrime tool | Microsoft On the Issues

Inside Microsoft's Global Operation to Disrupt Lumma Stealer's 2,300-Domain Malware Network | The Microsoft Threat Intelligence Podcast

Lumma Stealer: Breaking down the delivery techniques and capabilities of a prolific infostealer | Microsoft Security Blog

**Countering nation-state and emerging threats** continued

## Deterrence in action: Building consequences for nation-state actors

As infrastructure essential to daily life—including water, food, healthcare, communications, and transportation systems—becomes increasingly dependent on digital technology, nation-state cyber operations targeting these systems cannot be permissible; in particular those prepositioning for disruptive or destructive cyberattacks in case of future conflicts.

Defensive actions alone to protect critical infrastructure are unlikely to deter nation-state threat adversaries. These are politically motivated activities that must be addressed with political solutions as well. To protect critical infrastructure, political institutions, and civilian systems, governments must build frameworks that signal credible and proportionate consequences for malicious activity that violate international rules.

Over the past year, there has been a marked increase in recognition of the need for such cyber deterrence, with governments and industry aligning more closely to response to malicious activity. For example:

- **NATO** has advanced coalition-based attribution frameworks and is exploring collective countermeasures in response to cyberattacks. In July, the alliance released a statement recognizing and condemning malicious cyber activities attributed to Russia by member states.

- The **US administration** has issued strong public statements and indictments tied to cyber operations and has publicly attributed cyberattacks in coordination with allies and partners.
- The **EU** is increasingly leveraging its Cyber Diplomacy Toolbox and sanctions regime to hold bad actors accountable, though implementation remains uneven.

Looking ahead, these are important foundations to build upon. To further strengthen a cyber deterrence framework, like-minded governments should work to:

- **Regularize public attributions.** States should more consistently issue public attribution statements, leveraging insights from other governments and partners in the private sector and establishing a more uniform process for doing so. Such statements should always indicate if international laws or norms were violated during a cyber incident.
- **Signal red lines.** States should make clear they will impose increasingly severe consequences in response to a spectrum of malicious nation-state cyber activity, ranging from espionage to prepositioning to disruptive and or destructive cyber operations.
- **Impose diverse consequences.** Responses to nation-state cyberattacks should not be constrained to the cyber domain or prescribed in a one-size-fits-all model. Different threat actors will be deterred by different consequences. These could include economic measures, diplomatic sanctions, naming and shaming, posturing, or targeted declassification.

- **Prohibit retaliatory cyber operations.** Private companies are not in the position to independently hack back against malicious nation-state actors, and doing so can risk unintended escalation and harm. While industry can support attributions and partner with government to take action, imposing consequences for internationally wrongful behavior by states will always need to be led by governments.

A viable model for cyber deterrence is a necessity for the stability of the online world and will require innovations in statecraft and diplomacy in the years ahead. This is why Microsoft is supporting ongoing research by the Royal United Services Institute (RUSI) to explore novel approaches to deterring malicious activity online.

## Addressing the geopolitical enablers of ransomware operations

Many of the most prolific ransomware groups avoid consequences by targeting victims in other countries while their own governments turn a blind eye. Whether they are state-affiliated groups or their government simply ignores their activity, the result is the existence of "safe haven" states that enable ransomware attacks abroad and violate international norms of due diligence which oblige governments to take action to prevent illegal cyber activity within their borders.

As a result, addressing ransomware operations requires a more coordinated international effort and political pressure that holds governments accountable for both direct and indirect support of ransomware attacks. Designating state sponsors of ransomware, for example, similar to state sponsors of terror, with associated stigmas and penalties, is one way to incentivize states to confront ransomware groups operating within their borders.

Other approaches to address escalating ransomware include:

- **Legal action:** Ransomware is a form of extortion which, in most cases, violates existing laws. These should be applied whenever possible. By designating state sponsors of ransomware, civilians might be able to take further legal action against those governments following ransomware attacks to seek damages in civil courts.
- **Public-private partnerships:** Encourage industry partnerships with law enforcement to improve cooperation against cybercrime. Examples include the International Counter Ransomware Initiative (CRI)[18] and the Institute for Security and Technology (IST) Ransomware Task Force.[19]
- **Deterrent consequences:** Governments should set clear expectations around what is responsible state behavior, reinforced by escalating consequences across domains sufficient to deter state-sponsored, or enabled, ransomware attacks.

## Combating cyber mercenaries: Closing the gaps in global regulation

Cyber mercenaries, private firms that sell offensive cyber capabilities, operate in legal gray zones, often across borders. Their cross-jurisdictional nature and a lack of oversight make them difficult to trace or prosecute, allowing them to act with near impunity. Many also rebrand frequently, shift operations across jurisdictions, and use complex financial networks to further evade detection and regulation.

To counter this growing threat, governments and industry must collaborate further to disrupt the enabling market through intelligence sharing, coordinated responses, and regulation. International norms should also prohibit the use of cyber mercenaries and close legal loopholes that allow them to persist. Governments need to put in place severe limitations—or outright bans—on the cyber mercenary market to ensure their products, including spyware, cannot be used in violation of domestic or international law, human rights, or to significantly undermine product security.

Examples already exist of states taking effective action. The US has placed restrictions on when federal agencies can solicit the services of cyber mercenaries and banned firms that operate irresponsibly, meaningfully impacting the bottom lines of some cyber mercenary firms. Meanwhile, the UK and France have made strides over the past year in their stewardship of the Pall Mall Process, an international multistakeholder dialogue that includes more than 20 government participants and

which seeks to regulate Commercial Cyber Intrusion Capabilities (CCIC) with shared frameworks. In April 2025, the Pall Mall Process produced a first-of-its-kind Code of Practice for governments to follow in order to limit harmful impacts of CCICs.[20]

Transparency is key. Governments should expose vendors and intermediaries, enforce sanctions, and lead by example by refraining from using cyber mercenaries themselves. Meanwhile, industry must enhance platform security, monitor abuse, and act swiftly to disrupt cyber mercenary operations. Through due diligence and collaboration, both sectors can help shrink the space in which cyber mercenaries operate—protecting national security, human rights, and global digital stability.

> ↗ **Learn more**
>
> Protecting users and reaffirming our commitment to combatting cyber mercenaries | Microsoft On the Issues
>
> Protecting the public from abusive AI-generated content across the EU | EU Policy Blog (March 2025)

**Countering nation-state and emerging threats** continued

## Intelligent signals: Accelerating incident response and recovery

Threat-informed defense strategies aren't just for large organizations; all organizations can implement threat-informed defense.

Understanding the threat landscape and curating relevant operational processes can be a great start for small organizations to enhance their security lifecycle. For example, start with the basics: understand the organization's attack surface and most applicable threats first, then build from there.

