
Microsoft
Microsoft Digital Defense Report 2024
The foundations and new frontiers of cybersecurity
A Microsoft Threat Intelligence report

Complex, challenging, and increasingly dangerous

- 600 million attacks every day
- Meeting the challenge: focus and commitment from individual users to executive level
- Governments must impose deterrent consequences
- AI can help defenders restore balance
- “All hands on deck” commitment

“This is a consequential time.”

Satya Nadella, Microsoft CEO

Our presence in the digital ecosystem positions us to observe key trends in cybersecurity. Microsoft's perspectives on cybersecurity are framed through **50 years of experience and insight.**

Society | Microsoft stakeholders | Microsoft Customers

Microsoft's unique vantage point

Billions of customers globally, from a broad and diverse spectrum of organizations, and consumers.

78 trillion security signals per day

1,500 unique threat groups tracked

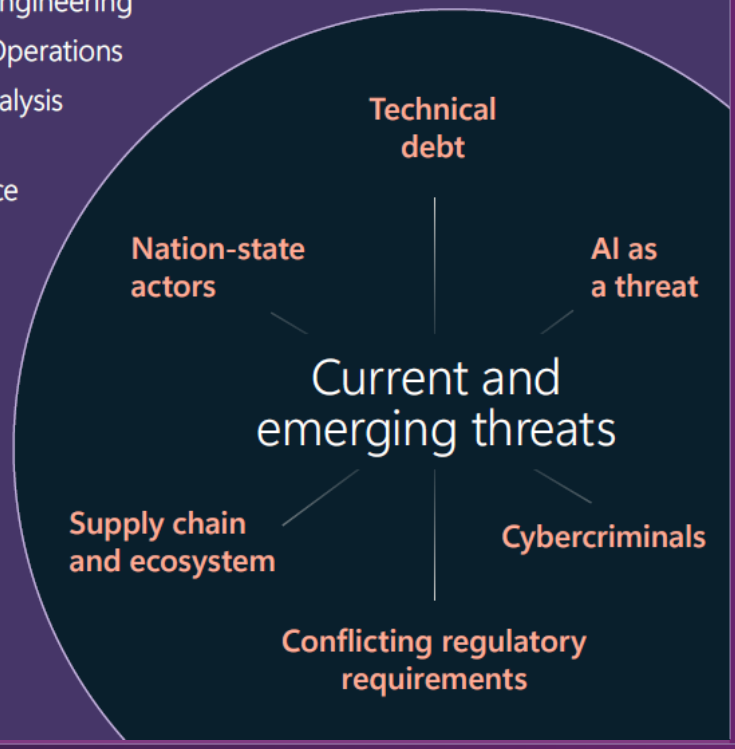
Microsoft's cybersecurity approach

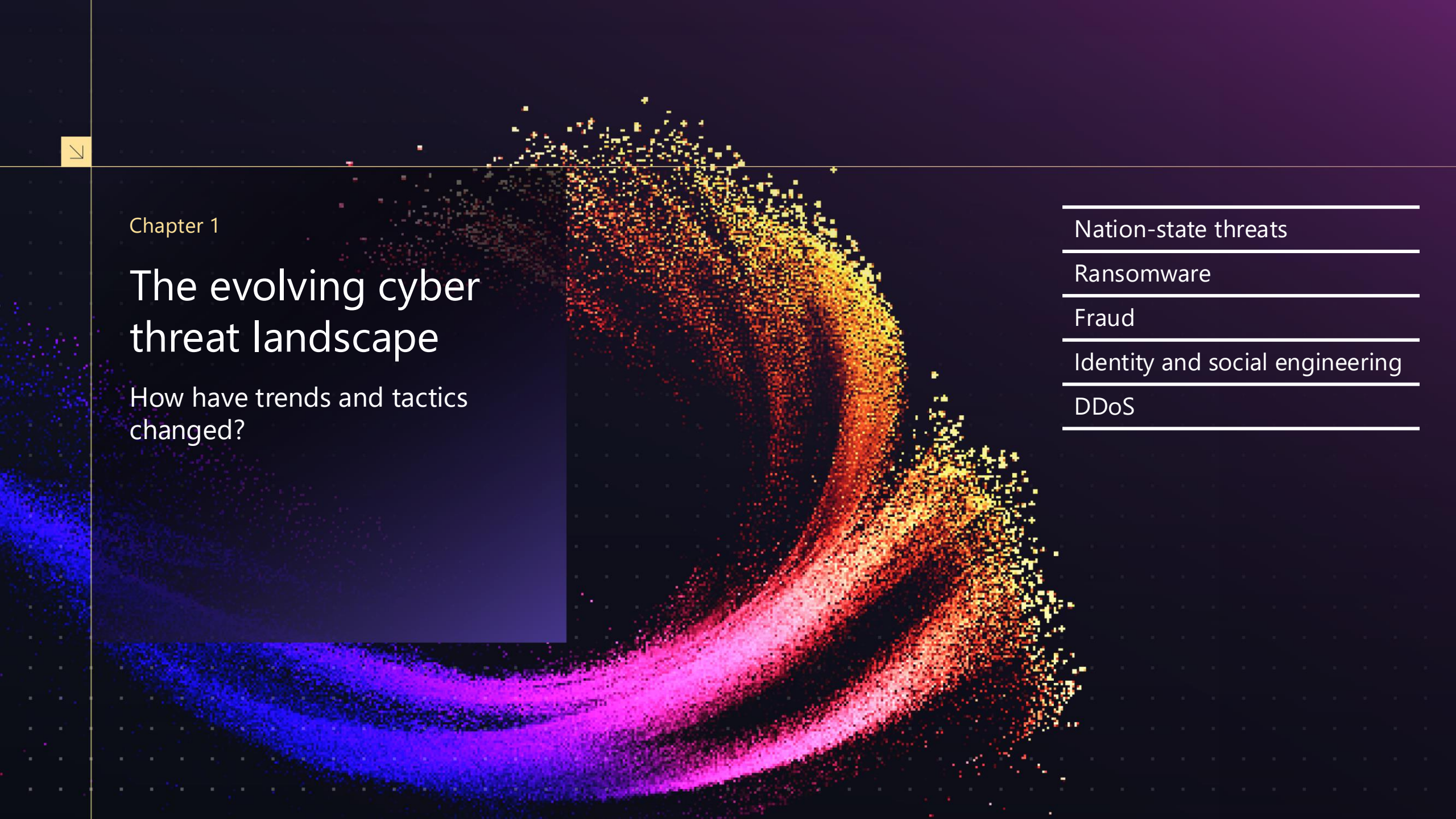
Microsoft security investments

- AI Red Teams
- Defending Democracy
- Detection and Response
- Digital Crimes
- Digital Safety
- Incident Response
- National Security
- Physical Security
- Public Awareness and Education
- Responsible AI
- Security Engineering
- Security Operations
- Threat Analysis
- Threat Intelligence

34,000 dedicated security engineers

focused full-time on the largest cybersecurity engineering project in the history of digital technology.





Chapter 1

The evolving cyber threat landscape

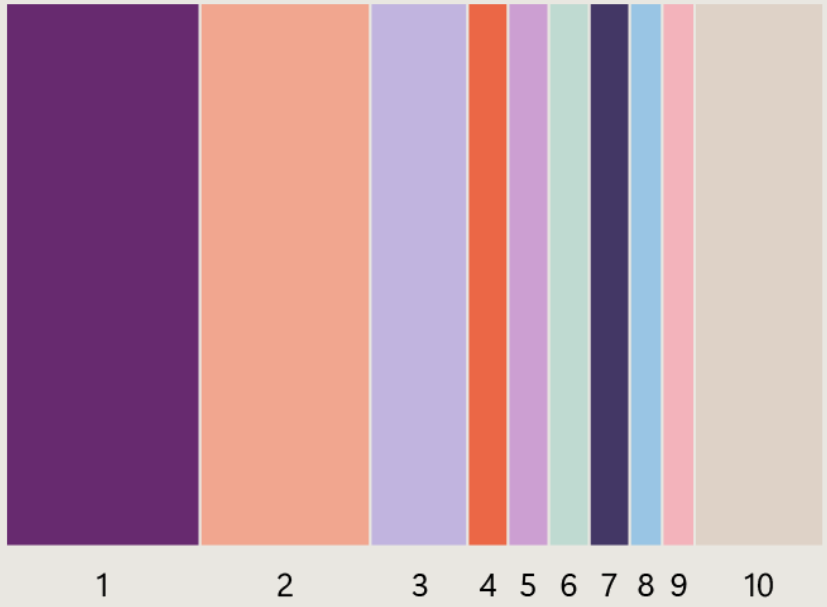
How have trends and tactics changed?

- Nation-state threats
- Ransomware
- Fraud
- Identity and social engineering
- DDoS

Nation-state threat activity by the numbers

- State-affiliated threat actors played a persistent supporting role in broader geopolitical conflicts.
- The Education and Research sector became the second most targeted by nation-state threat actors.

Top 10 targeted sectors worldwide



Sector		Percentage
1	IT	24%
2	Education and Research	21%
3	Government	12%
4	Think tanks and NGOs	5%
5	Transportation	5%
6	Consumer Retail	5%
7	Finance	5%
8	Manufacturing	4%
9	Communications	4%
10	All others	16%

Threat actors from Russia, China, Iran, and North Korea pursued access to IT products and services, in part to conduct supply chain attacks against government and other sensitive organizations.

Source: Microsoft Threat Intelligence, nation-state notification data

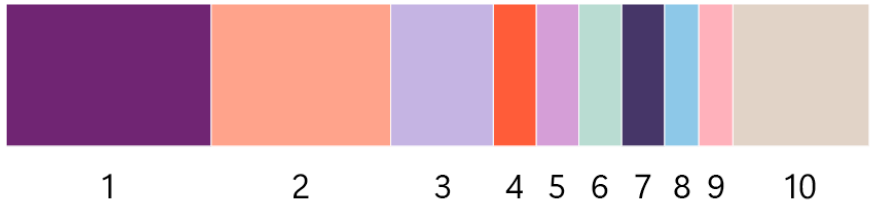
Source: Microsoft Threat Intelligence

Nation-state threat activity by the numbers

- State-affiliated threat actors played a persistent supporting role in broader geopolitical conflicts.
- The Education and Research sector became the second most targeted by nation-state threat actors.

Threat actors from Russia, China, Iran, and North Korea pursued access to IT products and services, in part to conduct supply chain attacks against government and other sensitive organizations.

Top 10 targeted sectors worldwide



	Sector	Percentage
1	IT	24%
2	Education and Research	21%
3	Government	12%
4	Think tanks and NGOs	5%
5	Transportation	5%
6	Consumer Retail	5%
7	Finance	5%
8	Manufacturing	4%
9	Communications	4%
10	All others	16%

Source: Microsoft Threat Intelligence

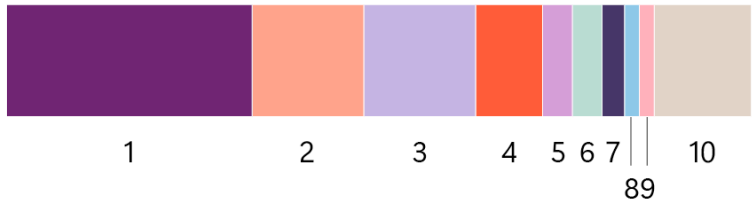
Russia: Nation-state threat actor activity

Targeting by region

Sector		Percentage
1	Europe & Central Asia	68%
2	North America	20%
3	Middle East & North Africa	5%
4	East Asia & Pacific	3%
5	Latin America & Caribbean	3%
6	South Asia	1%
7	Sub-Saharan Africa	1%

Approximately 75% of targets were in Ukraine or a NATO member state, as Moscow seeks to collect intelligence on the West's policies on the war. Ukraine remains the country most targeted by Russian actors.

