Cyber risks and artificial intelligence

05.10.2023 3 min read

Authors

Felipe Palhares

Related areas

Data Privacy and Cybersecurity

Topics

Artificial Intelligence | Data Privacy and Cybersecurity

Picture this: you, the company's CFO, receive a call from the CEO. The call shows up on your device as the CEO's personal phone number. The voice on the other end of the call is that CEO's, with the same timber, intonation, choice of words and manner of speaking.

The CEO is calling to ask you to arrange for an immediate transfer of funds to a new supplier, to meet an urgent demand. You make the transfer, like most people would. It turns out, however, that the voice on the other end of the line is not the CEO's but a digitally manipulated imitation.

This scenario is not merely hypothetical. Phone number and e-mail masking techniques and AI tools have made scams like these not only commonplace but increasingly elaborate, targeting businesses from many different sectors.

Addressing all AI-based cyber threats in a single article is impossible. Here we look at three of the most important, which should be on all companies' radar.

1. Deepfakes

Artificial intelligence systems learn from data and are excellent at replicating the data presented to them. While this feature makes them exceptional tools for replicating images, videos and sound, it also allows for the creation of deepfakes – ultrarealistic representations of a person's voice or image, like the CEO's voice in the scenario above.

Criminals have been using deepfakes to run scams and practice extortion, manipulating photos and videos to trick their victims into transferring funds, in the belief that a loved one is in trouble or because the request was legitimately made by someone in their organization.

With the advent of deepfakes, it is more important than ever for companies to establish authenticiation policies and strategies, especially for the members of their senior management.

2. Sharing of Information

In the era of ChatGPT and similar tools, which generate responses based on user input, including text excerpts and even entire documents, a fundamental security measure is monitoring confidential information that might be unwittingly shared by employees if they are unaware of the risks associated with the use of these tools.

Large language models are trained using an enormous volume of data and they store information entered by users. That information may then be used

later by the system, if it is useful in generating responses to other users' questions.

In other words, excerpts from documents that users input into generative AI systems, in order to have the text proofread, spellchecked or summarized, for example, can constitute a security incident and a breach of the document's confidentiality.

3. Third-party plugins

AI solutions often allow users to employ plugins developed by third parties to facilitate access to the solution or provide new functionalities.

Malicious actors can exploit the vulnerability of these plugins to gain access to commands entered by users of AI systems, including all information entered by the user or generated in response by the AI. In some cases, the third party is able to intercept the response and alter it, breaching not only the information's confidentiality but also its integrity.

Exploiting the exponential growth in demand for AI solutions, criminals have also developed their own plugins and pass them off as legitimate. Unsuspecting users then install the malicious plugins on their devices, allowing their creators to harm users or collect their data.

These are just some of the cybersecurity risks that come with the use of AI systems. However, awareness of the threats and robust cybersecurity policies and practices can help companies reduce their exposure to security incidents involving the complex and rapidly-evolving domain of artificial intelligence.

› PROFESSIONALS

Professionals who combine legal insight and market perspective

Felipe Palhares