The Microsoft Detection and Response Team (DART), leverages intelligent signals throughout an entire investigation to make calculated decisions based on the motivations and techniques of threat actor campaigns, intercepting and disrupting threat actor activity in hours, not days.
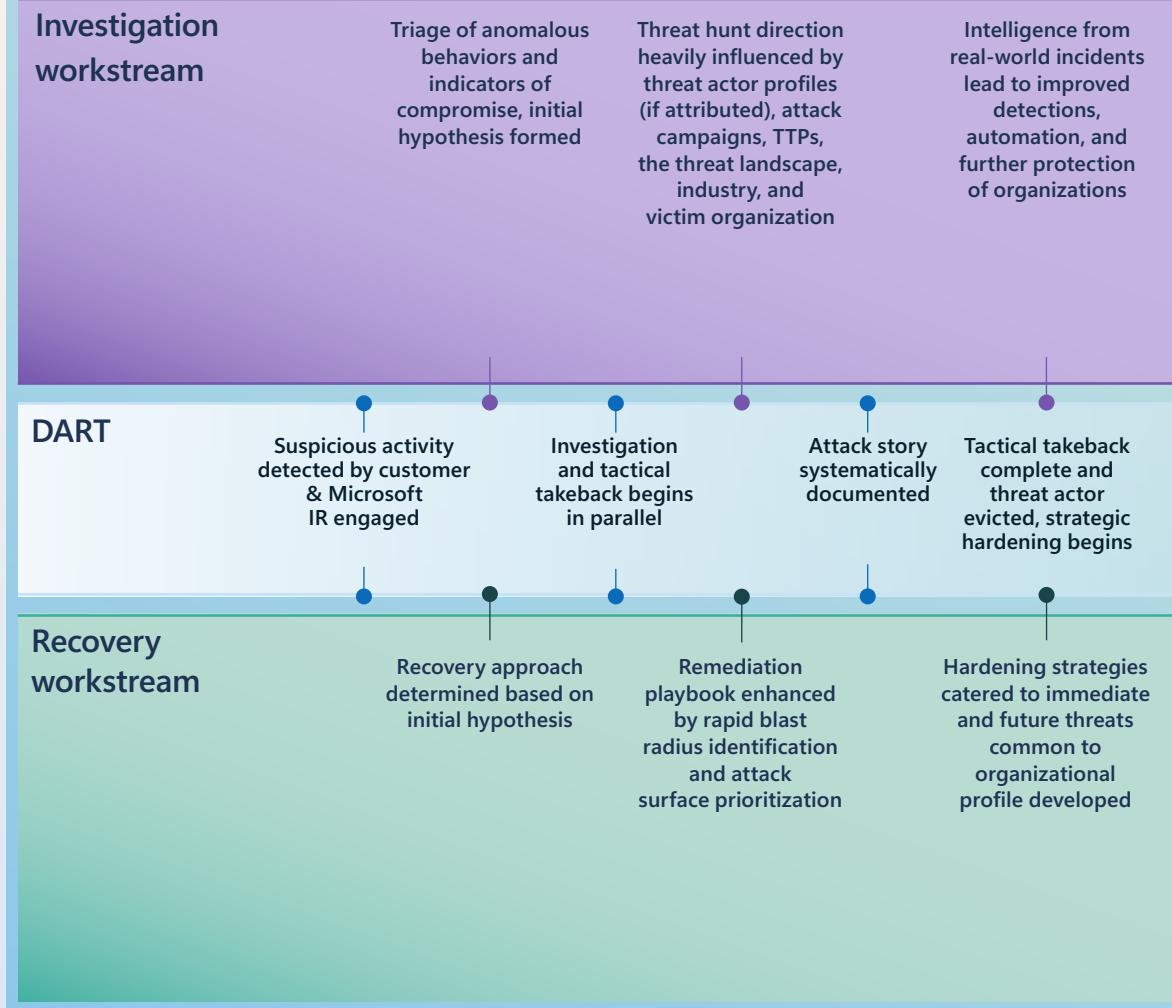
Applying diverse threat intelligence artifacts across multiple workstreams and stages of detection and response heavily influences the direction of threat hunting, tactical takeback efforts, remediation activities, and improved detection. Most importantly, this approach builds context-aware, tailored recommendations that can influence organizations' strategic security roadmaps and build towards a more secure future.

Organizations can enhance both proactive and reactive detection and response efforts by integrating a variety of threat intelligence artifacts holistically. Understanding the threat landscape, your own environment, and how high-quality threat intelligence can enhance detection and response includes:

- **Leveraging diversity in artifacts.** Threat intelligence comes in many forms. Atomic indicators of compromise and detection signatures should be paired with research into threat actor behavior. Threat hunters can't only rely on indicator-based hunting. Instead, they should have a broad understanding of threat actor motivations and TTPs.
- **Being industry and geographically threat aware.** Research and build a threat profile for the organization, such as its industry position, geographical location, and size. Use these data points to influence security roadmaps and prioritize implementation of security controls that directly mitigate prevalent threats.
- **Knowing what to protect and where.** Document the organization's internal security posture and relative attack surface. Highlight assets of value, those with trust dependencies and privileged pathways. Define a baseline of regular operation to rapidly highlight abnormalities should they arise.

### Intelligent signals in action
Incident response approach to rapid investigation and recovery

**Investigation workstream**

| Triage of anomalous behaviors and indicators of compromise, initial hypothesis formed | Threat hunt direction heavily influenced by threat actor profiles (if attributed), attack campaigns, TTPs, the threat landscape, industry, and victim organization | Intelligence from real-world incidents lead to improved detections, automation, and further protection of organizations |

**DART**

| Suspicious activity detected by customer & Microsoft IR engaged | Investigation and tactical takeback begins in parallel | Attack story systematically documented | Tactical takeback complete and threat actor evicted, strategic hardening begins |

**Recovery workstream**

| Recovery approach determined based on initial hypothesis | Remediation playbook enhanced by rapid blast radius identification and attack surface prioritization | Hardening strategies catered to immediate and future threats common to organizational profile developed |

While creating a dedicated threat detection and intelligence function is valuable, it can be costly. Nevertheless, security must be seen as an investment. Cybersecurity risks are business continuity risks. Building useful threat intelligence artifacts is a cyclical, collaborative effort. Knowledge sharing and partnerships are of paramount importance. Incident responders have a unique viewpoint of an organization's data—contextual artifacts can be continuously reported back into overall research efforts cyclically. This informs threat hunting, improves detections, and maintains cross-team awareness of the threat landscape. Extending collaboration with external partnerships also builds a stronger, collective defense against threats.

> "
> Security must be seen as an investment. Cybersecurity risks are business continuity risks.

## Collaboration as a counter measure: Breaking down fraud silos

Cyber fraud is growing more scalable and sophisticated, outpacing traditional defenses. A key vulnerability is the lack of robust, real-time data-sharing across sectors. Fragmented systems and siloed insights hinder early detection and coordinated response. One of the most effective countermeasures is structured collaboration between financial institutions, technology platforms, regulators, and law enforcement. Sharing fraud signals enables faster disruption of criminal activity—but it requires more than isolated partnerships. A unified approach that integrates diverse data sources is essential to expose abuse patterns and reduce harm.

Global efforts are gaining momentum. Initiatives like the Global Signal Exchange[21] promote standardized, privacy-conscious frameworks for multi-sector cooperation. Governments are responding to the trillions of USD lost to scams with legislation mandating reporting, liability reform, and stronger public-private collaboration.[22] Australia's Scam Prevention Framework Act 2025, for example, introduces sector-specific codes for fraud prevention.[23] The UK's national strategy, meanwhile, expands accountability to tech and telecom sectors and accelerates data-sharing mandates.[24] Singapore and Japan are tightening laws to counter digital payment fraud and cross-border scams.[25]

While approaches vary, these developments reflect a growing recognition of the need for more proactive, coordinated, and enforceable national responses to fraud. Looking ahead, we anticipate a significant acceleration in the implementation of these legislative frameworks as governments seek to close the regulatory gaps, enhance consumer protections, and build a more resilient digital economy.

