ManageEngine
AD360

Adopting Zero Trust to safeguard against generative AI cyberthreats

Adopting Zero Trust to safeguard against generative AI cyberthreats

Generative AI, as defined by the World Economic Forum, refers to algorithms that create new outputs based on the data they are trained on. The generative AI system isn't like traditional AI, which follows predetermined patterns and rules; instead, it learns from data without explicit instructions to create new content like audio and art. As generative AI advances, it also poses security risks, since they can be used to make convincing fake videos or phishing emails, potentially leading to misinformation, fraud, and privacy breaches.

In this e-book, we'll address generative AI and cybersecurity, but we'll begin by learning more about generative AI and how it works.

The rise of generative AI

Let’s take a look at the history of generative AI from past to present and beyond.

The root of generative AI can be traced back to the early days of machine learning. In the late 1950s, scientists and researchers began to explore the idea of using algorithms to create new data. At that time, one of the earliest examples of generative AI was the Markov chain, a statistical model capable of producing new sequences of data based on the input it received.

With advancements in computing power and algorithms, generative AI has made significant progress. During the 1980s, researchers developed more sophisticated techniques, like Bayesian networks and Hidden Markov models, to generate data and model complex relationships.

In the 1990s, deep learning emerged as a new form of machine learning that uses artificial neural networks to gain knowledge from data. Deep learning models comprehended much more complex relationships between data points than previous models, and this led to a new wave of AI progress.

The 2000s saw the birth of generative adversarial networks (GANs), a revolutionary deep learning model. Competing against each other, GANS are neural networks: one network generates new data, while the other distinguishes between real and generated data. This process helps the generator network learn how to create data that is realistic and can distinguish between real and fake data.

Recent years have seen huge advancements in generative AI, including the GPT-3 model from OpenAI. These models are used for creating art and music, developing new products, and improving healthcare. As technology advances and data access increases, generative AI is expanding and evolving, providing new opportunities for innovation and discovery.

Generative AI has captured interest and headlines recently, and holds immense potential for exponential growth in the future. AI-generated content will become even more sophisticated and creative as research continues, blurring the lines between human and AI-created content.

How does generative AI work?

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. It 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 content that AI models generate, 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.

Examples of generative AI

Here are some of the most popular AI interfaces.

ChatGPT: It is trained to converse in natural language. It can write stories, essays, poems, and recipes, as well as engage in back-and-forth conversations with you.

DALLE-E: It also uses natural language processing to generate new images. DALL-E incorporates descriptive text to generate photorealistic imagery based on the prompt. Images in varying perspectives and styles can also be generated.

BARD: Like ChatGPT, Bard is powered by Google's AI technology. The AI system is trained on a large collection of text documents and code, and is capable of generating text, translating languages, writing creative content, and answering questions.

Midjourney: It creates realistic and artistic images based on text prompts using a Discord bot. You can edit, upscale, and download your creations. It also helps create descriptions for images based on user input.

A double-edged sword

Any industry or organization can benefit from generative AI to increase productivity, automate tasks, enable new forms of creation, and develop synthetic data for training AI models. While this technology certainly has its advantages, some concerns need to be addressed. Let's explore the positive and negative aspects of generative AI.

The power of generative AI

- Automates and speeds up task performance to increase productivity.
- Removes or lowers skill and time barriers for generating content and developing creative applications.
- Enables exploring and analyzing complex data.
- Helps create synthetic data for training and improving other AI systems.

Don't forget the risks of generative AI

- Generative AI can create convincing deepfakes, which can be used to spread misinformation, defame people, or harm reputations.
- Using sensitive or proprietary data to train generative AI models can lead to inadvertent data leakage through generated content.
- Hackers can exploit vulnerabilities in generative AI apps to gain unauthorized access, manipulate data, or disrupt the system.
- Inadequate testing, weak access controls, and improper handling of data can lead to security breaches in generative AI models.

In Forrester's Top Cybersecurity Threats In 2023 report, AI applications such as ChatGPT are listed among the emerging threats, and Gartner analyst Avivah Litan highlights five major risks associated with generative AI: fabricated information, deepfakes, data privacy, copyright issues, and cybersecurity.

In terms of cybersecurity, generative AI poses a core concern because it could provide threat actors with great powers to develop malicious exploits and conduct more effective cyberattacks. So, how do hackers use generative AI in their attacks?

