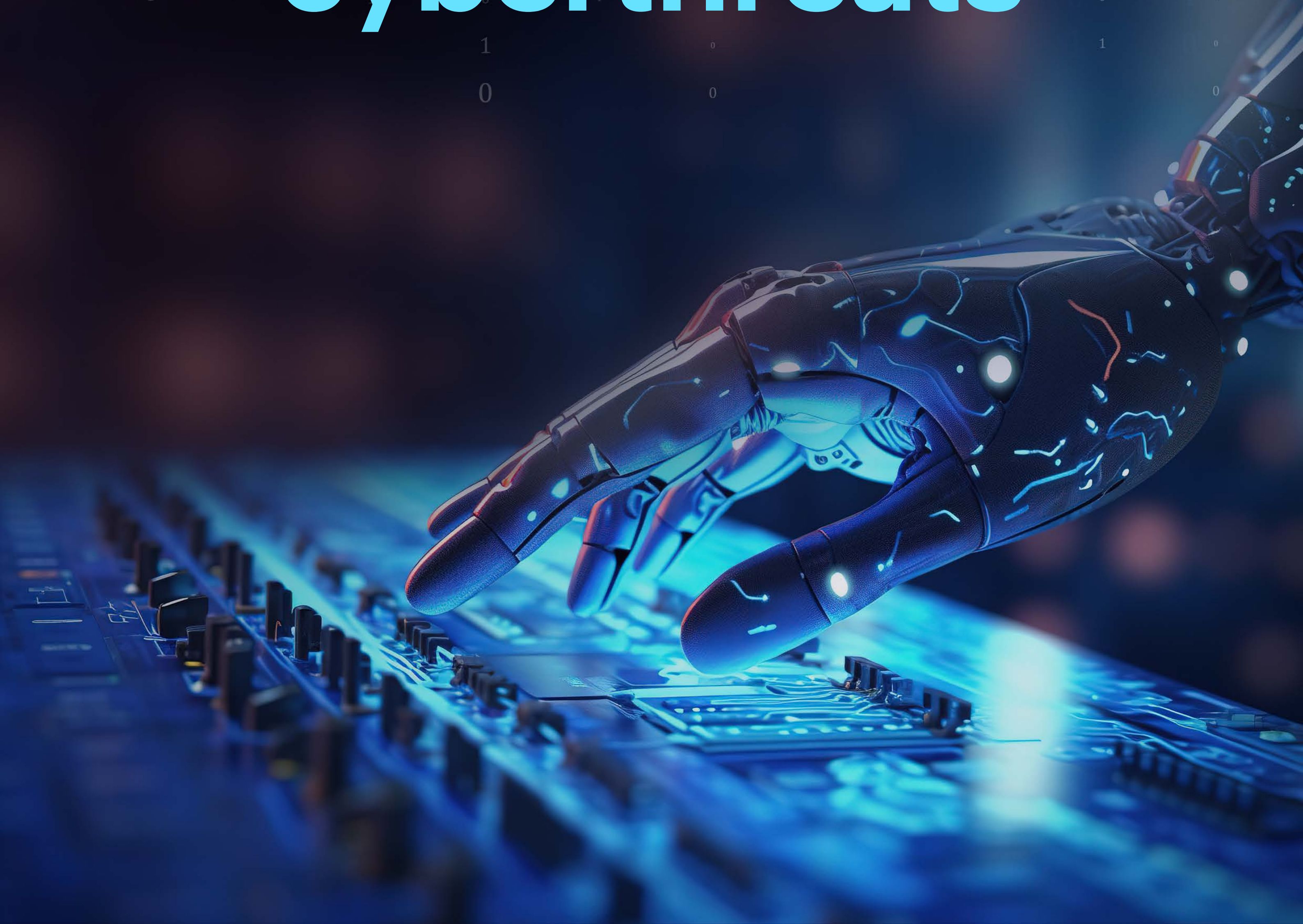
ManageEngine
AD360

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.

How generative AI works

Generative AI uses machine learning and neural networks to find patterns in data. AI models learn from large amounts of data fed to them during training. This data can be text, code, graphics, or anything relevant to the task.

Using the training data, an AI model analyzes the patterns and relationships within the data to understand the underlying rules that govern the content. As it learns, the AI model fine-tunes its parameters, enabling it to simulate human-generated content better. The more quality data AI models are fed, the more sophisticated and convincing the output will be.

To put it simply, users typically engage with generative AI by providing some type of prompt, which can be in any format the system can process. In response to the prompt, new content is returned to the user.

The most common AI techniques include:

- ✓ **Transformers:**
  AI models like GPT-4 and GPT-5 use transformer-based neural networks to generate human-like text by predicting the next word in a sequence.

- ✓ **Diffusion models:**
  Image-generation tools, like DALL·E 3 and Midjourney V2, use diffusion models to create realistic images from simple text prompts.

