Microsoft

Navigating cyberthreats and strengthening defenses in the era of AI

Microsoft Threat Intelligence

Cyber Signals

February 2024

Every day more than

2.5 billion

cloud-based, AI-driven detections protect Microsoft customers.

Introduction

Today the world of cybersecurity is undergoing a massive transformation. Artificial intelligence (AI) is at the forefront of this change, posing both a threat and an opportunity. While AI has the potential to empower organizations to defeat cyberattacks at machine speed and drive innovation and efficiency in threat detection, hunting, and incident response, adversaries can use AI as part of their exploits. It's never been more critical for us to design, deploy, and use AI securely.

At Microsoft, we are exploring the potential of AI to enhance our security measures, unlock new and advanced protections, and build better software. With AI, we have the power to adapt alongside evolving threats, detect anomalies instantly, respond swiftly to neutralize risks, and tailor defenses for an organization's needs.

AI can also help us overcome another of the industry's biggest challenges. In the face of a global cybersecurity workforce shortage, with roughly 4 million cybersecurity professionals needed worldwide, AI has the potential to be a pivotal tool to close the talent gap and to help defenders be more productive.

We've already seen in one study how Copilot for Security can help security analysts regardless of their expertise level—across all tasks, participants were 44 percent more accurate and 26 percent faster.

As we look to secure the future, we must ensure that we balance preparing securely for AI and leveraging its benefits, because AI has the power to elevate human potential and solve some of our most serious challenges.

A more secure future with AI will require fundamental advances in software engineering. It will require us to understand and counter AI-driven threats as essential components of any security strategy. And we must work together to build deep collaboration and partnerships across public and private sectors to combat bad actors.

As part of this effort and our own Secure Future Initiative, OpenAI and Microsoft are today publishing new intelligence detailing threat actors' attempts to test and explore the usefulness of large language models (LLMs) in attack techniques.

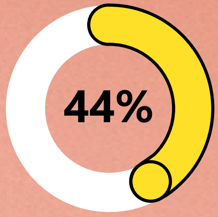
We hope this information will be useful across industries as we all work towards a more secure future. Because in the end, we are all defenders.

Bret Arsenault,
CVP, Chief Cybersecurity Advisor

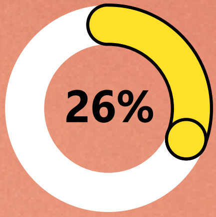
Security Snapshot

Traditional tools no longer keep pace with the threats posed by cybercriminals. The increasing speed, scale, and sophistication of recent cyberattacks demand a new approach to security. Additionally, given the cybersecurity workforce shortage, and with cyberthreats increasing in frequency and severity, bridging this skills gap is an urgent need.

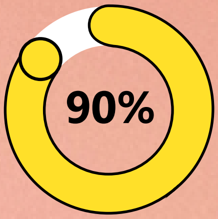
AI can tip the scales for defenders. One recent study of Microsoft Copilot for Security (currently in customer preview testing) showed increased security analyst speed and accuracy, regardless of their expertise level, across common tasks like identifying scripts used by attackers, creating incident reports, and identifying appropriate remediation steps.[1]



44 percent more accurate across all tasks for Copilot for Security users[1]



26 percent faster across all tasks for Copilot for Security users[1]



90 percent said they want Copilot next time they do the same task[1]

We are all defenders

Threat briefing

Attackers are exploring AI technologies

The cyberthreat landscape has become increasingly challenging with attackers growing more motivated, more sophisticated, and better resourced. Threat actors and defenders alike are looking to AI, including LLMs, to enhance their productivity and take advantage of accessible platforms that could suit their objectives and attack techniques.

Given the rapidly evolving threat landscape, today we are announcing Microsoft's principles guiding our actions that mitigate the risk of threat actors, including advanced persistent threats (APTs), advanced persistent manipulators (APMs) and cybercriminal syndicates, using AI platforms and APIs. These principles include identification and action against malicious threat actor's use of AI, notification to other AI service providers, collaboration with other stakeholders, and transparency.

Although threat actors' motives and sophistication vary, they share common tasks when deploying attacks. These include reconnaissance, such as researching potential victims' industries, locations, and relationships; coding, including improving software scripts and malware development; and assistance with learning and using both human and machine languages.

Forest Blizzard

Emerald Sleet

Crimson Sandstorm

Charcoal Typhoon

Salmon Typhoon

Nation-states attempt to leverage AI
In collaboration with OpenAI, we are sharing threat intelligence showing detected state-affiliated adversaries—tracked as Forest Blizzard, Emerald Sleet, Crimson Sandstorm, Charcoal Typhoon, and Salmon Typhoon—using LLMs to augment cyberoperations.

