Case Study: Combating AI-Generated Attacks

Thieves were able to steal $243,000 using an AI-powered voice changer. Why did this happen, and how could it be prevented?

In March 2019, attackers stole approximately €222,000 ($243,000) from an energy firm in the United Kingdom (UK). The attack relied heavily on AI.

The Story

The CEO of the UK-based firm received a phone call that appeared to come from the CEO of the firm's parent company in Germany. The UK's CEO knew the German CEO well enough to be familiar with his accent and manner of speech. The voice coming from the phone matched perfectly, and it never crossed the UK CEO's mind that the voice could be coming from an AI-Powered voice changer.

The speaking through the AI-powered voice changer, the imposter requested the UK CEO to transfer funds to an account they claimed belonged to a Hungarian supplier. They provided detailed internal information and plausible reasoning for transferring funds from the UK-based firm instead of the German-based one and promised reimbursement. By the time the CEO realized they had been fooled, it was too late.

How Did This Happen?

The attacker's use of an AI-powered voice-changer caught the UK CEO off guard. The possibility of copying someone's voice so closely had not occurred to the CEO or security team; even if it had, there had been no reported cases of attacks like this before.

The CEO was aware and familiar with the dangers of social engineering attacks via email or physical means. However, the UK CEO did not expect a machine to imitate critical aspects of the German CEO nor for an attacker to have detailed information about the parent firm. As a result, the CEO only became suspicious after the transfer had been made.

How Could This Have Been Avoided?

The Voice Changer

The attack could have been foiled if there had been some system to detect AI-generated content in any form. Unfortunately, this is easier said than done. Detectors need to be trained on specific models of AI, and even when faced with content from a model they were trained on, they sometimes misclassify the content.

However, there is still hope. AI developers are experimenting with making their AI's outputs more easily identifiable. Though this does not stop someone from training their model to bypass existing detectors, it does help strengthen security against AI-generated attacks.

Policy and Procedure

The energy firm's corporate policy details were unknown. However, we know that either the policy or the enforcement of the policy needed improvement since an individual could transfer a sizeable amount without further verification. There should be requirements about what information is required to initiate a financial transfer and more verification from the individual requesting the funds.

Conclusion

Attacks involving AI will likely become more common, and the best time to prepare is now. While new tools are being developed to detect AI-generated content, AI alone should not be the only line of

defense. Reviewing and strengthening existing defenses, especially against social engineering attacks, may be as important when defending against attacks using AI.