↗ **Learn more**

Cross-border collaboration: International law enforcement and Microsoft dismantle transnational scam network targeting older adults | Microsoft On the Issues (June 2025)

Disrupting Lumma Stealer: Microsoft leads global action against favored cybercrime tool | Microsoft On the Issues (May 2025)

# Policy, capacity, and future readiness

## Securing the digital frontier: Government's role in responsible use of AI in national security and cyber defense

As we've seen throughout this report, AI gives cyber defenders a significant boost in meeting security challenges. To fully realize these benefits, especially in national security contexts, the use of AI must be guided by robust policy frameworks that allow for a sustained commitment to trusted, secure innovation.

For governments, this includes establishing strong procurement and security protocols to ensure AI systems are securely designed, developed, deployed, and used, especially when handling sensitive or classified data. By supporting research, training, and commercialization—particularly for startups and subject matter experts developing cutting-edge AI and cybersecurity solutions—governments can also use security as a lever for economic growth.

Governments play a pivotal role in fostering experimentation and mission-driven innovation in the use of AI for cyber defense through public-private partnerships. The United Kingdom's Laboratory for AI Security Research (LASR),[26] announced in November 2024, is an example of one such initiative bringing together critical government agencies with academic and other multistakeholder partners to advance AI benefits for national cyber resilience. Microsoft welcomes the recent White House AI Action Plan and the Administration's commitment to appropriately balance the dissemination of AI technologies, for example to improve defense of critical infrastructure, with national security considerations for Frontier AI. And we continue to partner closely with the US government to effectively address security risks to US AI companies, talent, intellectual property, and systems.

As AI is increasingly integrated into national security, intelligence, and defense operations, its deployment must be governed by clear legal framework, such as NATO's Principles of Responsible Use and the US Department of Defense's Responsible AI Framework. Multilateral dialogue and engagements with stakeholder groups from industry, academia, and civil society are essential to promote responsible innovation that enhances rather than endangers global stability. Governments should set clear expectations for acceptable AI use in national security, grounded in the United Nations (UN) Charter, international humanitarian law (IHL), and international human rights law (IHRL). Increased international coordination will be needed to enforce existing norms and develop new ones that reflect the capabilities and risks of AI, especially as autonomous, agentic systems advance.

## Implementing responsible AI in national security

At Microsoft, we expanded our responsible AI tools to better assess and manage adversarial risks in model development and deployment. Microsoft launched our Frontier Governance Framework,[27] which serves as a monitoring function, tracking the emergence of new and advanced AI model capabilities that could be misused to threaten national security or pose at-scale public safety risks. It also sets out a process for assessing and mitigating these risks so that frontier AI models can be deployed in a secure and trustworthy way. We are also developing engineering guidance and responsible AI policies to support emerging agentic systems, as these will play a growing role in AI development and deployment.

Microsoft maintains a consistent risk review process across AI releases, including red teaming and pre-deployment assessments for high-impact systems. This includes all generative AI systems and models, including Azure OpenAI and Phi family of models to help product teams safely deploy their generative AI applications and models. Microsoft's Sensitive Uses and Emerging Technologies team continues to advise on high-risk AI and high-impact applications—especially in healthcare and science—helping teams navigate novel risks and shape internal guidance. To streamline documentation, we introduced an internal tool that brings together all responsible AI requirements outlined in the Responsible AI Standard.

To stay ahead of evolving regulations such as the EU AI Act, Microsoft has taken a layered approach to compliance, in line with the AI Act's staggered compliance deadlines. Microsoft has undertaken multiple initiatives to promote AI literacy in accordance with the Act, empowering our employees, customers, and others to responsibly leverage AI technologies.[28] Microsoft also proactively took a layered approach to prepare for compliance with the Act's prohibited practices provisions.[29] In July 2025, we signed the General-Purpose AI (GPAI) Code of Practice, which includes a set of guidelines for compliance with the AI Act's GPAI model provider obligations, which came into effect in August 2025.[30] Microsoft continues to engage with the central EU regulator, the AI Office, and other relevant authorities in EU Member States to share insights from our AI development, governance, and compliance experience, as well as insights we hear from our customers.

Microsoft also worked with global partners to support more consistent governance approaches aligned with technical standards, including working closely with industry partners in the Frontier Model Forum and the Coalition for Secure AI.

---

⬈ **Learn more**

Responsible AI Transparency Report | Microsoft

Securing AI and Cloud with the Zero Day Quest | MSRC Blog | Microsoft Security Response Center

Microsoft commits to skilling one million people for digital skills through Artificial Intelligence skilling initiative in South Africa - Source EMEA

Unlocking data to advance European commerce and culture | Microsoft On the Issues (July 2025)

Microsoft announces AI skilling opportunities for 2.5 million people in the ASEAN region by 2025 | Microsoft Stories Asia

Microsoft Elevate: Putting people first | Microsoft On the Issues (July 2025)

Unlocking AI's global potential: progress, productivity, and workforce development | Microsoft On the Issues (April 2025)

Microsoft announces ARC Initiative to strengthen cybersecurity in Kenya | Microsoft On the Issues (May 2025)

The Accra Call for Cyber Resilient Development | GC3B

Home - The GFCE

microsoft/llmail-inject-challenge · Datasets at Hugging Face

### Resilience by design: Strengthening critical infrastructure for the next wave of threats

In today's hyper-connected world, new vulnerabilities are constantly emerging. As a result, cybersecurity expectations, practices, and oversight must evolve to prioritize resilience.

Cyber–physical threats can arise from a variety of sources, including natural disasters, industrial accidents, human error, technical errors, or malicious activities such as cyberattacks, terrorism, or armed conflict. These threats have the potential to disrupt the business and operations of critical infrastructures.

Given the interconnected nature of these risks, cyber–physical resilience encompasses both technical and organizational measures. Its goal is to prevent, protect against, respond to, resist, mitigate, absorb, accommodate, and recover from incidents.[31]

Cyberattacks are inevitable. Whether due to sophisticated threat actors, human error, or system complexity, breaches will occur. The key question is therefore not if a system will be attacked, but how well it can withstand attacks and recover. This is the essence of cyber-physical resilience: the ability of systems to anticipate, withstand, recover from, and adapt to disruptions—regardless of the cause.

Leaders should shift from a purely defensive posture to one that embraces resilience as a core design principle. This means building systems that can continue to operate under duress, recover quickly, and evolve to meet future threats. For leaders, this is not just a technical issue—it's a strategic one. The resilience of our infrastructure directly impacts national security, economic stability, and public trust.

By embedding resilience into the DNA of an organization's infrastructure, we not only protect our assets but also enhance our ability to compete and thrive in a volatile world.

Cyber-physical resilience is not just a technical challenge, it's a leadership imperative.[32] CEOs and CFOs must recognize that downtime, data loss, and reputational damage from cyber incidents can have profound financial consequences. Simultaneously, government leaders must ensure that national infrastructure can withstand and recover from attacks that could otherwise disrupt societal functions at scale. Maintaining a robust defensive posture will be especially important for owners of critical infrastructure, many of whom operate with limited financial resources.

By embedding resilience into the DNA of an organization's infrastructure, we not only protect our assets but also enhance our ability to compete and thrive in a volatile world.

