Microsoft Security

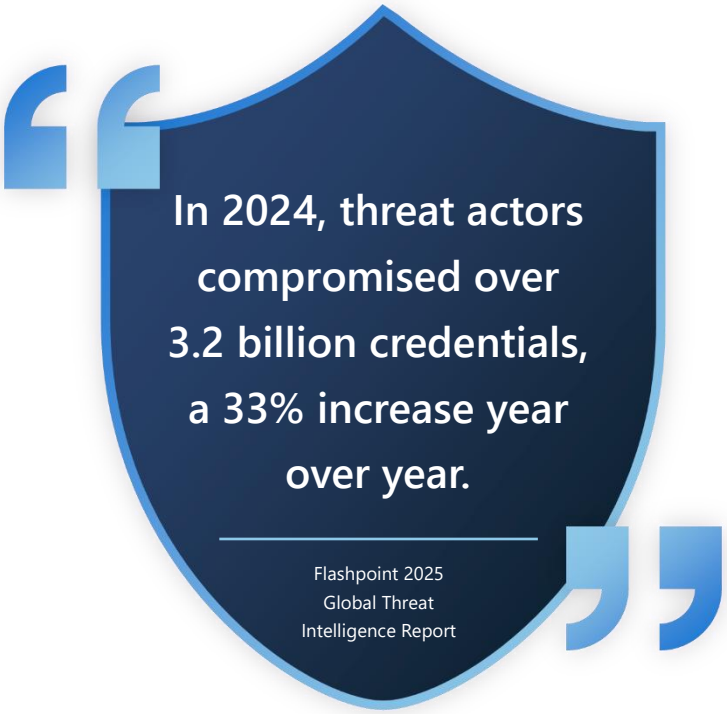
CUSTOMER SECURITY MANAGEMENT OFFICE

Security Foundations: Guarding Against AI-powered Attacks

Quick Reference Guide

This guide summarizes **key takeaways and best practices** from the internal Microsoft training course, *Security Foundations: Guarding Against AI-powered Attacks*. Use it for quick, just-in-time reminders to keep yourself, and your organization, secure.

Guide contents:



AI-powered attack methods

Threat actors are using AI to amplify their attacks, making them more **personalized, efficient, and effective**. Key attack methods include:

Targeted phishing
AI helps craft highly convincing phishing messages (email, chat, and text) tailored to you, often mimicking real communications to trick you into divulging information.

Automated malware creation
AI automates malware development, generating malicious software that's more complex and harder to detect by security systems.

Rapid vulnerability exploitation
AI-driven data analysis allows attackers to quickly find and exploit security weaknesses before defenses catch up.

Deepfakes
AI helps create images, videos, and audio that portrays events or people that simulate real people or events in false ways to earn your trust and get access to your data.

Spotting deepfakes

Over the last year, the use of AI-generated deepfakes by threat actors has increased tenfold. The tools to create them are inexpensive and readily available. To avoid becoming a victim of a cyberattack, look for these deepfake indicators and always verify the person's authenticity before engaging in a call, chat, or virtual meeting.

Listen for audio irregularities
Pay attention to unnatural speech patterns such as odd pauses, mismatched intonation, or background noise that doesn't quite fit. These can indicate an AI-generated voice.

Watch for visual inconsistencies
In video calls, look for strange image quality issues—for example, if a person's facial movements seem off or if a hand passing in front of their face causes distortion. These inconsistencies can be signs of a deep-fake video overlay.

Zero Trust security principles

The best offense is a good cybersecurity defense. Adopting a Zero Trust approach in your daily work makes life much harder for threat actors. Zero Trust is built on three principles:

1. **Assume breach**
   Operate as if a breach will happen. Don't wait for threats—proactively minimize potential damage. For example: Always store and back up data in secure cloud services (Teams, SharePoint, OneDrive for Business) instead of only locally. If your device is lost, fails, or gets malware, your data is likely recoverable.