Most targeted sectors



Sector		Percentage
1	Government	33%
2	IT	15%
3	Think tanks and NGOs	15%
4	Education and Research	9%
5	Inter-governmental organization	4%
6	Defense Industry	4%
7	Transportation	3%
8	Energy	2%
9	Media	2%
10	All others	13%

Blizzard

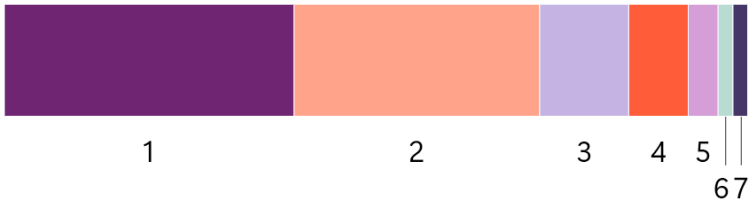
Russia

Russian actors focused their targeting against European and North American government agencies and think tanks, likely for intelligence collection related to the war in Ukraine. Actors like Midnight Blizzard also targeted the IT sector, suggesting it was in part planning supply chain attacks to gain access to these companies' client's networks for follow-on operations.

Source: Microsoft Threat Intelligence

China: Nation-state threat actor activity

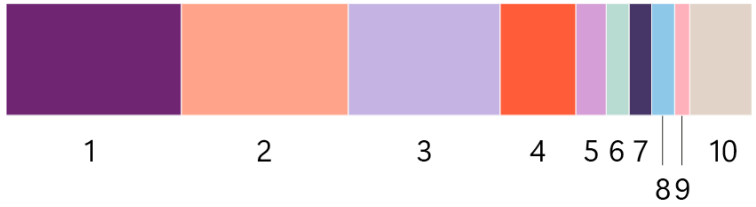
Targeting by region



Sector		Percentage
1	East Asia & Pacific	39%
2	North America	33%
3	Europe & Central Asia	12%
4	Latin America & Caribbean	8%
5	South Asia	4%
6	Middle East & North Africa	2%
7	Sub-Saharan Africa	2%

Chinese threat actors' targeting efforts remain similar to the last few years in terms of geographies targeted and intensity of targeting per location. While numerous threat actors target the United States across a wide variety of sectors, targeting in Taiwan is largely limited to one threat actor, Flax Typhoon.

Most targeted sectors



Sector		Percentage
1	IT	24%
2	Education and Research	22%
3	Government	20%
4	Think tanks and NGOs	10%
5	Manufacturing	4%
6	Defense Industry	3%
7	Communications	3%
8	Finance	3%
9	Transportation	2%
10	All others	9%

Typhoon

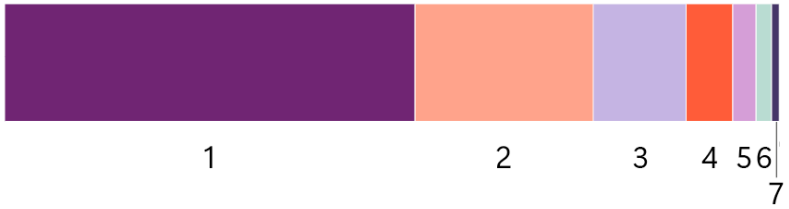
China

Most Chinese threat activity is for intelligence collection purposes and was especially prevalent in ASEAN countries around the South China Sea. Granite Typhoon and Raspberry Typhoon were the most active in the region, while Nylon Typhoon continued to target government and foreign affairs entities globally.

Source: Microsoft Threat Intelligence

Iran: Nation-state threat actor activity

Targeting by region



Sector		Percentage
1	Middle East & North Africa	53%
2	North America	23%
3	Europe & Central Asia	12%
4	South Asia	6%
5	East Asia & Pacific	3%
6	Latin America & Caribbean	2%
7	Sub-Saharan Africa	1%

Iran placed significant focus on Israel, especially after the outbreak of the Israel-Hamas war. Iranian actors continued to target the US and Gulf countries, including the UAE and Bahrain, in part because of their normalization of ties with Israel and Tehran's perception that they are both enabling Israel's war efforts.

Most targeted sectors

Sector		Percentage
1	Education and Research	19%
2	IT	11%
3	Government	7%
4	Transportation	6%
5	Finance	4%
6	Communications	4%
7	Energy	3%
8	Commercial Facilities	3%
9	Manufacturing	3%
10	All others	42%

Sandstorm

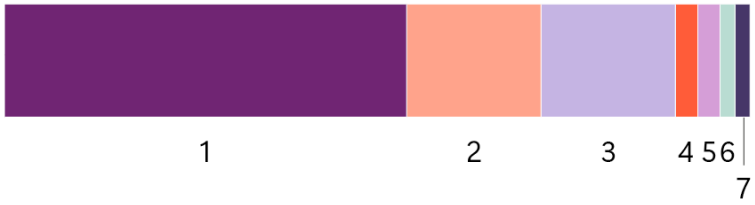
Iran

Iranian targeting focused on education, IT, and government as part of strategic intelligence collection. Iranian actors often target the IT sector to gain access to downstream customers, including those in government and the defense industrial base (DIB). "All others" includes media and think tanks or NGOs, which Iran often targets to gain insights into dissidents, activists, and persons who can impact policymaking.

Source: Microsoft Threat Intelligence

North Korea: Nation-state threat actor activity

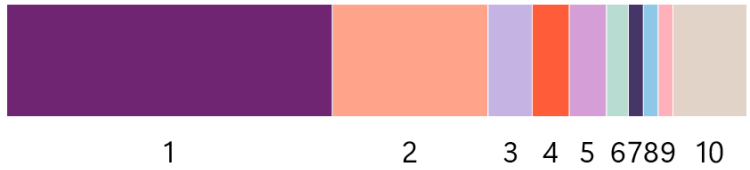
Targeting by region



	Sector	Percentage
1	North America	54%
2	East Asia & Pacific	18%
3	Europe & Central Asia	18%
4	Latin America & Caribbean	3%
5	Middle East & North Africa	3%
6	South Asia	2%
7	Sub-Saharan Africa	2%

The United States remained the most heavily targeted country by North Korean threat actors, but the United Kingdom rose up the ranks this year to second place. There were an additional 44 countries targeted by North Korean threat actors.

Most targeted sectors



	Sector	Percentage
1	IT	44%
2	Education and Research	21%
3	Manufacturing	6%
4	Consumer Retail	5%
5	Finance	5%
6	Think tanks and NGOs	3%
7	Communications	2%
8	Government	2%
9	Health	2%
10	All others	10%

Sleet

North Korea

North Korean threat actors targeted the IT sector the most, particularly to conduct increasingly sophisticated software supply chain attacks. They also continued to heavily target experts in the education sector for intelligence collection. The "All others" category comprised seven other sectors.

Source: Microsoft Threat Intelligence

Nation-state threat activity by the numbers

Russia

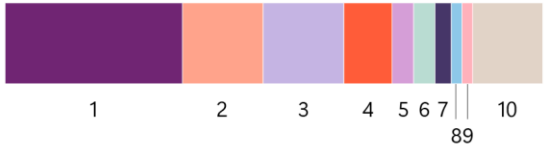
Nation-state threat actor activity

Targeting by region

	Sector	Percentage
1	Europe & Central Asia	68%
2	North America	20%
3	Middle East & North Africa	5%
4	East Asia & Pacific	3%
5	Latin America & Caribbean	3%
6	South Asia	1%
7	Sub-Saharan Africa	1%

Approximately 75% of targets were in Ukraine or a NATO member state, as Moscow seeks to collect intelligence on the West's policies on the war. Ukraine remains the country most targeted by Russian actors.

Most targeted sectors



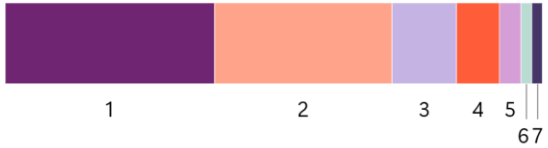
	Sector	Percentage
1	Government	33%
2	IT	15%
3	Think tanks and NGOs	15%
4	Education and Research	9%
5	Inter-governmental organization	4%
6	Defense Industry	4%
7	Transportation	3%
8	Energy	2%
9	Media	2%
10	All others	13%

Russian actors focused their targeting against European and North American government agencies and think tanks, likely for intelligence collection related to the war in Ukraine. Actors like Midnight Blizzard also targeted the IT sector, suggesting it was in part planning supply-chain attacks to gain access to these companies' client's networks for follow-on operations.

China

Nation-state threat actor activity

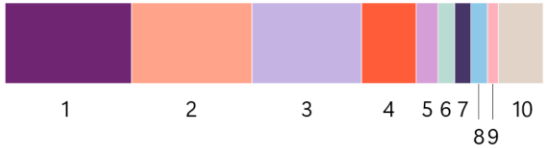
Targeting by region



	Sector	Percentage
1	East Asia & Pacific	39%
2	North America	33%
3	Europe & Central Asia	12%
4	Latin America & Caribbean	8%
5	South Asia	4%
6	Middle East & North Africa	2%
7	Sub-Saharan Africa	2%

Chinese threat actors' targeting efforts remain similar to the last few years in terms of geographies targeted and intensity of targeting per location. While numerous threat actors target the United States across a wide variety of sectors, targeting in Taiwan is largely limited to one threat actor, Flax Typhoon.

Most targeted sectors



	Sector	Percentage
1	IT	24%
2	Education and Research	22%
3	Government	20%
4	Think tanks and NGOs	10%
5	Manufacturing	4%
6	Defense Industry	3%
7	Communications	3%
8	Finance	3%
9	Transportation	2%
10	All others	9%

Most Chinese threat activity is for intelligence collection purposes and was especially prevalent in ASEAN countries around the South China Sea. Granite Typhoon and Raspberry Typhoon were the most active in the region, while Nylon Typhoon continued to target government and foreign affairs entities globally.

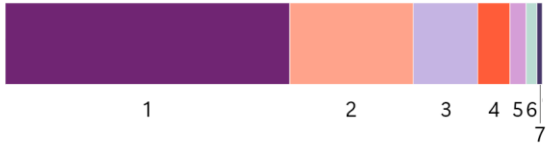
Source: Microsoft Threat Intelligence

Nation-state threat activity by the numbers

Iran

Nation-state threat actor activity

Targeting by region



	Sector	Percentage
1	Middle East & North Africa	53%
2	North America	23%
3	Europe & Central Asia	12%
4	South Asia	6%
5	East Asia & Pacific	3%
6	Latin America & Caribbean	2%
7	Sub-Saharan Africa	1%

Iran placed significant focus on Israel, especially after the outbreak of the Israel-Hamas war. Iranian actors continued to target the US and Gulf countries, including the UAE and Bahrain, in part because of their normalization of ties with Israel and Tehran's perception that they are both enabling Israel's war efforts.

Most targeted sectors

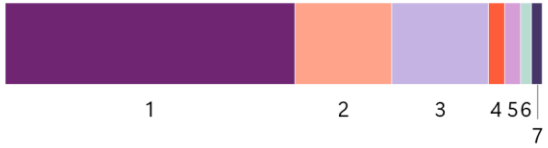
	Sector	Percentage
1	Education and Research	19%
2	IT	11%
3	Government	7%
4	Transportation	6%
5	Finance	4%
6	Communications	4%
7	Energy	3%
8	Commercial Facilities	3%
9	Manufacturing	3%
10	All others	42%

Iranian targeting focused on education, IT, and government as part of strategic intelligence collection. Iranian actors often target the IT sector to gain access to downstream customers, including those in government and the defense industrial base (DIB). "Other" includes media and think tanks or NGOs, which Iran often targets to gain insights into dissidents, activists, and persons who can impact policymaking.

North Korea

Nation-state threat actor activity

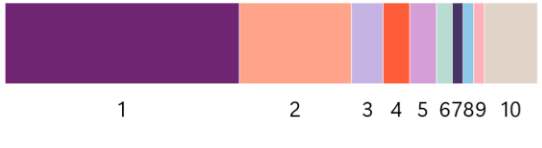
Targeting by region



	Sector	Percentage
1	North America	54%
2	East Asia & Pacific	18%
3	Europe & Central Asia	18%
4	Latin America & Caribbean	3%
5	Middle East & North Africa	3%
6	South Asia	2%
7	Sub-Saharan Africa	2%

The United States remained the most heavily targeted country by North Korean threat actors, but the United Kingdom rose up the ranks this year to second place. The "Other" category comprised 44 other countries targeted by North Korean threat actors.