The objective of Microsoft's research partnership with OpenAI is to ensure the safe and responsible use of AI technologies like ChatGPT, upholding the highest standards of ethical application to protect the community from potential misuse.

Forest Blizzard (STRONTIUM), a highly effective Russian military intelligence actor linked to The Main Directorate of the General Staff of the Armed Forces of the Russian or GRU Unit 26165, has targeted victims of tactical and strategic interest to the Russian government. Its activities span a variety of sectors including defense, transportation/logistics, government, energy, NGOs, and information technology.

Emerald Sleet (Velvet Chollima) is a North Korean threat actor Microsoft has found impersonating reputable academic institutions and NGOs to lure victims into replying with expert insights and commentary about foreign policies related to North Korea.

Emerald Sleet's use of LLMs involved research into think tanks and experts on North Korea, as well as content generation likely to be used in spear phishing campaigns. Emerald Sleet also interacted with LLMs to understand publicly known vulnerabilities, troubleshoot technical issues, and for assistance with using various web technologies.

Crimson Sandstorm (CURIUM) is an Iranian threat actor assessed to be connected to the Islamic Revolutionary Guard Corps. The use of LLMs has involved requests for support around social engineering, assistance in troubleshooting errors, .NET development, and ways in which an attacker might evade detection when on a compromised machine.

Charcoal Typhoon (CHROMIUM) is a China-affiliated threat actor predominantly focused on tracking groups in Taiwan, Thailand, Mongolia, Malaysia, France, Nepal, and individuals globally that oppose China's policies. In recent operations, Charcoal Typhoon has been observed engaging LLMs to gain insights into research to understand specific technologies, platforms, and vulnerabilities, indicative of preliminary information-gathering stages.

Another China-backed group, **Salmon Typhoon**, has been assessing the effectiveness of using LLMs throughout 2023 to source information on potentially sensitive topics, high-profile individuals, regional geopolitics, US influence, and internal affairs. This tentative engagement with LLMs could reflect both a broadening of its intelligence-gathering toolkit and an experimental phase in assessing the capabilities of emerging technologies.

Our research with OpenAI has not identified significant attacks employing the LLMs we monitor closely.

We have taken measures to disrupt assets and accounts associated with these threat actors and shape the guardrails and safety mechanisms around our models.

Other AI threats emerge

AI-powered fraud is another critical concern. Voice synthesis is an example of this, where a three-second voice sample can train a model to sound like anyone. Even something as innocuous as your voicemail greeting can be used to get a sufficient sampling.

Much of how we interact with each other and conduct business relies on identity proofing, such as recognizing a person's voice, face, email address, or writing style.

It's crucial that we understand how malicious actors use AI to undermine longstanding identity proofing systems so we can tackle complex fraud cases and other emerging social engineering threats that obscure identities.

AI can also be used to help companies disrupt fraud attempts. Although Microsoft discontinued our engagement with a company in Brazil, our AI systems detected its attempts to reconstitute itself to re-enter our ecosystem.

The group continually attempted to obfuscate its information, conceal ownership roots, and re-enter, but our AI detections used nearly a dozen risk signals to flag the fraudulent company and associate it with previously recognized suspect behavior, thereby blocking its attempts.

Microsoft is committed to responsible human-led AI featuring privacy and security with humans providing oversight, evaluating appeals, and interpreting policies and regulations.

Educate employees and the public about cyberrisks:

Use conditional access policies: These policies provide clear, self-deploying guidance to strengthen your security posture that will automatically protect tenants based on risk signals, licensing, and usage. Conditional access policies are customizable and will adapt to the changing cyberthreat landscape.

Train and retrain employees on social engineering tactics: Educate employees and the public to recognize and react to phishing emails, vishing (voicemail), smishing, (SMS/text) social engineering attacks, and apply security best practices for Microsoft Teams.

Rigorously protect data: Ensure data remains private and controlled from end to end.

Leverage Generative AI security tools: Tools like Microsoft Copilot for Security can expand capabilities and enhance the organization's security posture.

Enable multifactor authentication: Enable multifactor authentication for all users, especially for administrator functions, as it reduces the risk of account takeover by over 99 percent.