---

## Key recommendations for leaders

**Invest in resilience by design**
Encourage the development of infrastructure that is inherently resilient. This includes modular systems, redundancy, and fail-safes that allow for graceful degradation and rapid recovery.

**Foster public-private collaboration**
Resilience is a shared responsibility. Governments and industries must work together to set standards, share threat intelligence, and coordinate responses to disruptions.

**Support innovation and workforce development**
Resilience requires cutting-edge technologies and a skilled workforce. Leaders should champion investments in research and development and education to build national capacity.

**Incentivize resilience through policy and regulation**
Financial and regulatory frameworks should reward organizations that prioritize resilience, much like how safety and environmental standards are incentivized today.
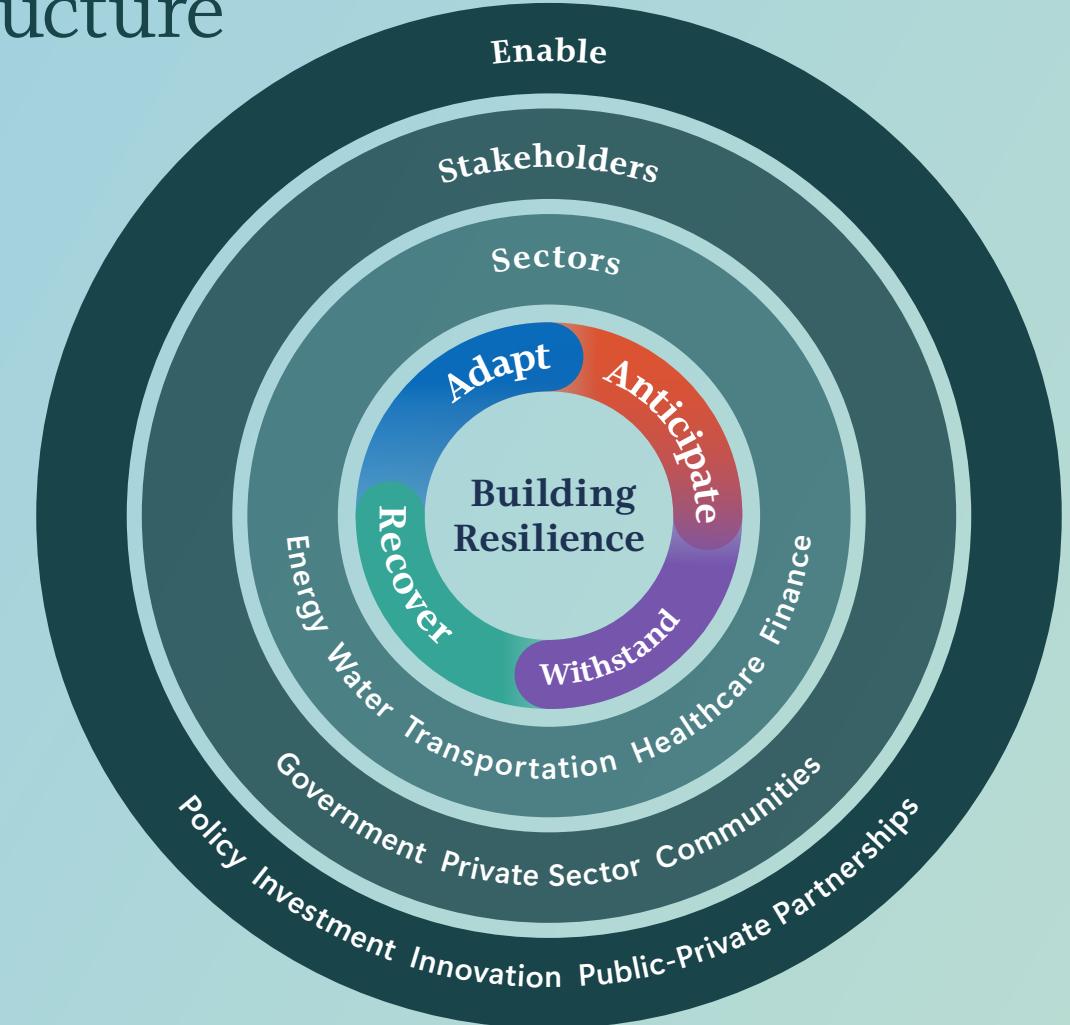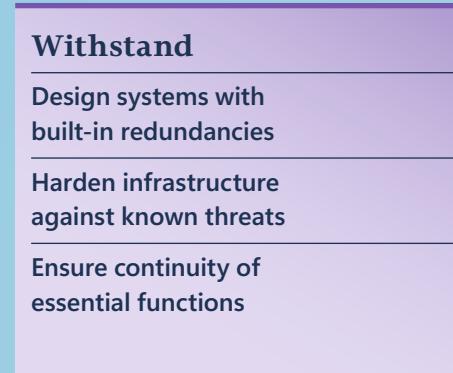
**Measure and monitor resilience**
Establish clear metrics and benchmarks to assess the resilience of critical systems. Transparency and accountability are essential for continuous improvement.

**Policy, capacity, and future readiness** continued

# Building resilience in critical infrastructure
A strategic lifecycle, four core phases...

### Anticipate

Identify vulnerabilities and emerging threats

Conduct risk assessments

Model potential disruptions

### Withstand

Design systems with built-in redundancies

Harden infrastructure against known threats

Ensure continuity of essential functions

### Recover

Rapid response and restoration protocols

Minimize downtime and service disruption

Communicate transparently with stakeholders

### Adapt

Learn from incidents and near-misses

Update systems and policies

Invest in innovation and workforce training

## Microsoft's strategic path to quantum safety

Much of modern cryptography relies on mathematical puzzles that are practically impossible for classical computers to solve—for instance, cracking the standard encryption behind a secure website or messaging app would take millions of years with today's computers.

Quantum computing is novel that can consider many possibilities at once, allowing quantum computers to process complex problems much faster than classical systems.

Quantum computing poses a serious threat to current cryptographic systems. While still an emerging technology, the expected development of a powerful cryptographically relevant quantum computer (CRQC) means that if organizations don't update our cryptography in time, we risk a scenario like the early days of the internet, when websites were on unencrypted HTTP and attackers could eavesdrop on information in transit. In the lead up to this potential data exposure, Harvest Now, Decrypt Later (HNDL) is a real concern: attackers can hoard encrypted data today so they can decrypt it in the future with quantum power.

Every organization should inventory its cryptography (keys, certificates, and protocols) and establish a roadmap to replace vulnerable algorithms with Post-Quantum Cryptography (PQC) standards as they become available. At Microsoft, there is a dedicated program to make sure our own products and services—and customers—stay safe in the quantum era. Microsoft established the Quantum Safe Program (QSP) to coordinate all its quantum security efforts across the company and achieve quantum readiness by gradually integrating PQC algorithms into Microsoft's services. As part of our efforts:

- We updated SymCrypt, Microsoft's core cryptographic library, to support new post-quantum algorithms. SymCrypt is like the engine that handles encryption under the hood in Windows, Azure, and many Microsoft products. We also enabled PQC support in Windows and Azure Linux (using SymCrypt OpenSSL).
- Microsoft Research has contributed to the design and analysis of PQC algorithms. Through blogs and publications, Microsoft shares these developments with the community, helping to lead the conversation on how to protect information in the quantum age.