Most targeted sectors



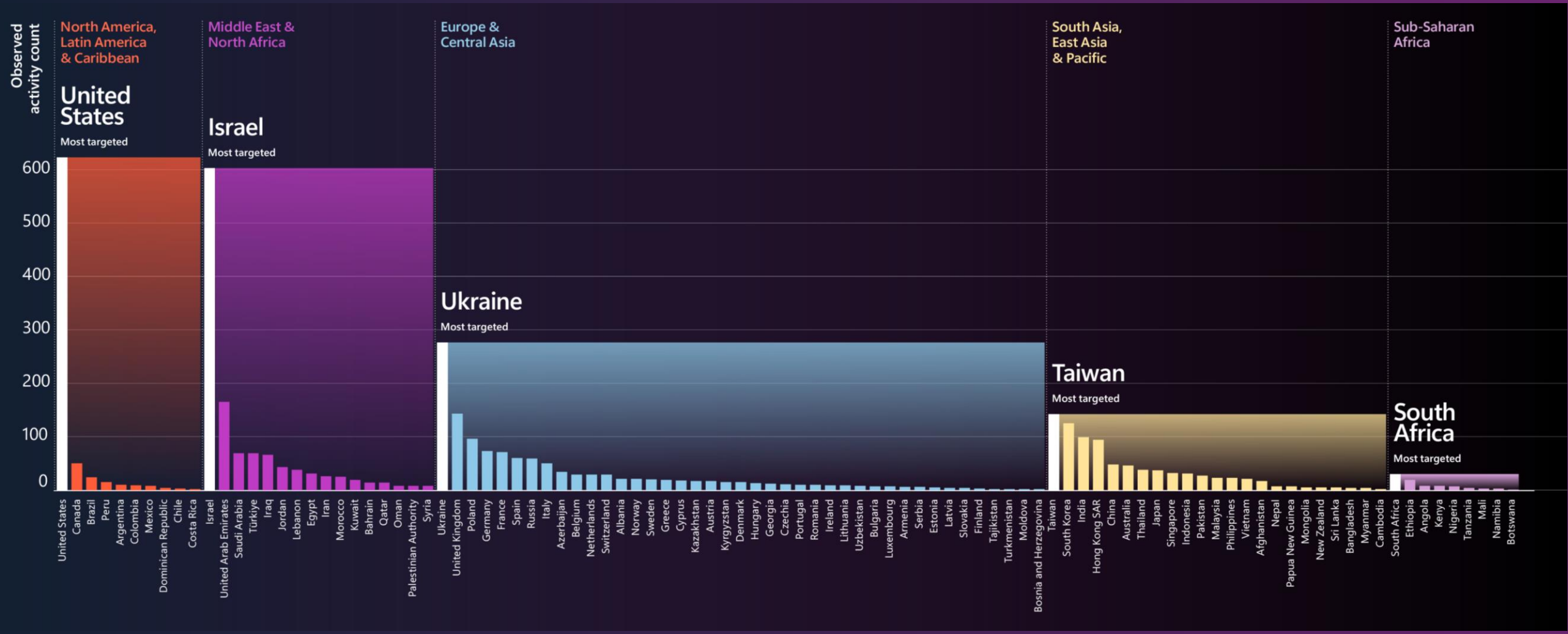
	Sector	Percentage
1	IT	44%
2	Education and Research	21%
3	Manufacturing	6%
4	Consumer Retail	5%
5	Finance	5%
6	Think tanks and NGOs	3%
7	Communications	2%
8	Government	2%
9	Health	2%
10	All others	10%

North Korean threat actors targeted the IT sector the most, particularly to conduct increasingly sophisticated software supply chain attacks. They also continued to heavily target experts in the education sector for intelligence collection. The "Other" category comprised seven other sectors.

Source: Microsoft Threat Intelligence

Nation-state threat activity by the numbers

Regional sample of activity levels observed



Source: Microsoft Threat Intelligence data

Election interference

Defending elections against influence campaigns—as well as opportunistic cybercriminal efforts—demands a collective commitment from industry, media, and governments alike.

The convergence and parallel nature of nation-state operations throughout 2024 underscores just how persistent adversarial states are in their attempts to exert influence over US elections and outcomes. Left unchecked, this poses a critical challenge to US national security and democratic resilience.

Election-related influence operations timeline

China (December 22, 2023)

PRC-linked influence actor Taizi Flood uses AI-generated audio files to allege then Taiwanese Democratic Progressive Party presidential candidate was an informant in the 1980s.

China (January 13, 2024)

Taizi Flood promotes faked AI-generated audio recording of former presidential candidate and Foxconn founder Terry Gou endorsing then Taiwanese Nationalist Party presidential candidate Hou Yu-ih.

Russia (February 23, 2024)

Russia-affiliated actor Ruza Flood registers a series of US election-themed news websites. The websites are amplified over social media by inauthentic accounts using website redirect networks to mask the actors' infrastructure and likely use AI tools to generate content.

Russia (April 19, 2024)

Russia-affiliated influence actor Storm-1516 produces fake video that attempts to frame Ukraine for interference in the 2024 US presidential election.

China (May 2024)

Sophisticated PRC-linked sockpuppet accounts position on new social media platforms to spread divisive messaging, particularly surrounding protests on US college campuses ahead of the US presidential election.

Iran (June 15, 2024)

Iran sends spear phish to presidential campaign, likely in preparation stage for influence operations targeting the US elections. (Source: Microsoft data)

China (July 2024)

July 10: Deceptively edited short-form video from PRC-linked sockpuppet account masquerading as US conservative voter reaches 1.5 million views.

July 13: PRC state media foment speculation of "deep state involvement" in Trump attempted assassination.

Presidential elections
Taiwan
Jan 2024

Presidential elections
US
Nov 2024

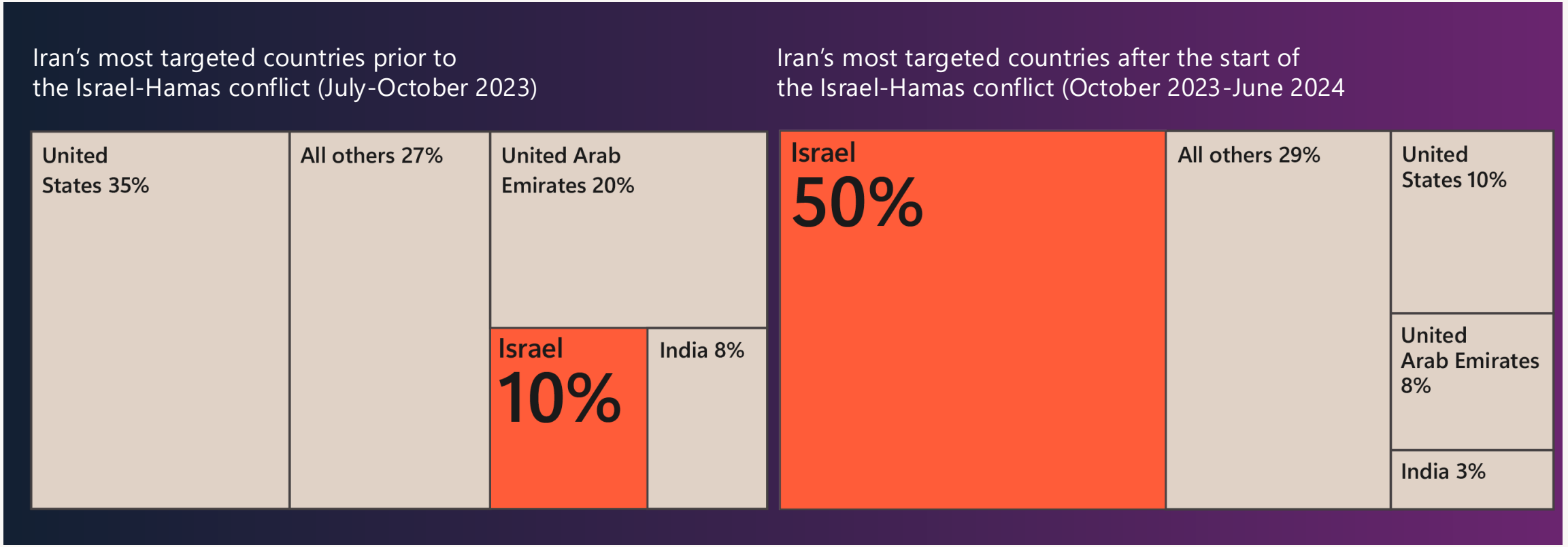
On the right are key elections the influence actors were likely seeking to influence.
The flags represent the nation-state affiliation of observed influence actors.

Source: Microsoft Threat Analysis Center

The many faces of hybrid war

Following the outbreak of the Israel-Hamas war, Iran surged its cyber, influence, and cyber-enabled influence operations against Israel.

From October 7, 2023, to July 2024, nearly half of the Iranian operations Microsoft observed targeted Israeli companies.



Source: Microsoft Threat Intelligence

Deterring the most advanced threats

A more robust deterrent framework will help to promote stability, protect critical infrastructure, and avoid some of the most harmful cyberattacks.

1. Strengthen international norms and diplomacy.

2. Sharpen government attributions of malicious activity.

3. Impose deterrent consequences.

Election interference

The goal of some nation-state-backed threat actor groups is to influence and undermine the results of democratic elections.

A collective commitment from industry, media, and governments is needed to defend against these threats.

Elections create impersonation opportunity

Microsoft Digital Crimes Unit monitors for domains related to elections around the world, to detect impersonations.

Target domain	Homoglyph domain	Payload delivered
crd.org	crd.com	Phish
crd.org	crd.com	Malware
gop.com	qop.com	Phish
gop.com	gops.com	Phish
gop.com	go.com	Phish
rnc.org	rnc.com	Phish
rnc.org	rnc.com	Malware
dnc.org	dnc.com	Phish
dnc.org	dn.org	Phish
dccc.org	dccc.com	Phish
nrcc.org	nrcc.com	Phish
sjrsa.com	sjrs.com	Phish

Source: Microsoft Threat Intelligence

Ransomware trends and insights

2.75x

Increase year over year in human-operated ransom-linked encounters

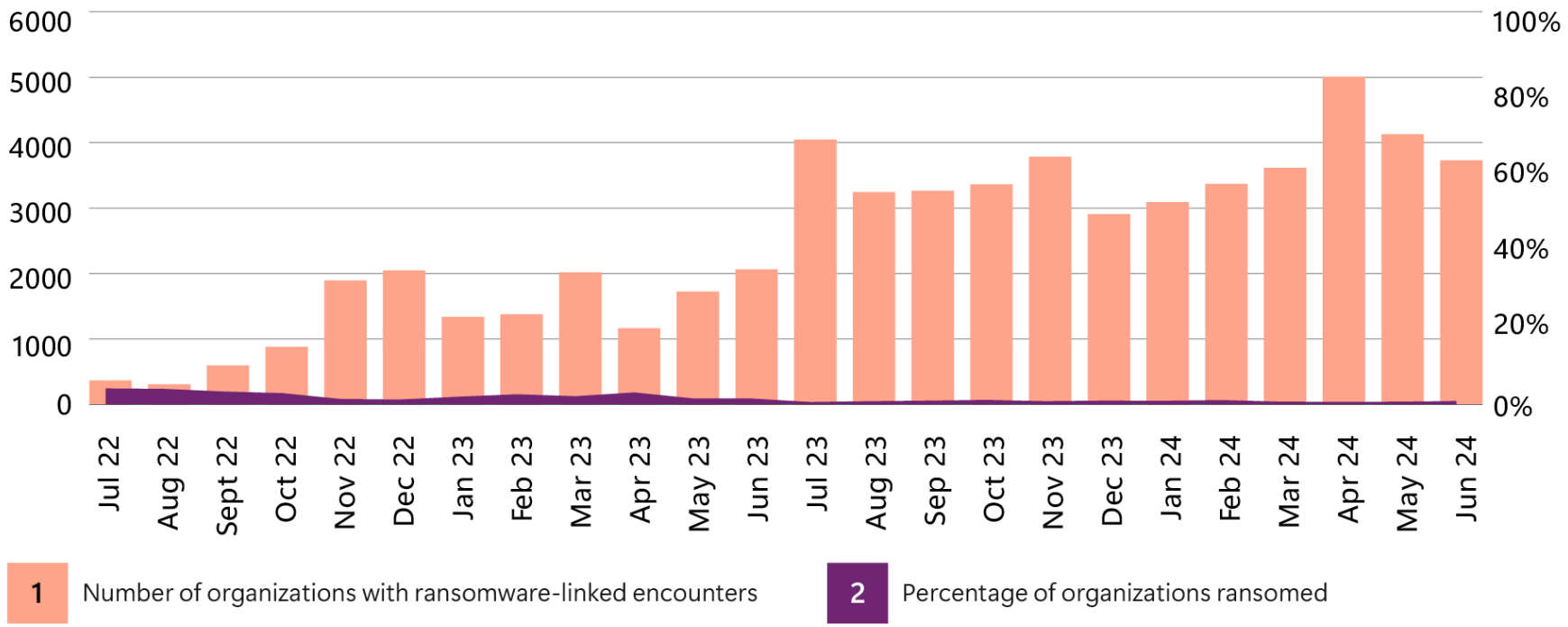
92%

Of successful ransom attacks leveraged an unmanaged device in the network

3x

Threefold decrease in ransom Attacks reaching encryption stage over the past two years

Organizations with ransom-linked encounters continues to increase while the percentage of those ransomed is decreasing (July 2022-June 2024)



Although organizations with ransom-linked encounters continues to increase, the percentage that are ultimately ransomed (reaching encryption stage) decreased more than threefold over the same time period.