99% risk reduction

Defending against attacks

Microsoft stays ahead with comprehensive security combined with Generative AI

Microsoft detects a tremendous amount of malicious traffic—more than 65 trillion cybersecurity signals per day. AI is boosting our ability to analyze this information and ensure that the most valuable insights are surfaced to help stop threats. We also use this signal intelligence to power Generative AI for advanced threat protection, data security, and identity security to help defenders catch what others miss.

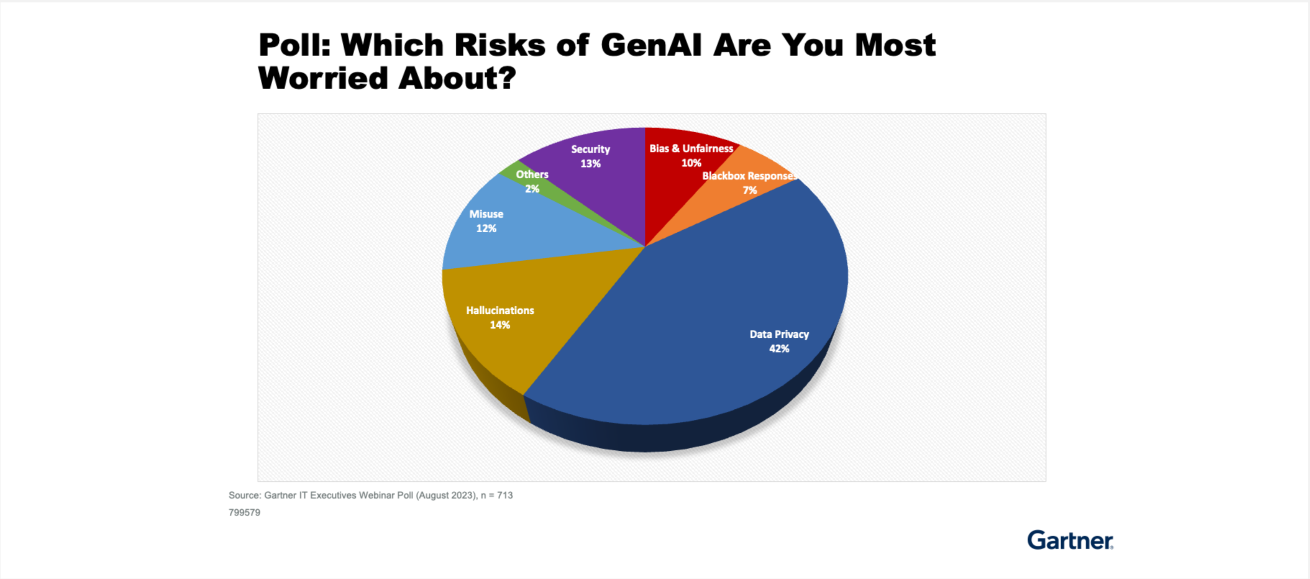
Microsoft uses several methods to protect itself and customers from cyberthreats, including AI-enabled threat detection to spot changes in how resources or traffic on the network are used; behavioral analytics to detect risky sign-ins and anomalous behavior; machine learning (ML) models to detect risky sign-ins and malware; Zero Trust models where every access request must be fully authenticated, authorized, and encrypted; and device health verification before a device can connect to a corporate network.

Because threat actors understand that Microsoft uses multifactor authentication (MFA) rigorously to protect itself—all our employees are set up for MFA or passwordless protection—we've seen attackers lean into social engineering in an attempt to compromise our employees.

Hot spots for this include areas where things of value are being conveyed, such as free trials or promotional pricing of services or products. In these areas, it isn't profitable for attackers to steal one subscription at a time, so they attempt to operationalize and scale those attacks without being detected.

Naturally, we build AI models to detect these attacks for Microsoft and our customers. We detect fake students and school accounts, fake companies or organizations that have altered their firmographic data or concealed their true identities to evade sanctions, circumvent controls, or hide past criminal transgressions like corruption convictions, theft attempts, etc.

The use of GitHub Copilot, Microsoft Copilot for Security, and other copilot chat features integrated into our internal engineering and operations infrastructure can help prevent incidents that could impact operations.



Poll: Which Risks of GenAI Are You Most Worried About?

Source: Gartner IT Executives Webinar Poll (August 2023), n = 713
799579

Defending against attacks

To address email threats, Microsoft is improving capabilities to glean signals besides an email's composition to understand if it is malicious. With AI in the hands of threat actors, there has been an influx of perfectly written emails that improve upon the obvious language and grammatical errors which often reveal phishing attempts, making phishing attempts harder to detect.