Governments and industries worldwide are actively preparing for the quantum era by upgrading their cryptographic algorithms to quantum-resistant alternatives. Standards bodies like National Institute of Standards and Technology (NIST) and International Organization for Standardization (ISO) have been running global competitions to select robust PQC algorithms, and international groups are working on standards to integrate these algorithms into our software so that everyone's systems can work together. In everyday terms, it's like the world has agreed to upgrade all its locks and keys and is now in the process of implementing the change.

During the last year, multiple governments have also published guidance and requirements to spur the transition, with most identifying 2035 as the deadline for completing transition. In the United States, European Union, and Australia, changes to some of the highest risk systems should be made by 2030, while in Canada and the UK, that date is 2031.

# Recommendations

Governments play a critical role in enabling a quantum-safe future through strong collaboration with industry and effective policies. To accelerate readiness, we recommend governments take the following actions:

**Establish quantum safety as a national cybersecurity priority.** Position quantum-safe cryptography as a strategic imperative and embed it into national cybersecurity frameworks.

**Align quantum-safe strategies across jurisdictions.** Harmonize public policies, standards, and transition timelines. The G7 should lead by expanding its financial sector post-quantum cryptography workstream to align G7 members' broader quantum-safe strategies.

**Adopt international standards.** Support global standards development and avoid fragmented, region-specific approaches that hinder interoperability, innovation, and security.

**Set early and progressive timelines.** Drive action well before 2030. For instance, the US Committee on National Security Systems Policy 15 (CNSSP -15) mandates quantum-safe algorithms in all new products and services for national security systems by January 2027.

**Lead by example with transparent transition plans.** Publish and regularly update government transition roadmaps—including timelines, milestones, and budgets—to foster knowledge sharing and best practices.

**Raise awareness and build workforce capacity.** Educate the public and critical infrastructure sectors on quantum risks and readiness. Invest in skilling programs to equip the workforce for a quantum-safe transition.

**Modernize through cloud adoption** Promote cloud migration as a strategic enabler. Cloud platforms can streamline the transition by embedding quantum-safe capabilities, reducing the burden on individual organizations.

↗ **Learn more**

Post-quantum resilience: building secure foundations | Microsoft On the Issues

Quantum-safe security: Progress towards next-generation cryptographyQuantum-safe security: Progress towards next-generation cryptographyQuantum-safe security: Progress towards next-generation cryptographyQuantum-safe security: Progress towards next-generation cryptographyQuantum-safe security: Progress towards next-generation cryptography

https://quantum.microsoft.com

# Strategic vision and global commitments

**Secure Future Initiative: Progress and priorities**

Microsoft's Secure Future Initiative (SFI) is our multi-year effort to revolutionize how we design, build, test, and operate our products and services to achieve the highest security standards. Released in April 2025, the third edition of our public progress report continued our tradition of transparency, articulating improvements to Microsoft's internal security posture and sharing innovations that help better protect customers by design and by default.

As we highlight in our report, we continue to foster a robust internal security culture. Every Microsoft employee now has a Security Core Priority within their performance objectives, fostering personal accountability and a stronger security mindset. To strengthen governance, we've established a regulatory governance council of Deputy Chief Information Security Officers (dCISO) embedded across critical product and business areas, driving risk management alignment, accountability, and resilience at scale.

> **"** Transparency and clarity remain central to our mission, and through regular reports and additional guidance, we aim to share our learnings to collectively move our ecosystem toward a safer future.

At the engineering level, progress has been made across our twenty-eight aligned objectives covering six engineering pillars, protecting identities, secrets, tenants, and networks, isolating production systems, securing engineering systems, monitoring and detecting threats, and accelerating response and remediation. While there will always be more work to do, we have made meaningful progress across all areas. This structured approach aligns closely to Zero Trust architecture, enabling consistent, risk-based prioritization and continuous improvement.

We continue to deliver product innovations that translate our internal learnings into customer value, across Microsoft Azure, Microsoft 365, Windows, and our security portfolio, including Microsoft Entra, Defender, and Purview. For instance, Azure's integrated Hardware Security Modules (HSMs), Microsoft 365's Copilot Control System (CCS), and the widespread deployment of phishing-resistant MFA reflect our commitment to protecting customers. Grounded in our core principles of Secure by Design, Secure by Default, and Secure Operations, this work reinforces our mission to strengthen security across Microsoft and empower customers with solutions that are more secure out of the box.

Our intent in reporting on SFI is not only to share progress, but also to offer clear and actionable guidance through patterns and practices to customers, partners, and the broader ecosystem. Transparency and clarity remain central to our mission, and through regular reports and additional guidance, we aim to share our learnings to collectively move our ecosystem toward a safer future.

**⬈ Learn more**



Secure Future Initiative | Microsoft Trust Center

SFI April 2025 Progress Report

SFI Customer Guidance: Patterns and Practices | Microsoft Security Blog

## Microsoft's commitment to strengthening global cybersecurity

Microsoft is deeply committed to supporting the global effort to counter cyber threats by fostering strong partnerships with governments and advocating for cybersecurity laws and regulations that promote a safer digital ecosystem for all.

### A regional focus: Europe's cybersecurity imperative

The EU has enacted the Cyber Resilience Act (CRA), a landmark regulation poised to become the gold standard for cybersecurity, much like the General Data Protection Regulation (GDPR) did for data privacy. The CRA is expected to elevate global security standards, influencing how secure products are built even beyond Europe's borders.

But regulation alone isn't enough. Protecting Europe's digital infrastructure requires deep collaboration between governments and industry. Microsoft is actively contributing to this shared mission by:

- Appointing a European dCISO to its cybersecurity governance council.
- Launching a European Security Program to provide EU governments with real-time threat intelligence and response capabilities.
- Contributing guidance to help manufacturers comply with the CRA—including the development of harmonized standards by European Standards Organizations and EU Commission guidance and supporting legislation through the CRA Expert Group.

These efforts reflect Microsoft's belief that collective security is only possible through trusted partnerships and shared responsibility.

### Global trends: Cybersecurity policies and laws

As governments accelerate efforts to manage cyber risk through new laws and policies, two key trends have emerged:

- **Regulatory expansion and enforcement**
  Governments are shifting from voluntary guidelines to enforceable standards, emphasizing accountability, risk management, and timely incident reporting.
- **Securing the digital supply chain**
  New mandates are driving secure by design principles, transparency through clearer support lifecycles and forward leaning efforts such as encouraging the generation of SBOMs, and robust post-market monitoring.

While regulatory expansion and enforcement and efforts to secure the digital supply chain are well intended, they can also introduce complexity. Fragmented regulatory frameworks can slow down incident response and ultimately weaken defenses.

As a global company, Microsoft sees firsthand how inconsistent cybersecurity regulations across jurisdictions can hinder resilience. That is why efforts to establish international regulatory cooperation, such as the effort led by Germany and South Korea, are important. To truly strengthen global cybersecurity, governments must pursue harmonized, risk-based approaches that promote interoperability and reduce duplication.

Key opportunities for regulatory alignment include:

- Incident reporting: Standardizing timelines, definitions, thresholds, and formats to enable faster, coordinated responses.
- Emerging technologies: Aligning approaches to AI and post-quantum cryptography to avoid innovation silos.
- Supply chains and vulnerability management: Encouraging technology suppliers to inventory their supply chain dependencies and strengthen practices of coordinated vulnerability disclosure to improve the identification, communication, and remediation of vulnerabilities across the supply chain promptly.

