# Generative AI accelerates sophistication of phishing attacks, warns IBM expert



The rapid evolution of generative artificial intelligence has raised pressing concerns across various sectors, particularly in cybersecurity. Stephanie Carruthers, IBM’s Global Lead of Cyber Range and Chief People Hacker, reflects on this transformation since the introduction of ChatGPT in 2022. Initially observed for its linguistic capabilities and vast knowledge base, the discourse soon pivoted to its potential ramifications for employment and security. Carruthers recalls the prevalent fears: “Are you scared AI is going to take your job?” Her conviction that AI would not fully replicate human ingenuity in targeted social engineering has largely shifted in light of recent developments.

As Carruthers notes, generative AI has advanced significantly within a short span, enabling the crafting of tailored phishing scams that can exploit individual vulnerabilities with minimal prompts. Initially, the era of early generative models produced generic phishing threats, but things have changed drastically. Today’s sophisticated large language models (LLMs) can autonomously search the web and gather specific information, heightening the potential danger from carefully orchestrated cyberattacks. “With very few prompts, an AI model can write a phishing message meant just for me,” she stated. “That’s terrifying.”

This alarming sentiment is echoed by various institutions, including the FBI, which has warned of malicious actors using AI to impersonate high-ranking officials. These operations are designed to infiltrate personal accounts, threatening sensitive information with far-reaching consequences. Reports indicate that such impersonation tactics are becoming ever more prevalent, demonstrating a sophisticated level of execution not seen in previous phishing attempts.

Research underscores AI's role as a tool for nation-state hackers, particularly from countries like China, Iran, North Korea, and Russia. These entities are reportedly leveraging platforms such as ChatGPT to craft customised phishing messages and gather intelligence on their targets. The growing capacity for these automated systems to create believable communications signifies a marked escalation in the tactics employed by cybercriminals, further corroborating Carruthers' concerns.

Moreover, the implications for the financial sector are stark. The IRS included generative AI-enhanced scams in its 2023 Dirty Dozen list of top tax scams, highlighting the sophisticated strategies used to defraud individuals through fake refund offers and phishing emails. Law enforcement agencies are finding it increasingly challenging to combat these advanced tactics, as cybercriminals adapt rapidly to the technological landscape.

Recent experiments conducted by IBM reveal just how effective these tools can be. In a study involving 1,600 employees at a global healthcare company, 14% fell victim to a phishing email crafted by ChatGPT within a mere five minutes. Such findings are indicative of a growing trend where phishing attempts have become harder to identify, further complicating the responsibilities of cybersecurity specialists.

Industry experts from firms such as Darktrace have pointed out that the rising sophistication of AI-generated phishing emails effectively dismantles traditional detection methods. Criminals are now able to create longer, more complex messages that evade spam filters and deploy communications that resonate more effectively with their targets, intensifying the urgency for enhanced security measures.

As generative AI continues to evolve, the interplay between its capabilities and cybersecurity risks presents an intricate challenge. The pressing need for organisations to bolster their defences against AI-enhanced cyberattacks highlights a larger conversation about the implications of AI in society. Preparedness is not just a defensive mechanism; it is a requisite in navigating this new landscape where technology and security must coalesce to safeguard against emerging threats.

**Reference Map:**

* Paragraph 1: [[1]](https://www.ibm.com/think/insights/generative-ai-social-engineering)
* Paragraph 2: [[2]](https://www.ibm.com/think/insights/generative-ai-social-engineering), [[6]](https://www.axios.com/2023/10/24/chatgpt-written-phishing-emails)
* Paragraph 3: [[3]](https://www.reuters.com/world/us/malicious-actors-using-ai-pose-senior-us-officials-fbi-says-2025-05-15/)
* Paragraph 4: [[4]](https://www.axios.com/2024/02/14/state-hackers-ai-chatbot-use)
* Paragraph 5: [[5]](https://www.axios.com/2023/04/13/tax-scams-irs-ai-chatbot-chatgpt)
* Paragraph 6: [[6]](https://www.axios.com/2023/10/24/chatgpt-written-phishing-emails)
* Paragraph 7: [[7]](https://www.theguardian.com/technology/2023/mar/29/ai-chatbots-making-it-harder-to-spot-phishing-emails-say-experts)

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

1. <https://www.ibm.com/think/insights/generative-ai-social-engineering> - Please view link - unable to able to access data
2. <https://www.ibm.com/think/insights/generative-ai-social-engineering> - Stephanie Carruthers, IBM's Global Lead of Cyber Range and Chief People Hacker, discusses the evolution of generative AI in social engineering. She highlights how AI models can now craft highly targeted phishing messages with minimal prompts, posing significant cybersecurity risks. The article emphasizes the need for enhanced defenses against AI-driven cyberattacks and the importance of understanding AI's role in modern social engineering tactics.
3. <https://www.reuters.com/world/us/malicious-actors-using-ai-pose-senior-us-officials-fbi-says-2025-05-15/> - The FBI warns that malicious actors are using AI to impersonate senior U.S. officials through text and AI-generated voice messages. These impersonations aim to gain unauthorized access to personal accounts of government officials and their contacts, leading to potential exploitation of sensitive information or funds. The FBI's advisory underscores the growing threat of AI in cyberattacks targeting high-profile individuals.
4. <https://www.axios.com/2024/02/14/state-hackers-ai-chatbot-use> - Research from Microsoft and OpenAI reveals that nation-state hackers from China, Iran, North Korea, and Russia are employing AI chatbots like ChatGPT to write phishing emails and gather intelligence on targets. This development highlights the accelerating capabilities of hackers using AI tools, confirming previous warnings about AI's potential to enhance cyberattack strategies.
5. <https://www.axios.com/2023/04/13/tax-scams-irs-ai-chatbot-chatgpt> - Generative AI programs like ChatGPT are amplifying tax scammers' efforts, as indicated in the IRS' 2023 Dirty Dozen list of top scams. These scams, including fake refund pitches and phishing emails, are becoming more sophisticated through AI, allowing swindlers to create more believable messages. The article discusses the challenges law enforcement faces in containing internet scams and advises caution with personal data.
6. <https://www.axios.com/2023/10/24/chatgpt-written-phishing-emails> - IBM research shows that ChatGPT can generate highly convincing phishing emails. In an experiment with a global healthcare company's 1,600 employees, 14% fell for a phishing email created by ChatGPT in just five minutes. This raises concerns about the potential for AI tools to be weaponized by cybercriminals, highlighting the need for enhanced cybersecurity measures.
7. <https://www.theguardian.com/technology/2023/mar/29/ai-chatbots-making-it-harder-to-spot-phishing-emails-say-experts> - Data from cybersecurity experts at Darktrace suggests that phishing emails are increasingly being written by AI bots, allowing criminals to overcome language barriers and send longer, more complex messages less likely to be caught by spam filters. The article discusses the implications of AI in enhancing social engineering tactics and the challenges in detecting such sophisticated attacks.