Continued employee education and public awareness campaigns are needed to help combat social engineering, which is the one lever that relies 100 percent on human error. History has taught us that effective public awareness campaigns work to change behavior.

Microsoft anticipates that AI will evolve social engineering tactics, creating more sophisticated attacks including deepfakes and voice cloning, particularly if attackers find AI technologies operating without responsible practices and built-in security controls.

Prevention is key to combating all cyberthreats, whether traditional or AI-enabled.

Recommendations:

Apply vendor AI controls and continually assess their fit:
For any AI introduced into your enterprise, look for respective vendors' built-in features to scope AI access to employees and teams using the technology to foster secure and compliant AI adoption. Bring cyberrisk stakeholders across an organization together to align on defined AI employee use cases and access controls. Risk leaders and CISOs should regularly determine whether use cases and policies are adequate, or if they must change as objectives and learnings evolve.

Protect against prompt injections:
Implement strict input validation and sanitization for user-provided prompts. Use context-aware filtering and output encoding to prevent prompt manipulation. Regularly update and fine-tune LLMs to improve its understanding of malicious inputs and edge cases. Monitor and log LLM interactions to detect and analyze potential prompt injection attempts.

Mandate transparency across the AI supply chain:
Through clear and open practices, assess all areas where AI can come in contact with your organization's data, including through third-party partners and suppliers. Use partner relationships and cross-functional cyberrisk teams to explore learnings and close any resulting gaps. Maintaining current Zero Trust and data governance programs is more important than ever in the AI era.

Stay focused on communications:
Cyberrisk leaders must recognize that employees are witnessing AI's impact and benefits in their personal lives and will naturally want to explore applying similar technologies across hybrid work environments. CISOs and other leaders managing cyberrisk can proactively share and amplify their organizations' policies on the use and risks of AI, including which designated AI tools are approved for enterprise and points of contact for access and information. Proactive communications help keep employees informed and empowered, while reducing their risk of bringing unmanaged AI into contact with enterprise IT assets.

Expert Profile

Homa Hayatyfar

Principal Data and Applied Science Manager, Detection Analytics Manager

Homa Hayatyfar has seen how pathways to a career in cybersecurity are often nonlinear. She arrived at her career in cybersecurity by way of a research background in biochemistry and molecular biology—along with a passion for solving complex puzzles—and she believes that may be what the industry needs more of.

"The diversity of technical and soft skill competencies is how I build a team and is one of the biggest strengths I see at Microsoft. As Microsoft builds more diverse teams, which include those with the same demographics as attackers themselves, we continue to expand our threat intelligence capabilities," she says.

She says no day is the same working in cybersecurity, especially the intersection of cybersecurity and data science. As part of the data science arm to Microsoft's security operations team, Homa's work is to take an immense amount of data and transform those insights into practical steps to tackle potential risks head on, refining threat-detection methods and using machine learning models to reinforce Microsoft's defenses.

"Analyzing data has been the main catalyst propelling my career in cybersecurity. I specialize in securing digital landscapes and extracting insights from complex datasets."