- ✓ **Reinforcement learning:**
  AI chatbots and assistants improve responses based on user interactions, making them more intelligent over time.

While these capabilities fuel innovation, they also **introduce new security risks**. If AI models fall into the wrong hands, they can be manipulated to **generate deceptive, harmful, or biased content**. This is why organizations must implement **strict security measures to govern AI usage and protect sensitive data.**

Examples of generative AI in 2025

Generative AI is now embedded in everyday tools, from workplace automation to creative design. Some of the most advanced AI models in 2025 include:

ChatGPT-5: OpenAI's latest conversational AI model is equipped with real-time reasoning, enhanced memory capabilities, and multimodal processing, allowing it to handle text, voice, and image inputs seamlessly. It can maintain context over long conversations, generate complex documents, and even provide real-time voice synthesis, making it an essential tool for businesses and content creators alike.

DALL·E 3: This AI-powered image generation model produces photorealistic visuals from simple text prompts. Unlike earlier versions, DALL·E 3 provides finer control over artistic styles, composition, and even object placement. It's widely used in marketing, product design, and media industries for on-demand high-quality visuals.

Gemini 2: Developed by Google DeepMind, Gemini 2 is a cutting-edge generative AI model designed for advanced research and automated content creation. It specializes in processing and understanding large datasets, making it a valuable tool for scientific research, software development, and data analysis. It's also capable of generating contextually accurate, long-form content, helping businesses automate documentation and reporting.

Midjourney V2: A revolutionary AI for digital art and branding, Midjourney V2 allows companies and artists to generate professional-grade designs instantly. It offers precise control over visual elements, making it a preferred tool for branding agencies, UI/UX designers, and illustrators looking to create high-quality graphics without manual effort.

GitHub Copilot X: This AI-powered coding assistant goes beyond simple code suggestions. It now offers real-time debugging, test case generation, and intelligent code refactoring, allowing developers to write efficient and optimized code faster. Copilot X also integrates with major development environments, making it a go-to tool for software engineers working on complex projects.

DeepSeek-R1: Launched in early 2025 by Chinese AI startup DeepSeek, DeepSeek-R1 is an open-source reasoning model that challenges the dominance of proprietary AI systems. It uses reinforcement learning trained on millions of inference traces, making it exceptionally good at complex problem-solving, logical reasoning, and mathematical computations. Unlike other AI models focused on conversational abilities, DeepSeek-R1 is designed to assist in fields like scientific research, algorithm optimization, and AI-driven decision-making.

These advancements show how AI is reshaping industries—but they also highlight the need for **strict security measures** to prevent AI from being misused.

The Zero Trust approach to AI cybersecurity

Organizations must shift from traditional security models to a **Zero Trust architecture**. Unlike perimeter-based security, Zero Trust assumes **every access request is potentially malicious** until verified.

In NIST's Special Publication 800-207, Zero Trust is described as a cybersecurity paradigm that moves the focus from securing large networks to protecting individuals or small groups. In a Zero Trust model, no implicit trust is granted to assets or user accounts, regardless of their location—whether physical or network-based. Instead, authentication and authorization involve finite steps before an enterprise resource can be accessed.

With Zero Trust, organizations can minimize their attack surfaces, minimize the impact of breaches, and improve their overall security posture by continuously verifying the identities of users and devices.

Why Zero Trust is essential against AI cyberthreats

AI-driven cyberattacks have exposed the vulnerabilities of traditional security frameworks. Attackers are now:

- ✓ Using AI-powered **password-cracking tools** to bypass weak authentication measures.
- ✓ Deploying **AI-driven malware** that adapts to evade signature-based detection systems.
- ✓ Leveraging **AI-generated deepfake voices** to conduct social engineering attacks, such as impersonating executives to authorize fraudulent transactions

A **Zero Trust model disrupts these attack vectors** by continuously verifying identities, restricting access, and monitoring behavior in real time. Instead of assuming that users or devices within the network are trustworthy, Zero Trust treats every request as potentially hostile.

Core principles of Zero Trust

1

Verify every user and device

- ✓ Implement **multi-factor authentication (MFA)** and adaptive authentication to prevent unauthorized access.
- ✓ Use **continuous identity verification** to detect compromised accounts in real time.

2

Implement least privilege access

- ✓ Enforce **role-based access control (RBAC)** to ensure users only have access to the resources they need.
- ✓ Reduce the risk of lateral movement within a network by segmenting access permissions.

3

Monitor and analyze behavior

- ✓ Leverage **AI-driven User Behavior Analytics (UBA)** to identify anomalies that indicate insider threats or compromised accounts.
- ✓ Use **real-time behavioral analysis** to detect deviations from normal user activities.