2. **Verify explicitly**
   Never take anyone or anything at face value. Double-check everything before you act. For example: if you get an unusual request (like an unsolicited conference invite or a surprise email asking for project info) do the following:

   a. **Check the sender's email address**. Phishing emails are sneaky. They often come from addresses that look legitimate at first glance, but have small, easy-to-miss differences.
   b. **Look for signs of social engineering.** Be alert to urgent requests or offers that seem too good to be true. Attackers often manipulate emotions to fool you.
   c. **Verify via official channels.** Don't click links or reply immediately. Instead, independently verify the request. Go to the official website or contact the person through a known legitimate method to confirm the request's authenticity.
   d. Report suspicious activity. If something seems fishy, or you realize you may have interacted with a fraudulent request, report it immediately.

3. **Least privileged access**

   Limit access to only those with a need to know—both for systems and data. For example:

   a. **Restrict document access.** Apply sensitivity labels and sharing permissions so that only people who truly need a document or email can access it.
   b. **Lock down data about you.** Attackers often gather personal and work details from social media to target you. Help mitigate risk by doing the following:

- **Be selective with connections.** Don't accept unknown or unsolicited connection/friend requests. They could be an attacker looking for information about you.
- **Audit your contacts.** Regularly review your social network contacts and remove those you don't recognize or no longer trust with your info.
- **Review privacy settings:** Check social media app permissions and profile privacy settings. Limit what the public (non-connections) can see about you, such as your posts, friends list, or profile details.
- **Avoid oversharing:** Think twice before posting personal or work updates (like job changes, travel plans, birthdays) to social media. Even seemingly benign details can be weaponized to trick you later.

Identity protection practices

Think of identity protection practices as your personal firewall, constantly on watch to keep your private information safe from threat actors. Key identity protection practices to focus on include the following:

- Use a strong and unique password for each account.
- Use multifactor authentication (MFA) or passkeys for all personal accounts.
- Check regularly to see if any of your login credentials have been associated with a data breach.
- Immediately change passwords on accounts associated with a data breach.
- Keep your passwords secret.
- Store passwords only using encrypted tools like password managers.
- Slow down and approach with caution before engaging with an unsolicited message, chat, phone call, or prompt.
- Validate the identity of the sender through official channels before responding to unsolicited messages.
- Check for indicators of phishing when receiving unsolicited messages.
- Confirm that you initiated a multifactor authentication (MFA) request before approving.
- Unsubscribe from emails by blocking the sender; don't use unsubscribe links in unsolicited emails.
- Report suspicious emails, chats, phone calls, or prompts. Never respond to or engage.
- Deny unfamiliar or unsolicited social media connections or friend requests.
- Pause before posting information on social media to avoid oversharing.
- Review your social media privacy settings to manage who can view your information.
- Audit your social media connections and remove anyone you no longer want to have access to your profile and posts.

Improving your security practices

Ready to put it all into action? Here are some steps you can take right now to get started:

1. **Start with three.** Pick three security practices you want to improve. If you're not sure where to start, consider the essentials—enabling MFA, using encrypted password storage, and tightening social media privacy.
2. **Set goals.** For each chosen practice, set a specific goal for the next month. For example, "Enable MFA on all my financial accounts by the end of this month," or "Review and update privacy settings on all my social profiles this week."
3. **Practice regularly.** Consistency is key—practice the new habits until they become second nature.
4. **Check in weekly.** Once a week, evaluate your progress. Are the new habits sticking? Do you need to adjust your approach? Regular check-ins help ensure you don't forget your goals.
5. **Celebrate wins.** Acknowledge improvements and successes. Feel more confident with your security practices? Achieved one of your goals? Celebrating these milestones will motivate you to maintain and continue improving your security posture.

Every small step makes you, your organization, and everyone around you more secure. By steadily improving your practices, you contribute to a safer environment for all.

Secure. Together.

Security is everyone's responsibility. By using this guide and applying these practices, you play a crucial part in keeping yourself and your company safe from cyberattacks.

Content in this guide was developed by Microsoft's security education and awareness team with the assistance of Copilot.