Microsoft urges governments to prioritize regulatory harmonization and supports the Organisation for Economic Co-operation and Development (OECD) as a key convener in this effort. The organization's multilateral structure and digital security expertise make it well-positioned to:

- Develop principles for regulatory alignment.
- Establish a forum for regulators and experts representing various jurisdictions across the multistakeholder cybersecurity community.
- Commission research to map overlaps and gaps in global cybersecurity policy.

Earlier this year, Microsoft joined dozens of technology leaders in signing an open letter to the Group of Seven (G7) and OECD, calling for coordinated action to reduce cyber risk and foster innovation.

⤴ **Learn more**

Microsoft launches new European Security Program | Microsoft On the Issues

EU Data Resiliency | Microsoft Trust Center

The CyberPeace Institute is helping NGOs defend themselves—before it's too late (August 2025)
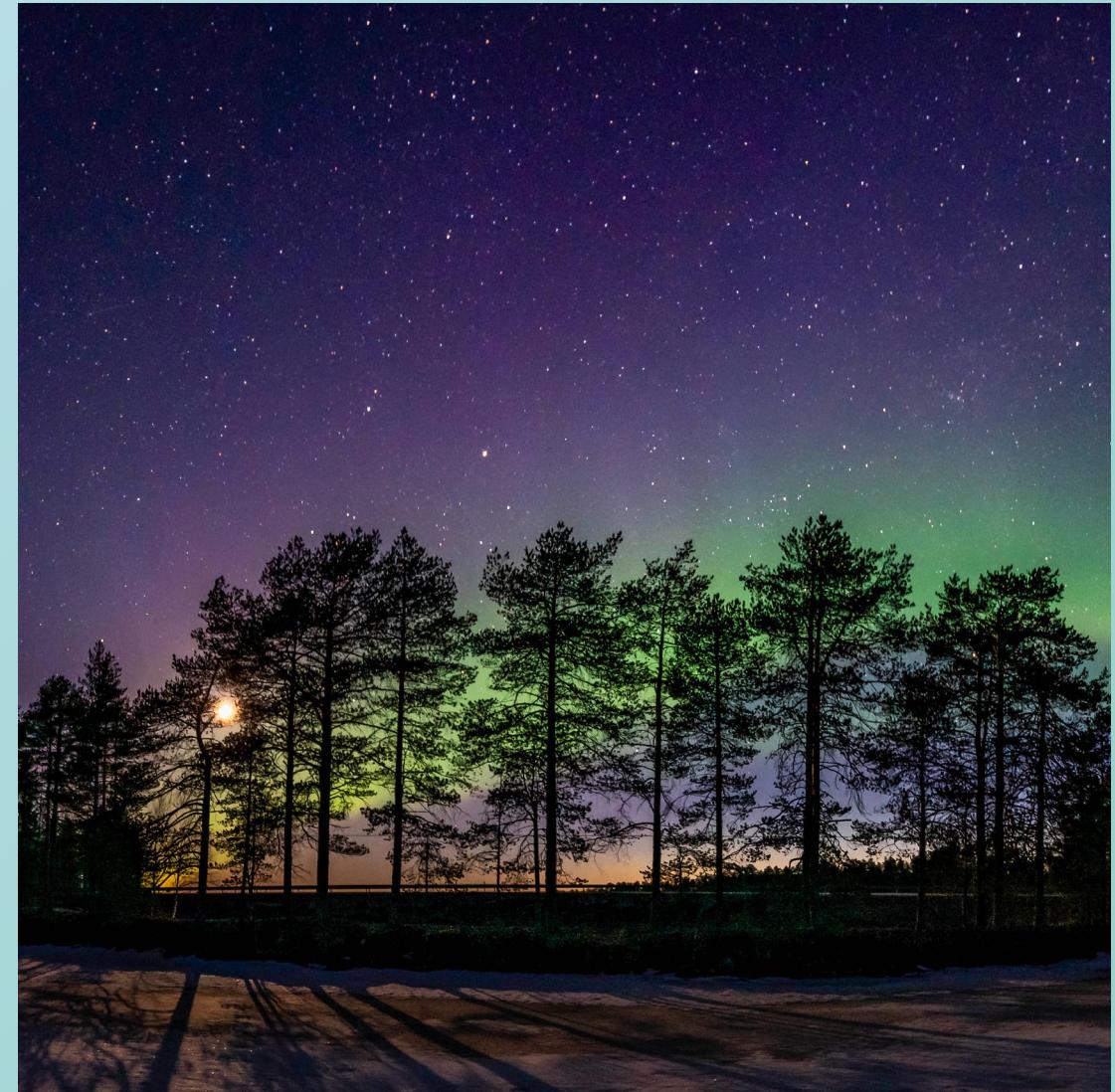
# Closing

As global regulatory frameworks evolve and legislative trends reshape the cybersecurity landscape, one truth remains constant: security is a shared responsibility.

Governments, industry leaders, civil society, and individual users each play a vital role in shaping a resilient digital ecosystem. The insights and data presented throughout this report underscore the urgency of collaboration—not only across borders but across sectors and disciplines.

Our commitment to lighting the path to a secure future is more than a campaign theme—it is a call to action. We believe that transparency, interoperability, and harmonized standards are foundational to progress. Whether through our threat intelligence, policy advocacy, or engineering innovations, we aim to empower defenders and decision-makers alike.

Thank you for reading this year's Microsoft Digital Defense Report. We invite you to explore our companion resources, share your feedback, and join us in building a secure, more trustworthy digital world.

# Appendix

# Glossary

**Access broker**

A cybercriminal who gains unauthorized access to organizations and sells that access to other criminals, enabling further attacks such as ransomware or data theft.

**AI deepfake**

Artificial Intelligence-generated audio, video, or images that convincingly mimic real people or events can be used to impersonate individuals, fabricate scenarios, or manipulate public perception—often contributing to fraud, misinformation, or disinformation.

**Attack surface**

The total set of points where an unauthorized user can try to enter or extract data from an environment.

**BEC (Business Email Compromise)**

A targeted attack where criminals gain access to business email accounts to defraud organizations, often by manipulating financial transactions.

**Botnet**

A network of computers infected with malware and controlled as a group to perform malicious activities, such as launching attacks or sending spam.

**Cloud security**

Protecting data, applications, and systems hosted in cloud environments. As organizations move to the cloud, attackers increasingly target cloud assets and identities.

**Cloud workload**

Applications, services, or processes running in a cloud environment, which can be targeted by attackers.

**Container (in cybersecurity context)**

A lightweight, standalone package of software that includes everything needed to run an application. Containers are widely used in cloud environments and can be targeted by attackers if not properly secured.

**Credential theft**

Stealing usernames, passwords, or other authentication information to gain unauthorized access to systems or data.

**Critical infrastructure**

Essential systems and assets (energy, water, transportation, healthcare, etc.) whose disruption would have significant societal impact.

**Cyber mercenary**

A private entity that sells hacking tools or services to governments or criminals, often operating in legal gray zones.

**Cyber resilience**

The ability of an organization to anticipate, withstand, recover from, and adapt to cyberattacks or disruptions.

**Cyber-enabled influence operations**

Efforts by threat actors to manipulate public opinion or behavior using digital tools, such as social media, fake news, or deepfakes.