4

Enforce microsegmentation

- ✓ Isolate critical assets and **restrict network access between systems** to minimize the impact of a breach.
- ✓ Prevent unauthorized movement within a network by dynamically adjusting access policies.

5

Automate Threat Detection and Response

- ✓ Deploy **AI-powered security solutions** that can identify and neutralize threats before they escalate.
- ✓ Implement **automated incident responses** to contain breaches within seconds.

Zero Trust in action: AI-driven security controls

To counter AI-generated cyberthreats, organizations must go beyond just adopting Zero Trust as a framework—they need to **integrate AI-powered security tools** that align with Zero Trust principles. For example:

Identity threat detection systems that use AI to detect and block credential stuffing attacks.

Autonomous security orchestration that automatically adjusts access policies based on real-time risk assessments.

AI-driven deception technologies that create synthetic honeypots to mislead attackers and collect intelligence on evolving tactics.

By combining **Zero Trust with AI-powered security solutions**, organizations can build a resilient cybersecurity posture that adapts to emerging threats and prevents AI-driven cyberattacks before they cause damage.

How AD360 helps enforce a Zero Trust model

The **Zero Trust security model** operates on the principle of **never trust, always verify**, ensuring that every access request is continuously authenticated, authorized, and monitored. **ManageEngine AD360** provides the necessary controls to establish and enforce a **Zero Trust architecture** across your IT environment.

Key AD360 features for Zero Trust security

Identity and access management (IAM): Enforcing least privilege access

A core principle of Zero Trust is **least privilege access**, meaning users should only have access to the data and applications required for their roles. AD360 provides:

- ✓ **Role-based access control (RBAC)** to enforce granular access policies.
- ✓ **Automated provisioning and deprovisioning** to ensure users only have access when needed.
- ✓ **Self-service access requests with approval workflows,** reducing IT overhead while maintaining security.

MFA: Strengthening authentication

Passwords alone are no longer sufficient to prevent breaches. AD360's **advanced multi-factor authentication (MFA) capabilities** enhance authentication security with:

- ✓ **Adaptive authentication**, which analyzes user behavior and risk levels before granting access.
- ✓ **A variety of MFA methods**, including **biometric authentication, one-time passwords (OTP), push notifications, and hardware tokens.**
- ✓ **Context-aware MFA**, enforcing stricter authentication for high-risk activities like accessing sensitive applications from unknown devices or locations.

UBA: Detecting anomalies and insider threats

A Zero Trust model requires **continuous monitoring and risk-based analytics**. AD360 integrates **AI-driven user behavior analytics (UBA)** to:

- ✓ **Detect unusual login patterns**, such as logins at odd hours.
- ✓ **Flag excessive failed login attempts**, potentially indicating brute-force attacks.

SSO: Reducing attack surfaces

Managing multiple passwords increases the risk of **credential-based attacks**. AD360's **single sign-on (SSO)** feature improves security by:

- ✓ **Providing a single authentication mechanism** for multiple enterprise applications.
- ✓ **Reducing password fatigue,** lowering the likelihood of credential reuse.
- ✓ **Enhancing visibility into user access across** cloud and on-premises environments.

Enforcing Zero Trust with AD360

By leveraging **AD360's comprehensive IAM framework**, organizations can:

- ✓ **Enforce strict access controls** with least privilege policies.
- ✓ **Continuously authenticate users** with adaptive MFA and UBA.
- ✓ **Minimize attack surfaces** with centralized identity governance.
- ✓ **Detect and respond to threats in real-time** with AI-driven analytics.

By leveraging AD360's comprehensive security framework, organizations can **enforce Zero Trust policies, safeguard AI-driven environments, and mitigate cyber risks.**

Securing the future

Generative AI has transformed the digital landscape, unlocking **endless possibilities—but also new security challenges.** With cybercriminals increasingly using AI for malicious purposes, businesses can no longer rely on **traditional security methods.**

To combat generative AI's malicious applications, it’s crucial to continually develop and implement robust defenses, enhance detection capabilities, and stay vigilant to emerging threats—before you become the next victim of an AI-generated attack.

Want to strengthen your security posture?
Let’s discuss how AD360 can help your organization implement Zero Trust today.

About AD360

ManageEngine AD360 is a unified identity platform that seamlessly connects people, technology and experiences while giving enterprises full visibility and control over their identity infrastructure. It offers automated life cycle management; secure SSO; adaptive MFA; and risk-based governance, auditing, compliance and identity analytics—all from a single, intuitive console. With extensive out-of-the-box integrations and support for custom connectors, AD360 easily integrates into existing IT ecosystems to enhance security and streamline identity operations. Trusted by leading enterprises across healthcare, finance, education, and government, AD360 simplifies identity management, fortifies security and ensures compliance with evolving regulatory standards.

For more information, please visit www.manageengine.com/active-directory-360/.

$ Get Quote ⤓ Download