Source: Microsoft Defender for Endpoint

Fraud landscape and trends

The ever-growing threat of cyber-enabled financial fraud

- Microsoft suspended nearly 64 million abusive service accounts in 2023.
- Working with industry and law enforcement partners to disrupt fraud actors in the real world.
- Working with law enforcement partners to improve intelligence exchange on cyber threats.

E-commerce fraud

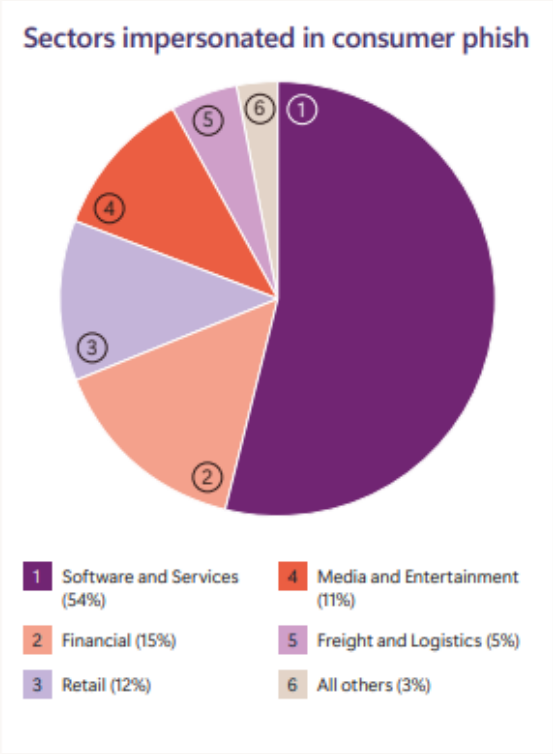
- The growth of the cybercrime-as-a-service economy gives criminals easier access to stolen data and fraudulent tools.
- Microsoft conducted over 1.6 billion risk evaluations for potential payment fraud and rejected $1.58 billion US in fraudulent transaction requests.

Fraud landscape and trends: Impersonation

Deepfakes

As deepfakes become more common in the business environment, organizations will have to implement countermeasures, such as requiring additional verification for transactions.

Corporate impersonation

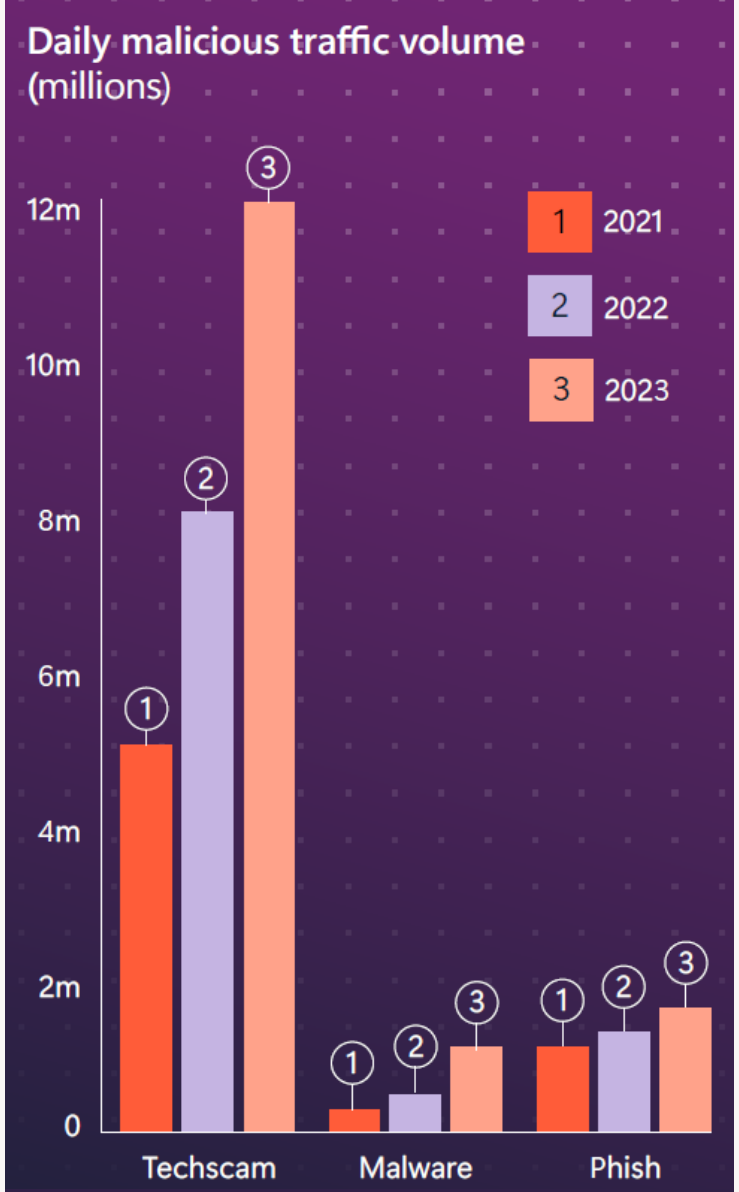


Account takeovers

Most ATOs still happen through simple methods like password spraying, phishing, keylogging, and using passwords from previous attacks found on the web.

Techscam

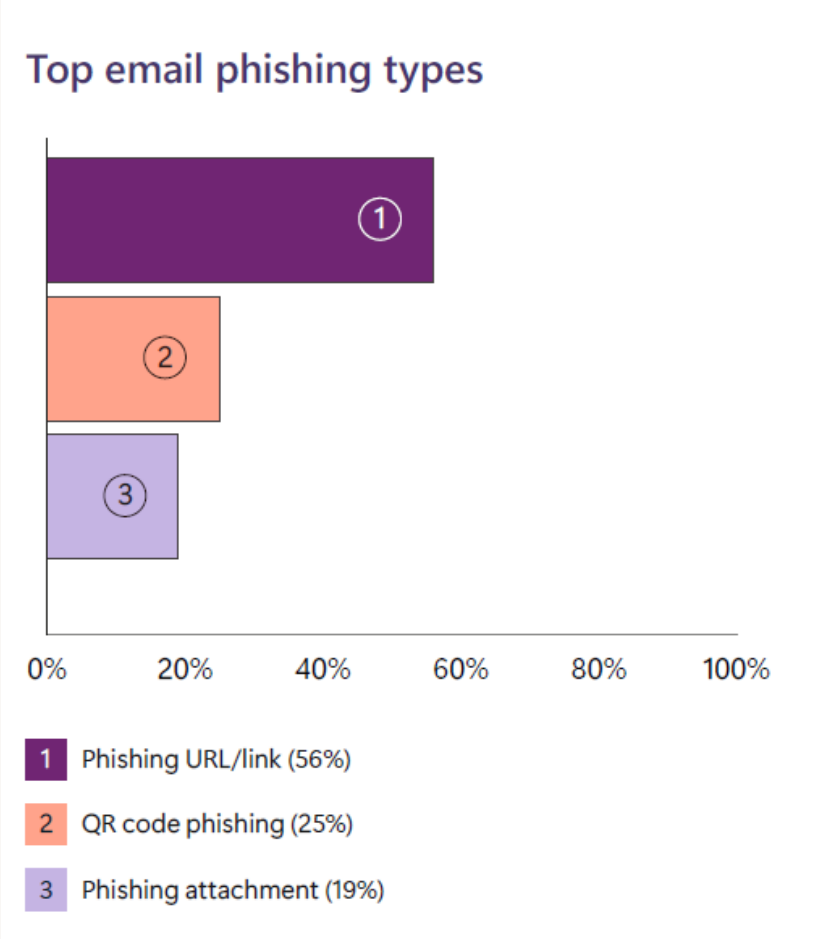
>70% of these malicious entities are active for less than two hours, meaning they may be gone before they're even detected.



The daily volume of techscam traffic escalated by 400% since 2022, in contrast to the 180% increase in malware and 30% in phishing.

Source: SmartScreen log data

Fraud landscape and trends: phishing

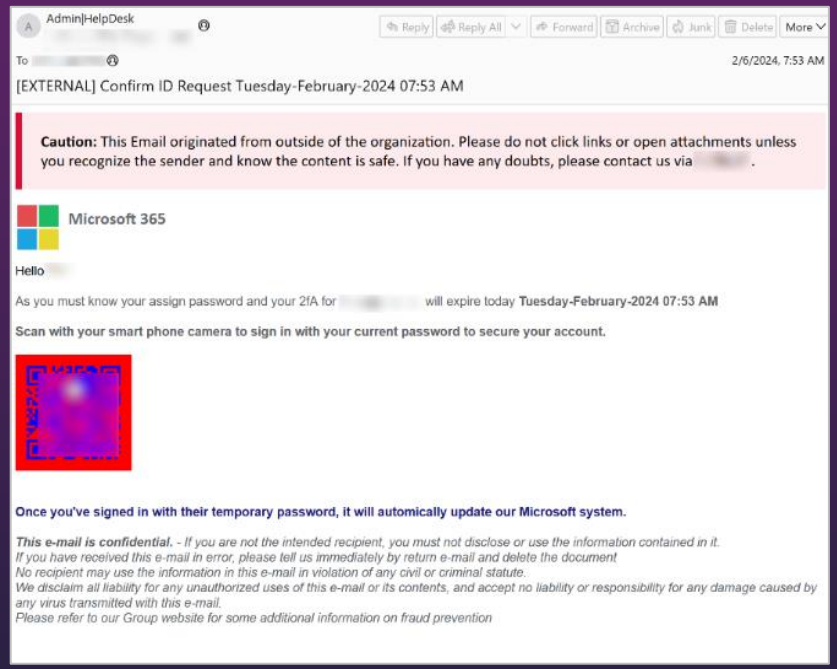


775 million

email messages contained malware

QR code phishing

By their nature, QR codes obscure the destination from the user, which creates a challenge for security



Identity attacks in perspective

Password-based attacks continue to dominate, but can be thwarted by using strong authentication methods.

