

Artificial Intelligence May Make Cyber Threats More Sophisticated

With the growing use of artificial intelligence in work environments, cyber threats have become more reliant on exploiting employees' behavior and trust when dealing with messages, links, and electronic requests. Risks no longer depend only on systems or technical tools; they may begin with a simple decision made by the user when reading a message, opening a link, sharing a file, or carrying out a request that appears familiar. Therefore, awareness of these methods helps in understanding the nature of modern threats and distinguishing between routine requests and those that may carry warning signs, especially when AI technologies are used to make attacks more convincing and harder to detect.

Artificial Intelligence in the Hands of Attackers: Greater Speed and Smarter Evasion

Artificial intelligence may not change the main objective of an attack, but it gives attackers advanced capabilities to develop their methods and make them more complex and effective. It makes it easier to gather information, craft precise deceptive messages, and personalize them to target the victim directly. These capabilities are mainly reflected in three key features:



Modifying attack techniques to appear different each time in order to avoid detection.



Searching for vulnerabilities or weaknesses in systems and services more quickly.



Analyzing large amounts of information rapidly to select more precise targets.



Manipulating the instructions of AI tools through prompt injection to change their responses and produce misleading or unsafe outputs.

These threats are not limited to a specific group. They may target anyone connected to the organization, whether an employee, manager, official in a certain department, or a third-party service provider. The goal is often to exploit trust or urgency to push the person into taking an action that may expose their data or the organization's data to risk.



Cyber Knowledge:

Do you know what **Prompt Injection** means?

Prompt Injection is one of the manipulation techniques that targets artificial intelligence applications. It occurs when the tool processes misleading instructions embedded within a text, message, file, link, or webpage, with the aim of influencing its response or changing the original purpose of the request.

The risk lies in the fact that the tool may treat these instructions as part of the request, which could lead to inaccurate outputs or unsafe

How does AI make cyber threats more advanced?

With the development of artificial intelligence technologies, cyber risks are no longer limited to traditional phishing messages or obvious deception attempts. They have become more diverse and more capable of exploiting trust and urgency within the workplace.

Type of Threat	Main Idea	How It May Evolve Using AI	How It May Appear to the Employee
 Phishing	Targeting employees to obtain login credentials or push them to perform an unsafe action.	Creating realistic and professional messages that imitate the organization's communication style and do not contain the usual mistakes that	An email requesting the employee to open a link, log in through a fake attachment, or update a password.
 AI-Powered Social Engineering	Influencing an employee's decision by exploiting trust or urgency.	Using publicly available information about the employee or the organization to create convincing and context-relevant requests.	A request that appears normal and comes from a known entity, such as sharing a file, confirming information, or carrying out an urgent administrative action.
 Impersonation and Deepfake	Making the communication appear as if it came from a trusted source.	Imitating writing style or voice, or creating a fake audio or video recording that appears to be from a manager, colleague, or official entity.	A message, phone call, or video clip that appears to come from a known person and requests a quick action or sensitive information.
 Business Email Compromise (BEC)	Targeting financial or administrative procedures within the work environment.	Creating accurate requests related to contracts, suppliers, bank accounts, or internal approval processes.	A request to make a financial transfer, change a supplier's bank account details, approve an invoice, or send official documents.

Pay Attention to These Warning Signs Before Taking Any Action

Even if the request appears official or comes from a known entity, stop and review the request if you notice any of the following signs:



Urgent or pressing requests that bypass the usual procedures or downplay the importance of verification.



A request to share sensitive information, carry out a financial or administrative action, or modify data such as bank account details or supplier information.



Unexpected links or attachments, or minor differences in the email address or links that may indicate a fraud attempt.



Always Remember:

As AI-powered threats continue to evolve, it is not enough for a message to appear official or for a voice to sound familiar. In the workplace, the risk may lie in the nature of the request itself, especially if it involves sharing sensitive information, changing payment details, opening an unexpected attachment, or carrying out an urgent financial or administrative action.

Therefore, always verify the source of the request through another official channel, and report any suspicious activity.

Why Should You Not Treat an AI Answer as a Final Source?



Stop
Before taking any action



Verify
The source through another channel



Report
When suspicious