**Data exfiltration**

The unauthorized transfer or theft of data from an organization, often as part of a cyberattack.

**Data theft**

Stealing sensitive or valuable information, such as intellectual property, personal data, or financial records.

**Device code phishing**

A phishing technique where attackers trick users into entering authentication codes on fake portals, allowing them to hijack accounts.

**Endpoint**

Any device (such as a computer, smartphone, or server) that connects to a network and can be targeted by cyberattacks.

**Espionage**

The act of spying to obtain confidential information, often for political, economic, or military advantage.

**Exploit**

A method or tool used by attackers to take advantage of vulnerabilities in software or systems.

**Fraud**

Deceptive practices intended to gain financial or personal benefit, often involving manipulation or impersonation.

**Human-operated attack**

A cyberattack where humans, rather than automated tools, actively control the intrusion, often adapting tactics in real time.

**Human-operated ransomware**

A ransomware attack in which cybercriminals actively control the intrusion, moving through networks, stealing data, and manually deploying ransomware for maximum impact. These attacks are more targeted and damaging than automated ransomware, often combining extortion with data theft or disruption of critical services.

**Glossary** continued

**Identity compromise**
When an attacker gains control of a user's digital identity, allowing unauthorized access to systems or data.

**Identity platform**
A system or service that manages digital identities, authentication, and access controls for users and devices.

**Incident response (IR)**
A structured approach to managing and mitigating the impact of cybersecurity incidents.

**Infostealer**
Malware designed to collect credentials, tokens, and other sensitive information from infected devices.

**Influence operations**
Coordinated efforts to affect public perception or behavior, often using digital channels and sometimes involving misinformation or manipulation.

**Infrastructure building**
A tactic where attackers use compromised systems to stage further attacks against other targets, often creating a base for future operations.

**Insider threat**
A risk posed by individuals within an organization who may intentionally or unintentionally cause harm by leaking data or facilitating attacks.

**LLM (Large Language Model)**
A type of AI model trained on vast amounts of text data to understand and generate human-like language. LLMs can answer questions, summarize documents, and assist with decision-making, but can also be targeted or manipulated by cyber attackers.

**Malvertising**
Malicious advertising that delivers malware to users through deceptive online ads.

**Malware**
Software designed to disrupt, damage, or gain unauthorized access to computer systems.

**MFA (multifactor authentication)**
A security process requiring two or more verification factors to access systems or data.

**Mule herding**
The recruitment and management of individuals ("money mules") who move or launder stolen funds on behalf of cybercriminals.

**Nation-state actor**
A cyber threat actor sponsored or directed by a government, often targeting other countries for espionage, disruption, or influence.

**Password spray attack**
A technique where attackers try common passwords against many accounts to gain unauthorized access.

**Phishing**
A cyberattack where attackers impersonate trusted entities to trick individuals into revealing sensitive information.

**Post-quantum cryptography (PQC)**
Encryption methods designed to be secure against quantum computing attacks.

**Prompt injection**
A type of attack on AI systems where malicious instructions are hidden in user input or data, causing the AI to behave in unintended or harmful ways.

**Quantum computing**
Advanced computing technology that could break current encryption methods, requiring new security standards.

**Ransomware**
Malicious software that encrypts data and demands payment for its release.

**Remote access tool**
Software that allows remote control of a computer, often used legitimately but also abused by attackers.

**Resilience by design**
Building systems and processes that can withstand, recover from, and adapt to cyberattacks or disruptions.

**Social engineering**
Manipulating people into performing actions or divulging confidential information, often used in phishing and fraud.

**SLM (Small Language Model)**
A more compact version of a language model, designed to perform language-related tasks efficiently with fewer computational resources. SLMs are often used for specific, focused applications where speed and efficiency are important, but they may have more limited capabilities compared to LLMs.

**Supply chain attack**
Targeting less secure elements in an organization's supply chain (vendors, partners) to gain access to the primary organization.

**Threat intelligence**
Information about current and emerging cyber threats, used to inform security strategies and improve defenses.

**Token theft**
Stealing authentication tokens (digital keys) to gain unauthorized access without needing a password.

**Vishing**
Voice phishing; using phone calls to trick individuals into revealing sensitive information or performing risky actions.

**Virtual credit card (VCC)**
A digital payment card generated for online transactions, often with unique details and limited lifespan to reduce fraud risk.

**Workload identities**
Digital identities assigned to applications, services, or automated processes (not people), which can be targeted by attackers if not properly secured.

# Contributing teams

### AI Safety and Security

AI Safety and Security is responsible for all aspects of as well as developing and deploying secure and safe AI, including pre-launch evaluation, incident response, building safety infrastructure, training, research, and policy.

### Central Fraud and Abuse Risk (CFAR)

Central Fraud and Abuse Risk detects and responds to nation-state actors, criminal syndicates, and common cyber criminals who wish to cause financial and reputational harm to Microsoft, its customers, and partners. The team also partners with law enforcement, industry affiliates, and customers to share fraud insights to make the world safer for all.

### Cloud Ecosystem Security

Cloud Ecosystem Security is responsible for the core cloud security platform, data security, compliance, governance, and privacy. The team also leads AI-powered threat and data intelligence, as well as AI security research and development.

### Corporate Standards Group

Corporate Standards Group represents Microsoft in multistakeholder organizations that are establishing standards on issues such as cybersecurity, AI, and data. The team works with governments, civil society, academia, and industry to create coherent international practices that can be used to develop, evaluate, and manage trustworthy technology.

### Customer Security and Trust

Customer Security and Trust drives continuous improvement of customer security in Microsoft products and online services. Working with engineering and security teams across the company, the team ensures compliance, enhances security, and drives transparency to protect customers and the global ecosystem.

### Cybersecurity Policy and Diplomacy (CPD)

Cybersecurity Policy and Diplomacy works on strengthening global cybersecurity by promoting responsible industry and state behaviour in cyberspace through sustained diplomatic and policy engagement and multistakeholder partnerships.

### Digital Crimes Unit (DCU)

The Digital Crimes Unit has been fighting cybercrime, protecting individuals and organizations, and safeguarding the integrity of Microsoft services since 2008, through strategic partnerships and engagements, the seizure of criminal infrastructure, and the disruption of global cyber threats and criminal networks.

### Digital Security & Resilience
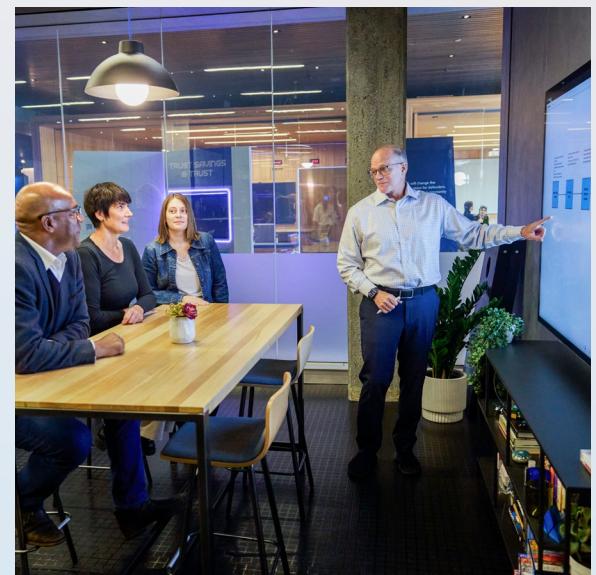
Digital Security & Resilience is dedicated to enabling Microsoft to build the most trusted devices and services, while keeping our company and customers protected.