Source: Microsoft Threat Intelligence

7,000
Password attacks per second

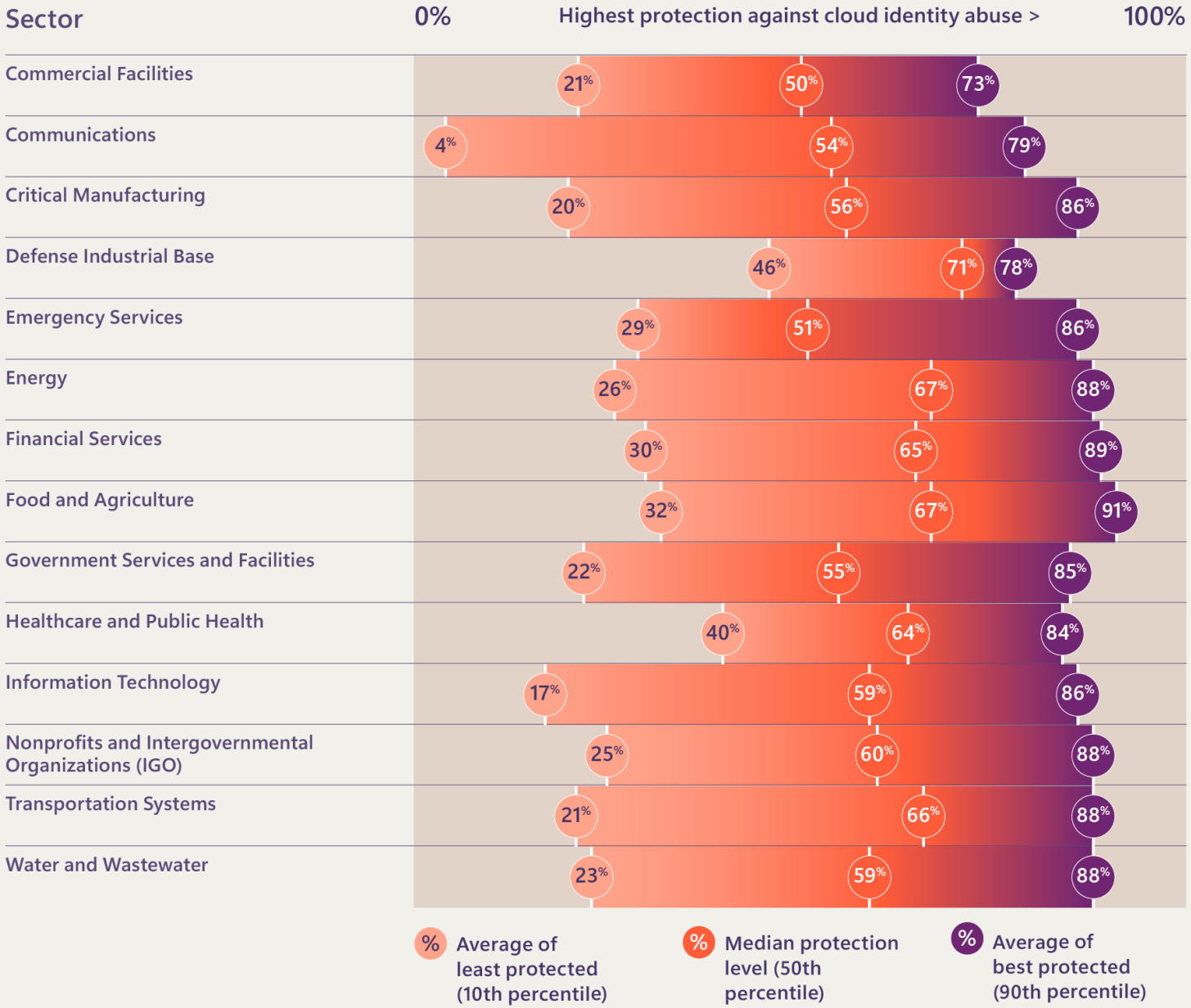
39,000
Token theft incidents per day

146%
Rise in AiTM phishing attacks

Stormy skies: the rise of cloud identity compromise

An attacker who manipulates identity can also manipulate any resource or process that identity is trusted to access, including email, other cloud services, or the on-premises environment.

Cloud identity abuse preparedness

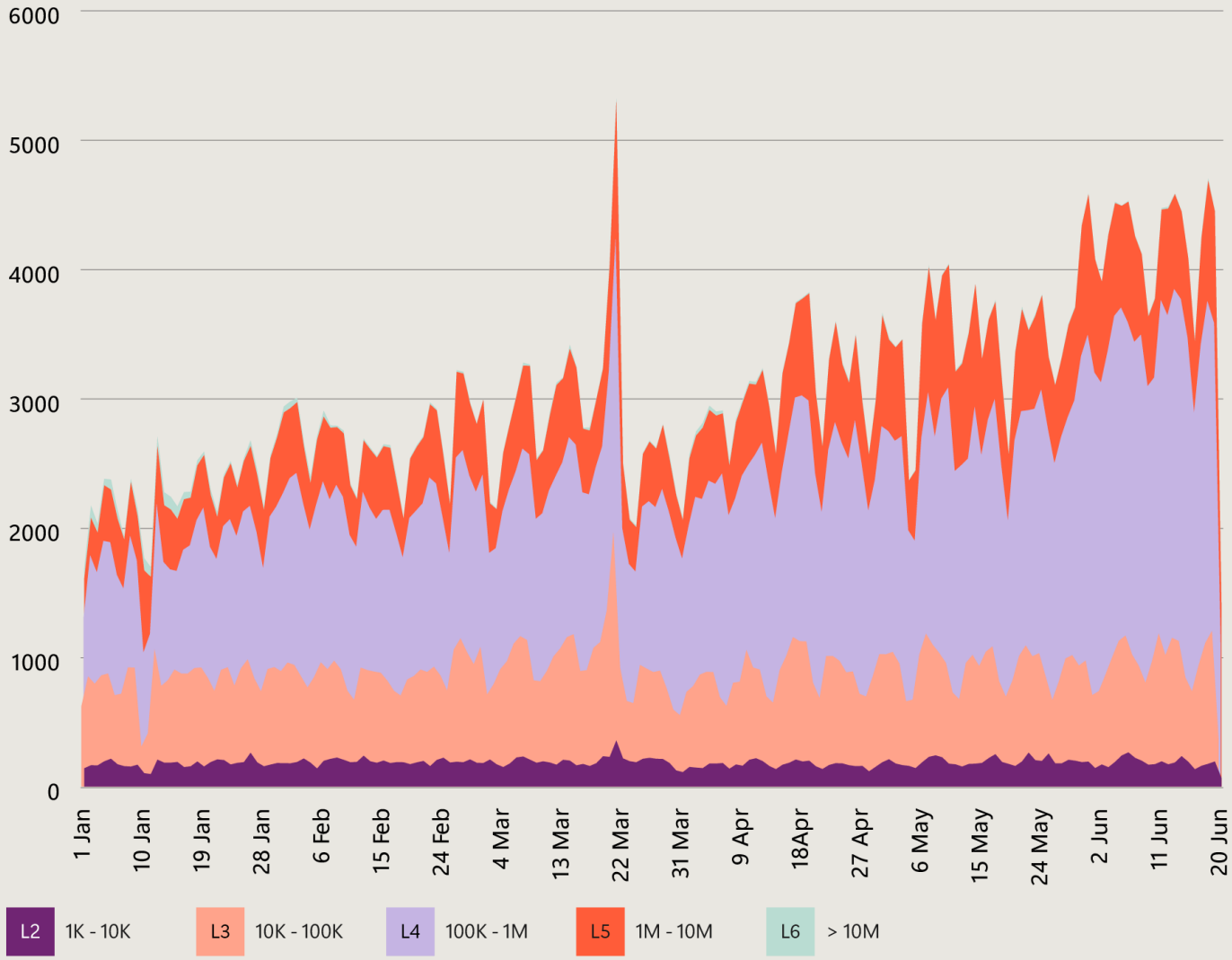


Sample size: 112,000 organizations representing a range of sizes and industries
Source: Microsoft Security Exposure Management

DDoS: Stealthier threats emerge

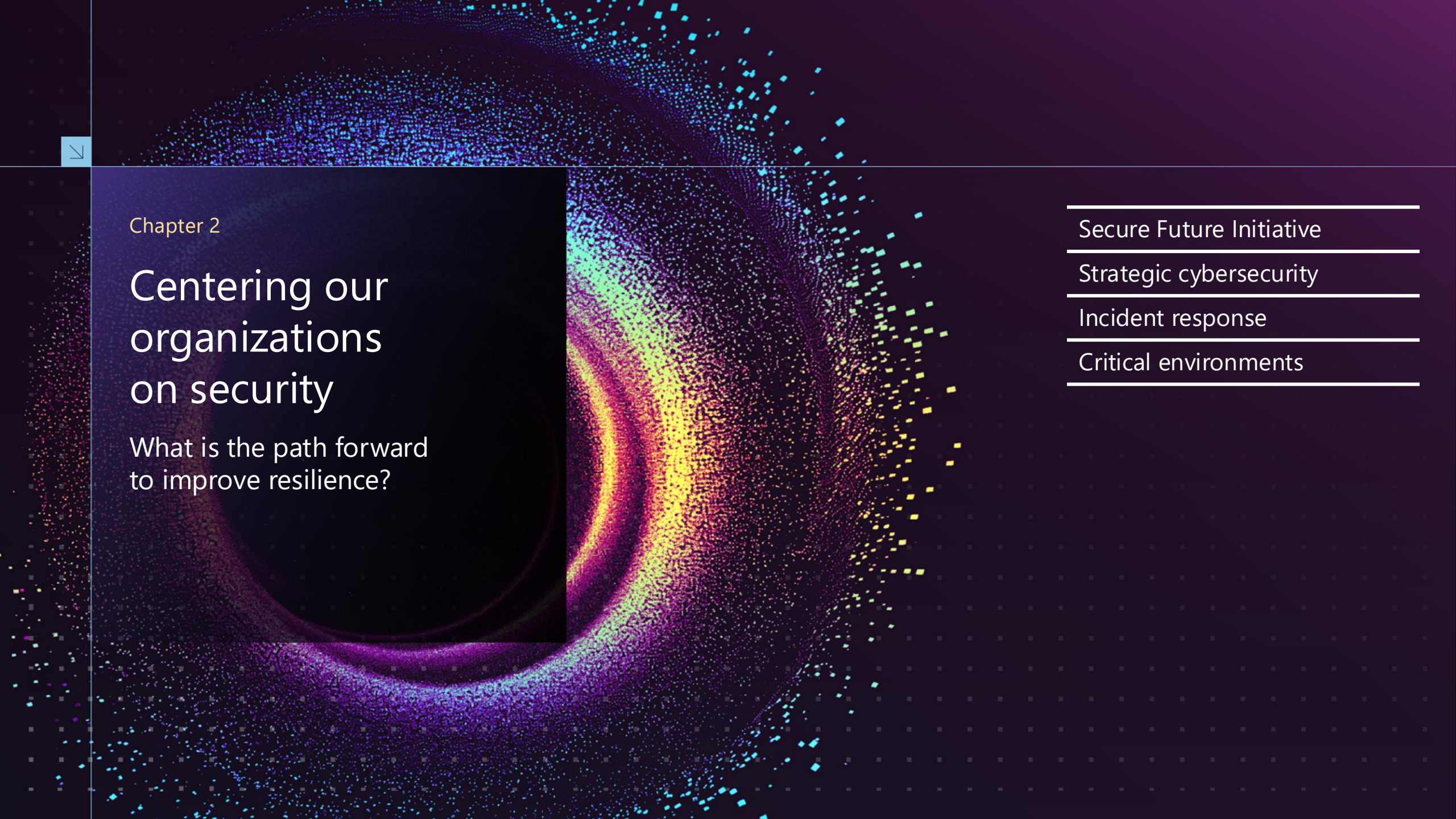
The increased focus of DDoS attacks on the application layer has created a greater risk of impact on business availability.

Number of network DDoS attacks (January-June 2024)



The number of DDoS attacks mitigated continues to increase, with a notable surge layer 4 (L4, application layer) attacks. Application layer attacks are more stealthy, sophisticated, and difficult to mitigate than network-level attacks. Layers in the key are in "packets per second (pps)".

Source: Microsoft Global DDoS Mitigation Operations



Chapter 2

Centering our organizations on security

What is the path forward to improve resilience?

Tackling technical debt and shadow IT for a secure future

Putting security above all else

The Microsoft Secure Future Initiative (SFI) is a multiyear initiative to evolve the way we design, build, test, and operate our products and services, to achieve the highest possible standards for security.