How generative AI is used in cyberattacks

As generative AI becomes more widespread, the potential for cyberattacks leveraging this technology also grows. The use of generative AI by cybercriminals might pose serious threats to organizations' security and data integrity. Let’s explore some of the risks of generative AI in cybersecurity:

Credential Phishing: AI is used in credential phishing to create emails that look like they came from a legitimate company or fake websites that look real. By doing so, users are more likely to reveal sensitive information or input their credentials.

Endpoint exploitation: This involves using AI to automate the process of finding and exploiting vulnerabilities, creating sophisticated attack vectors, and evading detection.

Business email compromise (BEC): Using AI in BEC attacks, the emails seem more real and convincing. AI could, for instance, analyze a CEO's email patterns and create an email that mimics their writing style. This increases the chances that the recipient will trust the email and follow the instructions.

Malware creation: AI is used in malware creation to create code that's harder to detect by antivirus software. For example, AI can be used to automate malware development. By analyzing existing malware samples and learning from their behavior, AI algorithms can create complex, polymorphic malware that is constantly evolving and adapting to escape detection by antivirus software.

To counter generative AI cyberattacks, organizations must adopt the Zero Trust security model. Zero Trust emphasizes the principle of "never trust, always verify," and requires constant authentication and authorization for every user, device, or application attempting to access the network or data.

Zero Trust adoption

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. Based on its physical or network location in Zero Trust, no implicit trust is granted to assets or user accounts. 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.

In 2023, a survey of security leaders in 31 countries, representing nearly every industry and the public sector, found that only 28% of the organizations deployed a full Zero Trust solution to remedy risks from cyberattacks. This means that organizations must prioritize the implementation of comprehensive Zero Trust strategies to bolster their security posture and protect sensitive data from sophisticated cyberthreats.

The Zero Trust approach involves the following security measures:

Identity and access management (IAM): IAM is crucial in information security since it determines who can access what resources and under what conditions. Each request is verified and authorized based on multiple factors, like identity, context, device, location, and behavior. In IAM, authentication, authorization, and access management are the key components.

Continuous monitoring and auditing: To detect suspicious activity or anomalies, Zero Trust monitors user and device behavior. It involves the collection and analysis of data from various sources, including network traffic, user activity logs, and endpoint security solutions. It gives organizations real-time data and implements security procedures like incident response, threat assessment, computer and database forensics, and root cause analysis.

User and entity behavior analysis (UEBA): By implementing UEBA solutions, anomalies such as unauthorized access attempts and unusual data transfers are detected. With UEBA tools prioritizing the intensity of the threat, organizations can determine to focus on high-risk issues first. In addition, UEBA helps detect a broader range of attacks, like DDoS attacks, brute-force attacks, data exfiltration, and insider attacks.

Incident response (IR): IR involves identifying attacks, understanding their severity, prioritizing them, investigating and mitigating them, restoring operations, and preventing their recurrence. Organizations can minimize the impact of incidents, reduce downtime, and protect their reputation by implementing IR plans.

Threat intelligence: The purpose of threat intelligence is to collect, analyze, and share information about security threats to an organization's services, networks, and data. This includes what hackers are doing, how they're doing it, and what vulnerabilities they're exploiting. Using threat intelligence, security teams can understand threats better and respond more proactively to protect their organization's assets.

How ManageEngine can help enforce a Zero Trust model

Zero Trust networks are an additional layer of security against generative AI cyberattacks. For example, in credential phishing, Zero Trust serves as a barrier thwarting cybercriminals from advancing their malicious goals. When MFA is required to access an account, a cybercriminal will have a difficult time logging into the account. By adhering to the Zero Trust paradigm, even if one account is compromised and linked to multiple applications, an attacker's ability to access other applications remains restricted. Furthermore, monitoring user behavior regularly minimizes external and insider threats by detecting potential threats early.

To enforce a Zero Trust model and strengthen your cybersecurity posture, ManageEngine provides these capabilities:

Multi-factor authentication (MFA): Adding a second layer of authentication on top of existing passwords helps thwart unauthorized access, even if attackers manage to obtain certain credentials through phishing attacks. With ManageEngine AD360, an enterprise IAM solution, you can implement MFA for:

- Windows, macOS, and Linux machines
- Top VPN providers like Fortinet, Cisco AnyConnect, Pulse, and more
- Endpoints supporting RADIUS authentication such as Citrix Gateway, VMWare Horizon, and Microsoft Remote Desktop Gateway (RDP)
- Outlook Web Access (OWA) logins

There are 19 different authentication factors including fingerprint, face ID, email verification, Duo security, Microsoft Authenticator, Google Authenticator, and YubiKey Authenticator. You can view the complete list of supported authenticators here.

