ManageEngine
AD360

Adopting Zero Trust to safeguard against generative AI cyberthreats

Introduction

Generative AI is no longer a futuristic concept—it's here, reshaping industries, accelerating innovation, and transforming the way we work. But with great power comes great vulnerability. The same AI models that generate realistic text, images, and even code are also being weaponized by cybercriminals, giving rise to sophisticated attacks like deepfake phishing and automated malware generation.

So how do organizations harness AI's potential without exposing themselves to new cyber risks? The answer lies in a **Zero Trust security model**—a proactive approach that assumes **no user or device can be trusted by default.**

This e-book explores how Zero Trust can help businesses stay ahead of evolving AI-powered threats while ensuring secure access to critical resources.

The current landscape of generative AI

Generative AI has seen **explosive growth** in just a few years. From text-based assistants to AI-powered design tools, these models have revolutionized productivity.

But this growth comes with challenges. **Cybercriminals are exploiting generative AI in new and alarming ways:**

AI-powered Phishing-as-a-Service (PhaaS)

Phishing has evolved beyond poorly worded scam emails. AI-driven tools now enable cybercriminals to create **context-aware phishing emails in real time** that mimic a company's internal communication style. These emails adapt dynamically based on the target's response, making detection even harder.

> In 2024, phishing attacks spiked by 58%, with cybercriminals adopting AI tools to craft sophisticated, multi-channel phishing campaigns. Notably, there was a significant rise in payloadless attacks relying solely on social engineering, accounting for nearly 17.3% of phishing attempts in the first quarter of 2024 alone, up from 5.4% in 2021.

Deepfake scams: Beyond video calls

Deepfake technology is no longer just a concern for manipulated videos—it's now being used in **live voice calls** and even text-based interactions. Attackers are training AI models on publicly available recordings to impersonate **executives, business partners, and even family members.**

> In 2024, an employee at a multinational firm in Hong Kong was tricked into wiring $25 million after participating in a video call with what appeared to be their CFO. In reality, it was a **deepfake-generated face and voice**, powered by AI.

AI-assisted malware that learns

Attackers are now integrating AI into malware, allowing it to **evade traditional detection mechanisms.** AI-powered malware can analyze security tools in real-time and alter its code or behavior to remain undetected.

In 2023, security researchers discovered WormGPT, which presents itself as a blackhat alternative to GPT models, designed specifically for malicious activities. Cybercriminals use such technology to automate the creation of highly convincing fake emails, personalized to the recipient, thus increasing the chances of success for the attack.

Synthetic identity fraud

AI can generate entire **synthetic identities,** complete with fake but realistic personal details, behavioral patterns, and social media footprints. These identities are used to open fraudulent bank accounts, apply for loans, and bypass know your customer (KYC) checks.

A man was sentenced to federal prison for his involvement in a nationwide fraud ring. This group used stolen Social Security numbers, including those belonging to children, to create synthetic identities.They then opened lines of credit and established shell companies, defrauding financial institutions of nearly $2 million.

A Forrester report predicts that cybercrime will cost $12 trillion in 2025. This highlights the need for **stronger security measures** that can keep up with AI-driven threats.