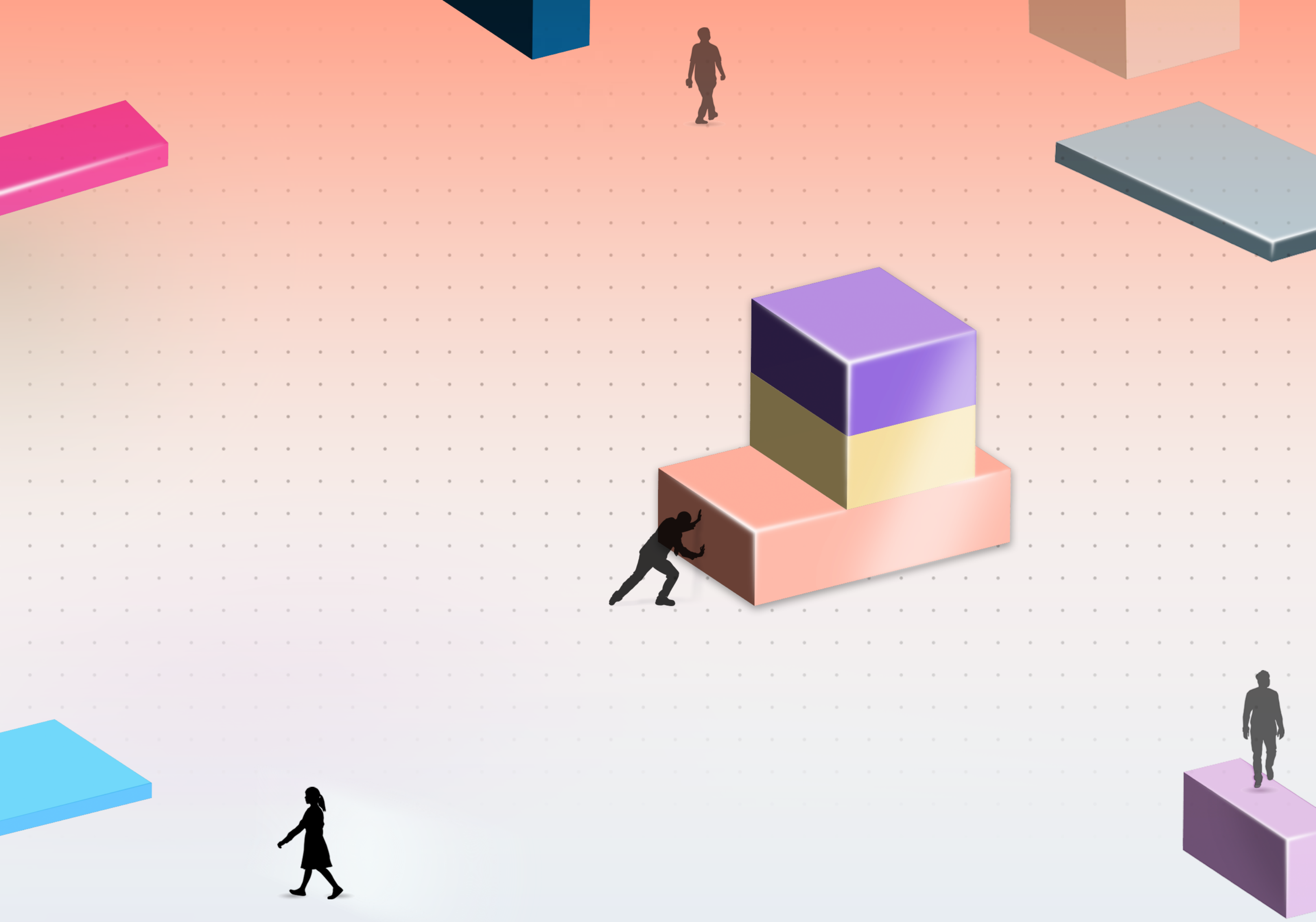
As adversaries have evolved, she says, so has Microsoft, consistently focusing on the perpetual evolution of adversaries to stay ahead. Being a frequently targeted company propels Microsoft to maintain a forward-looking stance to ensure resilience and readiness. "This reality propels us to maintain a forward-looking stance, ensuring our resilience and readiness," she says.

Homa says attackers will generally gravitate to what's easy and can be automated. The reason they persist with social engineering and traditional attacks like phishing, for example, is because it's effective—all it takes is one successful attempt to lure someone into sharing information, clicking a malicious link, or granting access to sensitive files. In the same way, attackers are increasingly looking to AI to help them do more.

"AI can help attackers bring more sophistication to their attacks, and they have resources to throw at it. We've seen this with the 300+ threat actors Microsoft tracks, and we use AI to protect, detect, and respond."

For companies looking to protect themselves, Homa stresses that the fundamentals matter: "Layering to add extra barriers such as applying Zero Trust principles, data protection, and multifactor authentication can protect against many of these attacks."

> **"This reality propels us to maintain a forward-looking stance, ensuring our resilience and readiness."**



Methodology: 1. Snapshot data represents a randomized controlled trial (RCT), where we tested 149 people to measure the productivity impact from using Microsoft Copilot for Security. In this RCT, we randomly gave Copilot to some analysts and not others, and then subtracted their performance and sentiments to get the effect of Copilot, separate from any base effect. Test subjects had basic IT skills, but were security novices, so we could test how Copilot helps "new in career" analysts. The Microsoft Copilot for Security RCT was conducted by Microsoft Office of the Chief Economist, November 2023. Additionally, Azure Active Directory provided anonymized data on threat activity, such as malicious email accounts, phishing emails, and attacker movement within networks. Additional insights are from the 65 trillion daily security signals gained across Microsoft, including the cloud, endpoints, the intelligent edge, our Compromise Security Recovery Practice and Detection and Response Teams, telemetry from Microsoft platforms and services including Microsoft Defender, and the 2023 Microsoft Digital Defense Report.