### Enterprise & Security

Enterprise & Security provides platform technologies and solutions to manage and harden platforms against attacks. The team also empowers company-wide security initiatives in Zero Trust, secure identity, secure devices, secure supply chain, and scale management from cloud.

### European Government Affairs

European Government Affairs represents Microsoft's positions towards European political institutions, governments and other political actors. The team oversees a large variety of digital policies across Europe, including AI, cloud, sustainability and cybersecurity policy.

**Contributing teams** continued

## Extended Security Posture Management

Extended Security Posture Management builds cross-domain pre-breach security solutions for attack surface management and threat exposure reduction. The team brings together posture management capabilities for devices, identities, cloud, and applications into a set of consolidated products serving security leaders and their teams.

## GUARD Detection Engineering team in the Cyber Defense Operations (CDO) under the CISO organization

The Security CTO office mission is to drive innovation, identify gaps across the security division, and promote opportunities related to organizational growth and talent. The team identifies systemic opportunities not only in product strategy but also across the division and Microsoft.

## Identity & Network Access

Identity & Network Access innovates and builds solutions that manage and govern identities and access, including the consumer sign-in experience.

## Insights, Data Engineering, Analytics, and Systems (IDEAS) and Insights, Data Engineering, and Analytics Momentum and Storytelling

Insights, Data Engineering, Analytics, and Systems (IDEAS) and Insights, Data Engineering, and Analytics Momentum and Storytelling curates metrics used in non-financial public disclosures. The team also helps craft the messages around those metrics, and ensures that the messages align with Microsoft's perspectives.

## Microsoft Defender Experts

Microsoft Defender Experts manage Threat Hunting and Extended Detection and Response service that proactively looks for threats 24/7/365 using Microsoft Defender data.

## Microsoft Incident Response—the Detection and Response Team (DART)

Microsoft Incident Response—the Detection and Response Team provides incident hunting, cyber resilience, and threat intelligence services to customers. The team maintains strategic partnerships with security organizations, governments, and internal Microsoft groups.

## Microsoft Threat Analysis Center

Microsoft Threat Analysis Center identifies and analyzes nation-state threats and influence operations, integrating intelligence with geopolitical context to deliver timely insights to Microsoft and its customers for effective response and protection.

## Microsoft Threat Intelligence Center (MSTIC)

Microsoft Threat Intelligence Center (MSTIC) discovers, tracks, and disrupts sophisticated cyber threat actors to protect Microsoft and its customers. MSTIC produces actor-centric threat intelligence and delivers high quality finished intelligence across Microsoft's security solutions.

## Microsoft Threat Protection Research

Microsoft Threat Protection Research combines the trillions of signals we see daily with world class security research into highly sophisticated and emerging threats to deliver prevention, detection,

response, and automated disruption capabilities to more than 1 billion devices across all domains (Endpoint, Identity, Office, Cloud, IoT/OT.)

## National Security Officers

National Security Officers advise on best practice cyber guidelines, support driving compliancy and certification of Microsoft's services and products in countries with particular national requirements.

## Office of Responsible AI (ORA)

Office of Responsible AI (ORA) collaborates with stakeholders across Microsoft to develop policies, practices, and governance systems to uphold our AI principles. ORA also helps to shape the new laws needed to ensure that the promise of AI technology is realized for the benefit of society at large.

## Office of the Chief Scientific Officer

Office of the Chief Scientific Officer leads strategic initiatives at the confluence of the sciences, technology, and society, including frontier efforts in AI.

## US Government Affairs

US Government Affairs advances collaborative discussions with US federal and state government representatives, policymakers, and third-party groups, as well as the UN and other international organizations. The team oversees a large variety of policy priorities including AI, cybersecurity, cloud, sustainability and competition.

# References

## Part I: The threat landscape

**1**  Cyber Signals: Cyberthreats in K-12 and higher education | Microsoft Security Blog

**2**  Mythical Beasts and where to find them: Mapping the global spyware market and its threats to national security and human rights | Atlantic Council

**3**  Intelligence-Driven Cyber Security | Intel 471

**4**  Have I Been Pwned: Check if your email address has been exposed in a data breach

**5**  Protecting customers from Octo Tempest attacks across multiple industries | Microsoft Security Blog (July 2025)

**6**  The latest marketing tactic on LinkedIn: AI-generated faces | NPR

**7**  Cybersecurity Information Sheet: Contextualizing Deepfake Threats to Organizations | National Security Agency

**8**  Synthetic ID document fraud is exploding worldwide thanks entirely to Generative AI: here's how to stay safe | TechRadarr

**9**  Grand View Research

**10**  Harvard Kennedy School et al., "Evaluating Large Language Models' Capability to Launch Fully Automated Spear Phishing Campaigns: Validated on Human Subjects," arXiv preprint, Nov. 30, 2024. Study found AI-automated phishing can be up to 50x more profitable than traditional methods when targeting large groups

**11**  States' use of non-state actors in cyberspace | Observer Research Foundation

**12**  Nation-State Threats | Cybersecurity and Infrastructure Security Agency (CISA)

**13**  2025 Ponemon Cost of Insider Threats Global Report

**14**  Communication Compliance | Microsoft Learn

**15**  Jasper Sleet: North Korean remote IT workers' evolving tactics to infiltrate organizations | Microsoft Security Blog

**16**  Protecting Quantum Science and Technology | FBI.gov

## Part II – The defense landscape

**17**  Secure employee access in the age of AI | Microsoft (2025)

**18**  International Counter Ransomware Initiative

**19**  Ransomware Task Force (RTF) | Institute for Security + Technology

**20**  The Pall Mall Process Code of Practice for States | GOV.UK

**21**  A global clearing house for real-time sharing of scam and fraud signals | Global Signal Exchange

**22**  International Scammers Steal Over $1 Trillion in 12 Months in Global State of Scams Report | Global Anti-Scam Alliance

**23**  Scams Prevention Framework – Protecting Australians from scams | Treasury.gov.au

**24**  APP fraud reimbursement protections | Payment Systems Regulator

**25**  Combatting Scams | Monetary Authority of Singapore

**26**  Mitigating AI Security Risks for UK Prosperity & National Resilience | LASR

**27**  Frontier Governance Framework | Microsoft (February 2025)

**28**  AI Literacy Starting Guide | Microsoft (June 2025)

**29**  The EU AI Act: Prohibited Practices | Microsoft (January 2025)

**30**  The General-Purpose AI Code of Practice: Shaping Europe's digital future | Microsoft AI Transparency Report

**31**  Cyber–Physical Resilience: Evolution of Concept, Indicators, and Legal Frameworks

**32**  Fortifying the Resilience of our Critical Infrastructure

Microsoft

# Microsoft Digital Defense Report 2025

**Lighting the path to a secure future**

For more news on cybersecurity, visit:
**https://microsoft.com/corporate-responsibility/cybersecurity**

For more report insights, visit:
**https://microsoft.com/mddr**

For more news on cybersecurity
policy, follow us on LinkedIn:
**https://aka.ms/MOILinkedin**

For insights and trends for security leaders, visit:
**https://www.microsoft.com/security/security-insider**

**A Microsoft Threat Intelligence report**
October 2025