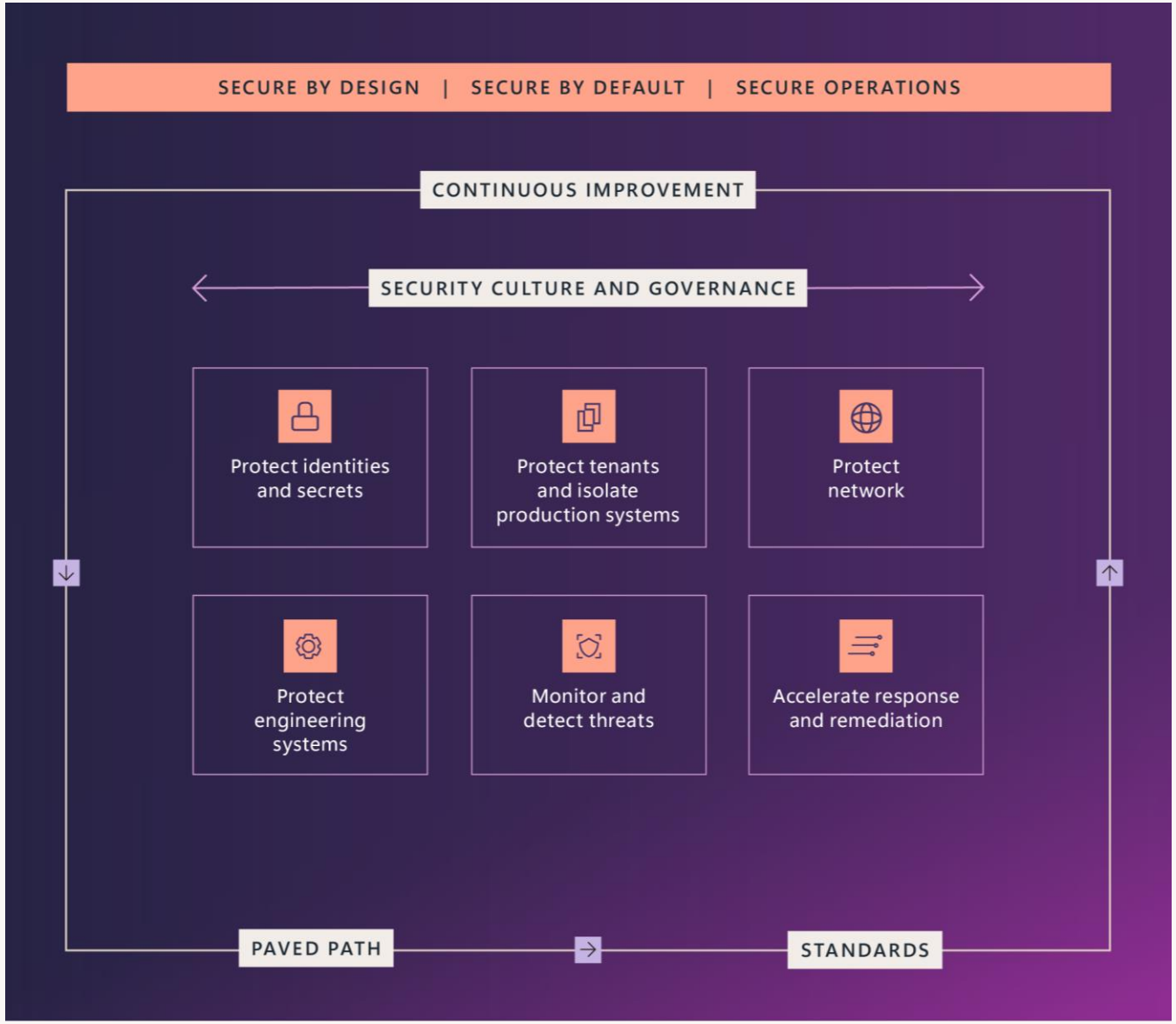
It's our long-term commitment to protect both the company and our customers in the ever-evolving threat landscape.

730k

SFI non-compliant apps eliminated

5.75 million

Inactive tenants eliminated, drastically reducing the potential cyberattack surface



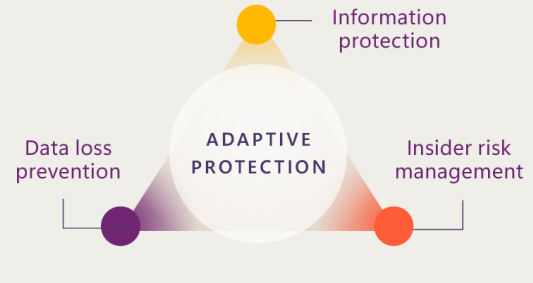
Strategic approaches to cybersecurity: data security

It is no longer enough to focus solely on the data; it's just as important to understand how that data moves within the organization, how users, customers or partners interact with it, and what level of risk is acceptable for the organization.

Data doesn't move on its own. It's moved by people.

An integrated approach to data security

- Classify and label sensitive data, and prevent its unauthorized use across apps, services, and devices.
- Understand the user intent and context around the use of sensitive data to identify the most critical risks
- Assign high-risk users to appropriate DLP, data lifecycle, and Conditional Access policies



Readiness levels: Protecting and governing data while benefitting from generative AI

1
Prepare data

Prepare your data for generative AI. Focus on labeling data, implementing controls, and educating users about data protection.

2
Limited implementation

Limited implementation of generative AI. Restrict access to sites that may contain sensitive files. Leverage tools that provide visibility into how users are using AI, which can inform stronger protection controls.

3
Used to enhance productivity

Generative AI used to enhance productivity. Optimize data governance and loss prevention. Use advanced capabilities for risk management and compliance.

4
Driving force for innovation

Generative AI used as a driving force for innovation. Expand adoption throughout the organization, continuously improve user behaviors and accountability, and extend data governance to cover all environments.

Hierarchy of cybersecurity needs

Drawing inspiration from Maslow's hierarchy of needs, this graphic illustrates a prioritization of cybersecurity, starting with the most basic need: protecting identities. AI has a role at each tier, underscoring its potential to enhance security measures.

Cultivating a robust security culture within the organization, helps ensure the technological defenses and human practices evolve in concert to mitigate threats effectively.

AUTOMATE SECURITY OPERATIONS
Automating security operations is the holistic approach to building on perspectives and insights across all layers in the pyramid.

IMPACT...
Automating processes at scale creates new opportunities for insights as well as relief for stressed defenders.

DETECT AND REMEDIATE THREATS
Monitoring your ecosystem to identify anomalous activity and contain threats.

IMPACT...
The ability to identify and respond quickly can limit lateral movement, contain damage to assets and deny persistence.

SECURE DIGITAL ASSETS
Digital assets, whether code, traditional data stores, and now generative AI models are all key components of modern workloads.

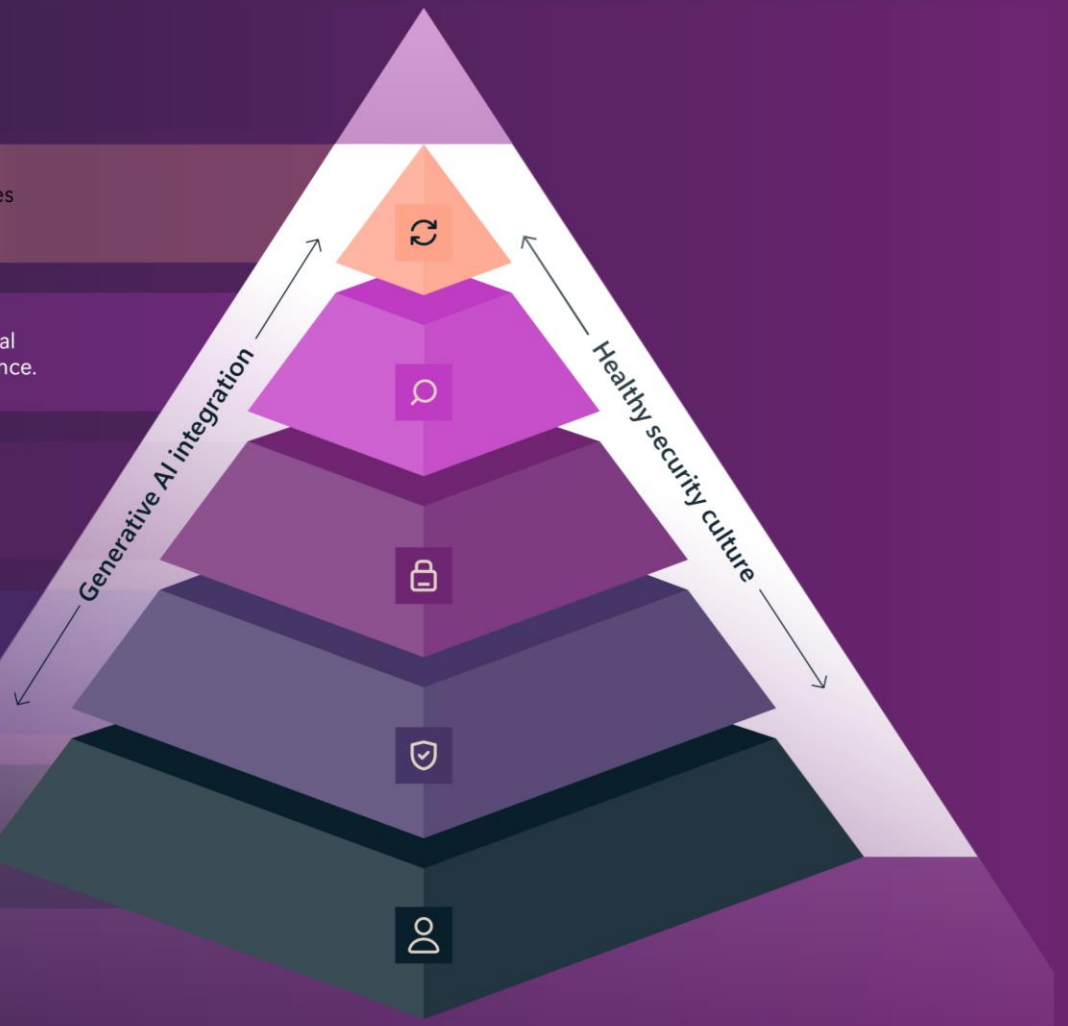
IMPACT...
Modern workloads deliver the value-add to end users who increasingly rely on their integrity and availability.

PROTECT ENDPOINTS
Protected endpoints include the multiple dimensions of devices in use today – from PCs and mobile devices, to network and operational technology (OT), and servers in datacenters.

IMPACT...
Effective endpoint protection can limit the repercussions of unauthorized access.

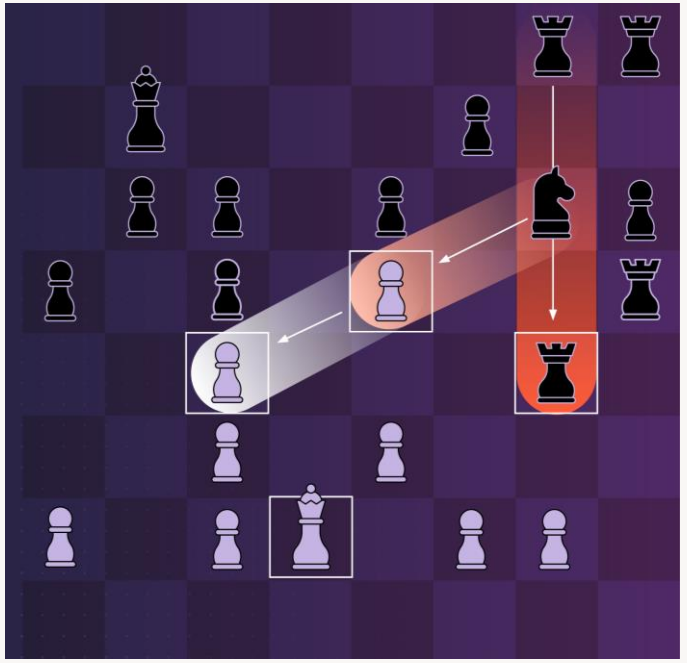
PROTECT IDENTITIES
"Attackers don't break in, they log in." Credentials for both individuals and machines are the perimeter of the modern attack surface.

IMPACT...
Strong identity security can greatly reduce risk exposure—particularly for ransomware attacks.



Threat-informed defense

Thinking differently to address threats



Siloed approach

- Rating bugs by severity and assigning teams to fix them within a set compliance window.