Passwordless authentication: Passwordless authentication eliminates the need to create passwords, reducing the risk of credential theft. ManageEngine AD360 provides enterprise SSO that gives users seamless, one-click access to:

- SAML-enabled applications
- OAuth and OpenID Connect-enabled applications
- Custom applications

You can also control who can access which cloud applications by creating policies based on domains, OUs, and groups.

Principle of least privilege: Granting each user access to only the essential resources and nothing more will narrow down the risk if the user's credentials are compromised in a security incident. ManageEngine AD360 provides:

- Prebuilt reports, which helps to view and manage efficiently both new and inherited access permissions to file shares.
- Templates that help admins grant only the necessary permissions to new hires. These templates automatically fill in the required permissions like group memberships, file server permissions, and more based on their role or designation.
- Automated time-bound permissions management that enables the temporary assignment of users to specific groups and to grant file server permissions.

User and entity behavior analytics: In UEBA solutions, anomalous user behavior is detected by analyzing patterns and deviations from normal behavior. In addition, real-time alerts can be set up to notify security teams of suspicious activities, which allows them to investigate and respond to incidents quickly. With ManageEngine Log360, a unified SIEM solution, you can:

- Resolve high-risk threats quickly by assigning risk scores to different security events.
- Detect anomalies more accurately and reduce false positives by grouping users based on their behaviors and setting alerts for deviations from established baselines.
- Reduce dwell time by getting real-time, behavior-based security alerts via SMS or email.

Incident detection and response: To minimize the effects of a security incident, an organization must have an effective incident management process in place. ManageEngine Log360's incident management system helps security teams respond to cybersecurity threats quickly and efficiently. Using this solution, you can:

- Improve incident management with an incident dashboard that displays active and unresolved incidents alongside recent and critical incidents, tracks key metrics, and prioritizes incident resolution for optimal SOC performance.

- Identify attack patterns such as SQL injection, denial-of-service, and firewall attacks through 30+ predefined correlation rules.
- Prioritize high-risk incidents and expedite threat resolution with a real-time alerting system that includes over 1,000 predefined alert criteria.
- Automate your response to security threats with workflow management that instantly addresses insider threats, compromised accounts, and data exfiltration attempts.
- Integrate with external help desk software like ManageEngine ServiceDesk Plus, Zendesk, ServiceNow, and Kayako to generate tickets when alerts are triggered.

Threat intelligence: In the age of generative AI cyberattacks, organizations need to incorporate threat intelligence into their security strategy to stay ahead of the game. ManageEngine Log360 helps you:

- Obtain threat insights from STIX/TAXII-based feeds like Hail A TAXII, AlienVault OTX, or custom sources.
- Automatically trigger workflows to permanently block blocklisted IPs by adding them to the firewall.
- Protect data by detecting outbound communications; notifying analysts about malicious IPs, domains, or URLs; and promptly blocking them.
- Mitigate false positives using a real-time event response system that differentiates suspicious activities from legitimate ones.

Securing the future

As AI advances, risks targeting these technologies are also evolving. The wide availability of tools like ChatGPT and Google Bard has granted malicious actors the ability to rapidly elevate attack complexity seemingly overnight. IT professionals need to proactively embrace strategies to shield the organization from these threats.

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.

Our Products

Log360 | ADManager Plus | ADAudit Plus | ADSelfService Plus

Exchange Reporter Plus | RecoveryManager Plus

ManageEngine
AD360

ManageEngine AD360 is a unified identity and access management solution that helps manage identities, secure access, and ensure compliance. It comes with powerful capabilities like automated identity life cycle management, secure SSO, adaptive MFA, approval-based workflows, UBA-driven identity threat protection, and historical audit reports of AD, Exchange Server, and Microsoft 365. AD360's intuitive interface and powerful capabilities make it the ideal solution for all your IAM needs, including fostering a Zero Trust environment.

$ Get Quote Download