Microsoft Security

How to use AI safely

Generative AI provides new opportunities for business productivity and personal fulfillment. However, with any new technology comes concerns around risk and safety.

Here are some tips to keep you and your organization safe.

3 key risk areas

1 Overreliance

What is overreliance?

Overreliance means accepting AI outputs as if they're authoritative. AIs are not search engines. They sometimes create outputs based on information that may not be factual.

How to stay safe

Research the source of the information

- Validate accuracy with independent sources.
- Check that the references listed in copilot say what you think they say. Confirm the outputs are correct.

2 Impersonation

What is impersonation?

People may use AI to impersonate colleagues or family members, hoping to trick you into giving up money or personal information. For example, an urgent request from the CFO in a video meeting asking you to approve a financial request.[1]

How to stay safe

Always verify who is communicating with you.

Set up code words with family or friends as a second factor of authentication.

Don't overshare online—social engineers will use personal details to hook you.

[1] Finance worker pays out $25 million after video call with deepfake 'chief financial officer' | CNN

3 Social manipulation

What is AI-generated social manipulation?

With AI, bad actors can create content and swarms of fake users across social and traditional media to influence viewers. This tactic is used for everything from fraud and industrial espionage to cyber influence operations.

How to manage social manipulation

Be critical of things you read on social media.

- Verify the source.
- Be wary of content posted by people you don't know.
- Validate information with sources you trust.

Be suspicious if it seems like "everyone agrees" with an idea or if there's a sudden public consensus about something. Bad actors use amplification and an appearance of trustworthiness to make content seem factual.

Be extra alert when strangers approach you online. Most fraud and influence operations start with a social media reply or a "wrong number" message.

- Assess an account's associates to help confirm their authenticity.
- Ask your network if they know the person and how.
- Help educate your network on the dangers of accepting fake friend requests.

Additional AI safety imperatives

Choose the AI deemed safe by your organization's IT or security team and adhere to all policies.

Protect sensitive data from AI usage with labels and encryption so it does not get leaked.

Break down tasks, as AIs work best and most accurately with smaller tasks. If you don't know what steps are best, ask AI for suggestions as your first step.

Explore more cybersecurity awareness topics and skilling opportunities at https://aka.ms/cybersecurity-awareness.

Microsoft Security