Strategic approach

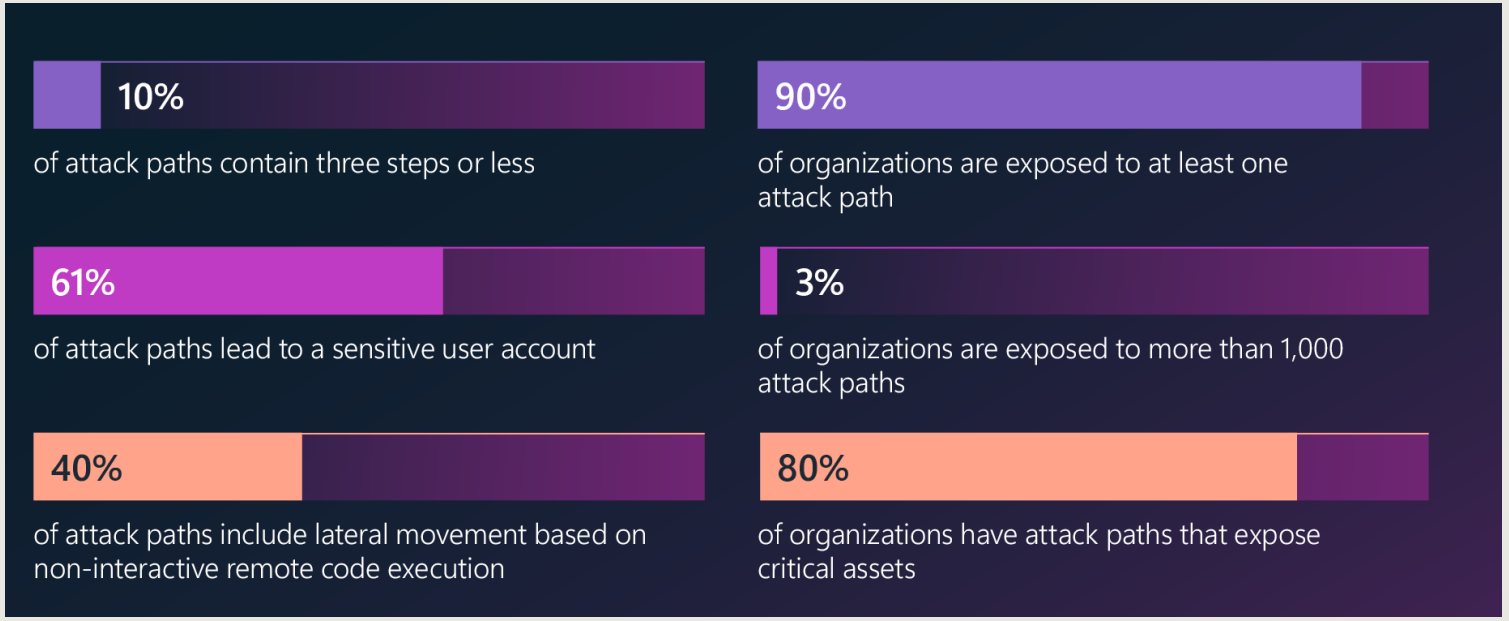
- Instead, we can use AI tools to see attack paths.
- Improve prioritization of mitigation efforts.

Attack path insights for threat-informed defense

Source: Microsoft Security Exposure Management

Scope: 112,000 organizations representing the full range of organization sizes.

Snapshot: June 2024



Security incident decisions: Dispatches from First Responders of the cybersecurity space

Real-world experience from Microsoft's Incident Response team can help to better prepare for cyber incidents.

> *Most common challenges we encountered during IR engagements:*
>
> - *Reporting lines not clearly defined*
> - *Roles and responsibilities not clearly defined*
> - *Lack of preparedness simulation exercises*

Preparation

Identify key decision makers, business-critical apps and services, roles and responsibilities, and response and recovery processes well in advance.

Communication

Tailor communication to each audience: company executives, regulatory bodies, employees, and the public.

Execution

Established playbooks with procedural plans to contain, recover, or remediate risks, and include actionable steps to address these tasks.

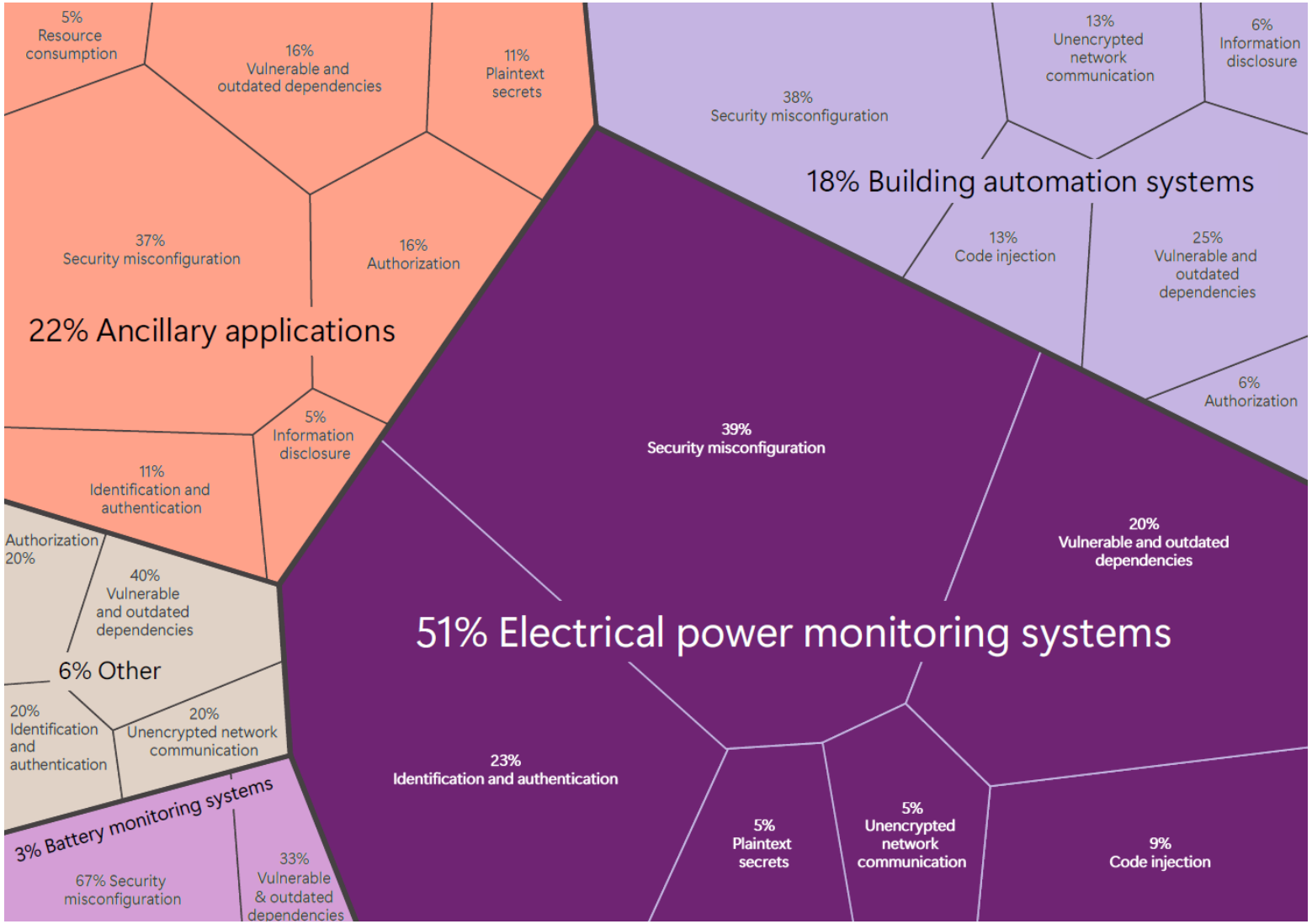
Critical environments: security stories from the OT frontline

Threat actors exploit OT devices to:

- Access critical networks
- Move laterally within the network
- Establish a foothold in a supply chain
- Disrupting OT operations

Microsoft third-party OT app review program:

- Identify and address potential vulnerabilities
- Help ensure robustness and reliability
- Enhance security standards across the industry



Source: Microsoft third-party OT application security assessments

Collective action: The digital transformation of defense and a call for partnership

Hybrid warfare, cyberattacks, and foreign influence operations pose grave risks not just to IT systems, but to the stability, prosperity and national security of society itself.

Cloud computing, AI, and quantum computing all have a role in cybersecurity, but their impacts on our collective defense can only be maximized through joint action and collaboration in defense innovation.

> *A deeper partnership between industry and governments is needed to implement the digital transformation of the defense sector.*

How Microsoft helps support democratic elections

During this unprecedented period of critical elections taking place around the world, Microsoft has worked to defend democratic institutions by combatting malicious schemes designed to disrupt or influence electoral processes and promoting a healthy information ecosystem.

Detection. Advanced tools and capabilities to monitor, analyze, and attribute malicious activities or campaigns that aim to disrupt, influence, or manipulate elections and elections infrastructure.

Response. The Digital Crimes Unit uses legal and technical expertise to disrupt the malicious activities intended to compromise, sabotage, or interfere with the elections. Microsoft Incident Response help respond to and recover from active cyber incidents.

Collaboration. Work with public and private stakeholders globally who share a similar goal of protecting the electoral process.

Chapter 3

Early insights: Ai's impact on cybersecurity

What do we know about new AI challenges and solutions today?

Two key insights about securing generative AI

1. Building is easy; testing is hard

Testing and tuning the system:

- Uncommon inputs
- Adversarial inputs
- Inputs from users who think differently from the developers

2. Generative AI security is nondeterministic

- Saying the same thing twice won't have the same effect
- Slight changes in phrasing may change the outcome
- This means "patching" is different than with traditional security vulnerabilities

Emerging techniques in AI-enabled attacks

Threat actors are using AI-enabled targeting aided by machine learning to target high-value individuals—those with access to trade secrets, financial systems, key strategies, and other sensitive and proprietary intellectual property.

AI-enabled spear phishing and whaling

- AI coupled with malware; a tool that lies dormant until it identifies its intended target and deploys
- Threat actors focus on highly specific targets and can exfiltrate only the most useful information
- The AI uses device cameras, speakers, and GPS for target verification

"Résumé swarming" and steganography

- Use AI to scrape keywords from job postings and develop "perfect" résumés
- Generate thousands of variations of highly qualified—but imaginary—candidates' résumés to apply for open positions
- Use steganography to embed invisible information to increase chances of passing automated screening tools
- Applicant selected for interviews and ultimately hired. In this way, threat actors emplace insiders within an organization to steal trade secrets, intelligence, or other sensitive information.

> *When it comes to AI-enabled human targeting, threats will be more difficult to detect and defend against—even with AI tools assisting defensive strategies.*

Nation-state threat actors using AI for influence operations

Adversarial use of AI in influence operations

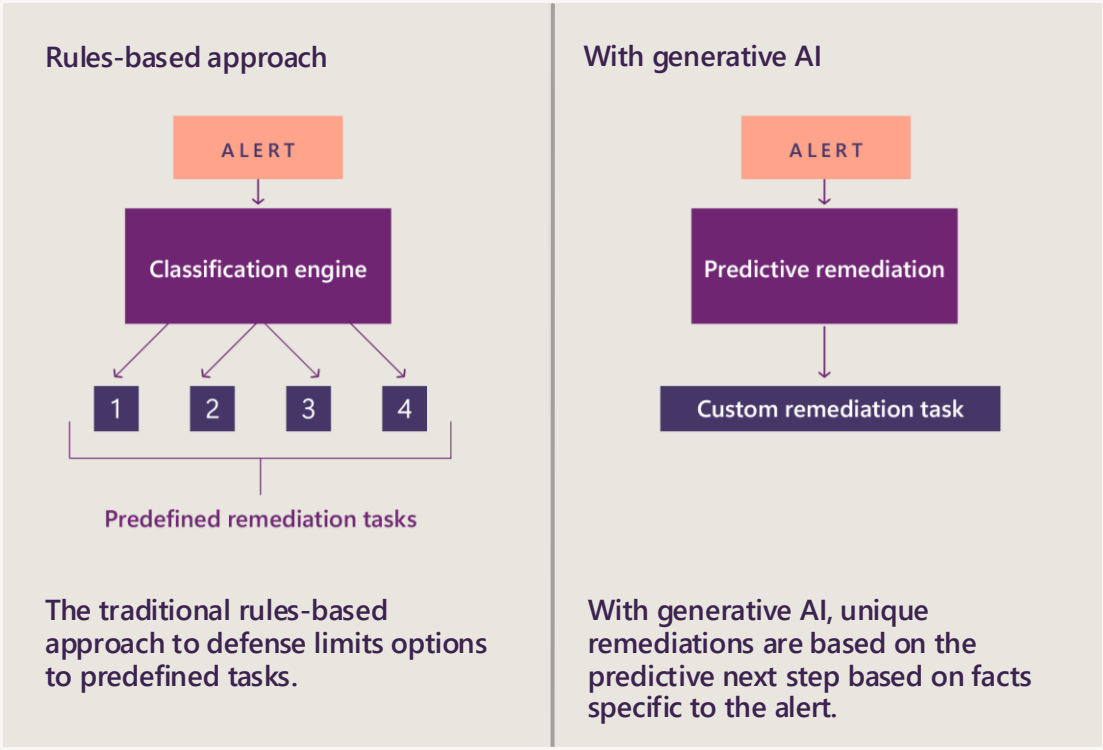
Capability	China	Russia	Iran & proxies
Text	MEDIUM / LOW	MEDIUM / LOW	LOW
Image	HIGH	HIGH	MEDIUM / LOW
Audio/video	HIGH	HIGH	LOW
Example	May 2024: Bespoke Taizi Flood AI-generated cartoon 	June 2024: AI-generated audio of Elon Musk narrating fabricated documentary 	April 2024: Likely AI-generated video leading up to Iranian military operation 

Nation-state threat actor groups, such as those backed by Russia, Iran, and China, are increasingly incorporating AI-generated or enhanced content into their influence operations in search of greater productivity, efficiency, and audience engagement.

AI for defense

Microsoft's significant investment in AI innovation is aimed at providing cybersecurity defenders with an asymmetric advantage over attackers.

Using generative AI to understand cyberattacks and create tailored mitigations



Harnessing AI to detect and disrupt cyberattacks

Detecting hidden attacks with AI

Hands-on-keyboard attacks, where cybercriminals directly interact with compromised systems, are hard to detect.

- LLMs fine-tuned to identify suspicious activities
- Can learn from the context and semantics and flag potential threats

Disrupting attacks by combining endpoint detection and response with AI

AI model alerts when it detects hands-on-keyboard attack.

MDE automatically:

- isolates affected device
- temporarily disables compromised user accounts

Extending AI across cybersecurity

AI models can analyze and find malicious activities using large and complex data sources such as network logs, email communications, web traffic, and social media.

Seven areas of security operations efficiencies

AI can enhance threat detection, response, analysis, and prediction. It can process large volumes of unstructured data to gain insights, answer questions and make informed decisions. Microsoft is leveraging AI in seven key areas in order to support our security operations:

1 **Triaging.** Teams in a security organization receive large volumes of requests and tickets. Depending on the complexity of the logic that determines how these items are dispositioned, AI can help triage some of the items and increase the efficiency and effectiveness of responding teams. Saving at least 20 hours per person, per week.

2 **Prioritizing work items.** Assess the priority of a given item based on how similar items were prioritized in the past.

AI can ensure that the prioritization criteria are up to date with the ever-evolving compliance requirements.

3 **Knowledge gathering from diverse external sources.** Augmenting proprietary in-house datasets with online content. AI can scrape online content and extract security-related information at scale.

One of our internal teams identifies and processes 50 articles per week.

Before: 2 hours per article

With AI: Reports within minutes

4 **Knowledge retrieval.** Integrate information about security policies, best practices, and the remediation actions necessary for compliance. LLMs can generate tailored answers to questions and follow-ups.

5 **Risk assessment.** Assimilate information from diverse sources, to conduct informed risk assessment.

Leverage unstructured organizational knowledge and historical precedents to enrich the set of factors determining risk.

6 **Learning from the past.** Security operations generate large volumes of diverse artifacts (tickets, reports, playbooks). LLMs can ingest data about previous incidents, violations, and remediations to uncover valuable learnings.

7 **Reporting.** Distill artifacts such as documents and slides into reports tailored to the audience and report goal.

Governments and industries advancing global AI security

- There is a consensus on the importance of safety and security in the development, deployment, and use of AI.
- Governments have pursued different approaches in implementing security requirements.

Policy approaches vary in scope and scale

Differences among governments' policy initiatives reflect:

- Core values of the governments' leadership
- Countries' legal and constitutional frameworks
- The state of the technology industry and its potential for future growth.

International standards

International standards can help mitigate the fragmentation of AI security regulation.

- There are two ISO standards (42001 and 27090) that relevant to AI security
- The US government's National Institute of Standards and Technology has a risk management framework and crosswalk that address AI and security intersections

Collaborative policy initiatives for AI security

Organizations around the world are collaborating to advance government policy initiatives on enhanced AI security.

July 2023

- Microsoft, Anthropic, Google and OpenAI launched Frontier Model Forum, an industry body focused on ensuring safe and responsible development of frontier AI models.[63]

August 2023

- The White House announces the AI Cyber Challenge, for cybersecurity researchers to spur the use of AI to identify and fix software vulnerabilities.[64] Microsoft committed to host competition on Microsoft Azure.

November 2023

- The UK launched the world's first safety institute to spur collaboration on AI's safety with leading AI companies and nations.[65]
- The US Department of Commerce, through National Institute of Standards and Technology (NIST) announced the US Artificial Intelligence Safety Institute (USAISI) to lead the US Government's efforts on AI safety and trust, including working with partners in academia, industry, government, and civil society to advance AI safety.[66]
- The Bletchley Agreement for collaboration resulted from an AI Safety Summit convened by the UK and including the US, EU, and China, likeminded AI companies, and 28 country delegations.[67]
- Microsoft contributed to the development of secure AI system guidelines alongside the UK National Cyber Security Centre (NCSC), and the US Cybersecurity and Infrastructure Security Agency (CISA),[68] among others. It was co-sealed by 23 domestic and international cybersecurity organizations. This publication marked a significant step in addressing the intersection of AI, cybersecurity, and critical infrastructure.

January 2024

- CISA's cross-sector analysis of sector-specific AI risk assessments completed by sector risk management agencies. Microsoft provided recommendations through the IT Sector Coordinating Council - a public private partnership for collaboration between IT sector and the Department of Homeland Security (DHS).

February 2024

- The Japanese government launched a new AI Safety Institute within the Information-technology Promotion Agency (IPA) in collaboration with relevant ministries and agencies.[69] The Institute aims to examine evaluation methods and standards related to AI. Japan plans to collaborate with the UK and the US.

March 2024

- The US Department of Treasury released a report on the current state of AI-related cybersecurity and fraud risks in financial services, including an overview of current AI use cases, trends of threats and risks, best-practice recommendations, and challenges and opportunities.[70]

April 2024

- In April 2024, building on the NCSC secure AI development guidelines release in 2023, the US National Security Agency's Artificial Intelligence Security Center published the joint Cybersecurity Information Sheet Deploying AI Systems Securely[71] in collaboration with CISA, the US Federal Bureau of Investigation, the Australian Signals Directorate's Australian Cyber Security Centre, the Canadian Centre for Cyber Security, the New Zealand National Cyber Security Centre, and the United Kingdom's National Cyber Security Centre.
- The US Department of Homeland Security (DHS) released Safety and Security Guidelines for Critical Infrastructure Owners and Operators.[72] Microsoft contributed to the cross-sector risk assessments that informed the DHS guidance.
- Microsoft joined the DHS AI Safety and Security Board (AISSB).[73] The AISSB advises the DHS Secretary, the critical infrastructure community, other private sector stakeholders, and the broader public on the safe, secure, and responsible development and deployment of AI technology in our nation's critical infrastructure.

May 2024

- The second global AI summit secured safety commitments from companies. It is a new agreement[74] between 10 countries and the EU to establish an international network similar to the UK's AI Safety Institute,[75] the world's first publicly backed organization to accelerate the advancement of AI safety science. The network will promote a common understanding of AI safety and align its work with research, standards, and testing. Australia, Canada, the EU, France, Germany, Italy, Japan, Singapore, South Korea, the UK, and the US have signed the agreement.[76]
- Microsoft released a blueprint for mutual prosperity through AI governance in Korea.[77]

June 2024

- Microsoft funded the Securing Critical Infrastructure in the Age of AI workshop led by Georgetown University's Center for Security and Emerging Tech (CSET). CSET will publish a report based on findings from the workshop offering policy recommendations for AI security in critical infrastructure. Expected publication date: September 2024.
- Microsoft hosted and participated in the first federal AI security tabletop exercise led by CISA JCDC.AI,[78] convening more than 50 AI experts from US and international agencies and industry partners focused on effective and coordinated responses to AI security incidents.

Mitigating the most advanced risks in the age of AI

Russia

Forest Blizzard: Research into satellite communications protocols

North Korea

Emerald Sleet: Identify experts focused on Asia-Pacific defense issues

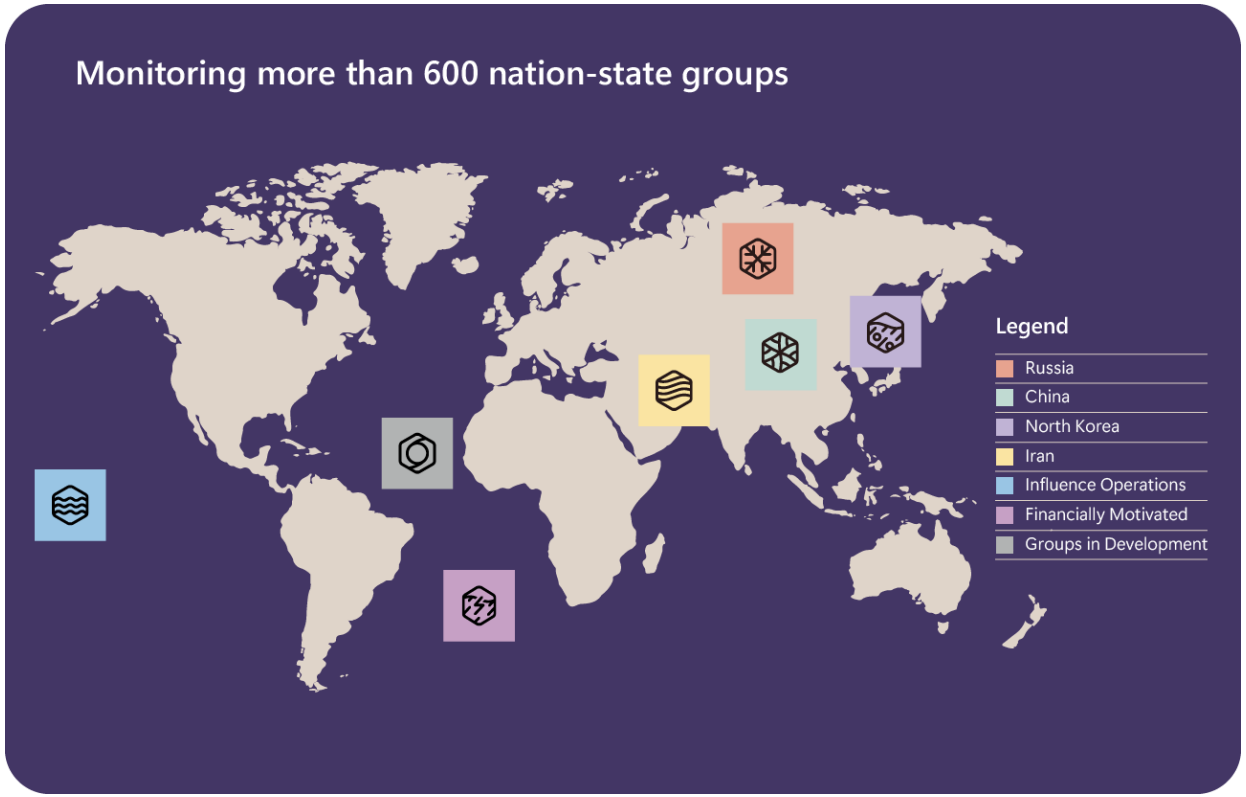
Iran

Crimson Sandstrom: Research ways malware can evade detection

China

Charcoal Typhoon: Create content likely for use in phishing

Salmon Typhoon: Translate technical papers and assist with coding



Microsoft policy to stay ahead of threat actors:

- **Identification and action:** We will disrupt their activities if we detect use of our AI APIs, services or systems by threat actors
- **Notification to other providers:** We will notify other providers if we detect threat actor use of their AI, AI APIs, services, and/or systems
- **Collaboration with other stakeholders:** We collaborate with others to regularly exchange information about detected threat actors' use of AI.
- **Transparency:** We will inform the public and stakeholders about actions taken under these principles, including details on the use of AI within our systems and measures taken against them.

Source: Staying ahead of threat actors in the age of AI (in collaboration with OpenAI). Microsoft Security Blog | February 2024

Microsoft

Microsoft Digital Defense Report 2024

We'd love to hear your feedback!

Scan the QR Code or press the links in the announcements