Solution Brief

SkopeAI for ChatGPT and Generative AI

Introduction

The emergence of artificial intelligence (AI)-based SaaS applications has brought about a transformative shift in how corporate users engage with their daily work. Generative AI applications, such as ChatGPT, have opened countless possibilities for organizations and their employees to enhance business productivity, simplify tasks, improve services, and streamline operations. With ChatGPT, teams and individuals can conveniently leverage its capabilities to generate content, translate text, process data, build financial plans, and debug and write code, among other uses. Generative AI applications also create enormous and unprecedented data security risks.

The Data Security Challenge

While AI applications have the potential to improve work efficiency, they also introduce new risks and expose sensitive data to external threats. Organizations need to address these challenges to ensure the confidentiality, integrity, and security of their data. Here are some examples of how sensitive data can be exposed to ChatGPT and other cloud-based AI applications:

- Text containing Personally Identifiable Information (PII) can be posted and therefore exposed on the chatbot to request email ideas, customer responses, personalized letters, or sentiment analysis.
- Confidential health information, including individualized treatment plans and medical imaging data, may be entered into the chatbot, potentially compromising patient privacy.
- Software developers may upload unreleased proprietary source code for debugging, code completion, and performance improvements.
- Software developers could even directly connect a corporate app, containing source code or a database, to generative AI apps via API. This cross-app data movement enables automatic synchronization of information in the cloud and facilitates routine tasks such as refining code structure and improving readability. However, it is important to note that such access could potentially expose confidential data to an unsafe third party application.
- Files of confidential company documents, such as earnings report drafts, M&A documents, and pre-release announcements, might be uploaded for grammar and writing checks, negligently risking potential data leaks.
- Financial data, including corporate transactions, undisclosed revenue, credit card numbers, and customer credit ratings, can be processed by ChatGPT for financial planning, compliance, fraud detection, and customer onboarding, without any security measures.
- Within the marketing department, an employee could in the future integrate the complete customer database in Salesforce.com with a ChatGPT and other generative AI powered plug-ins and many other unsanctioned apps via an OAuth integration. This cross-app integration will empower the employee to leverage the capabilities of GPT, enabling them to potentially automate the process of composing emails to contacts whose contracts are nearing expiration. This is another example of cross-app data movement that cannot be detected by inline network solutions like firewalls and secure web gateways (SWG).

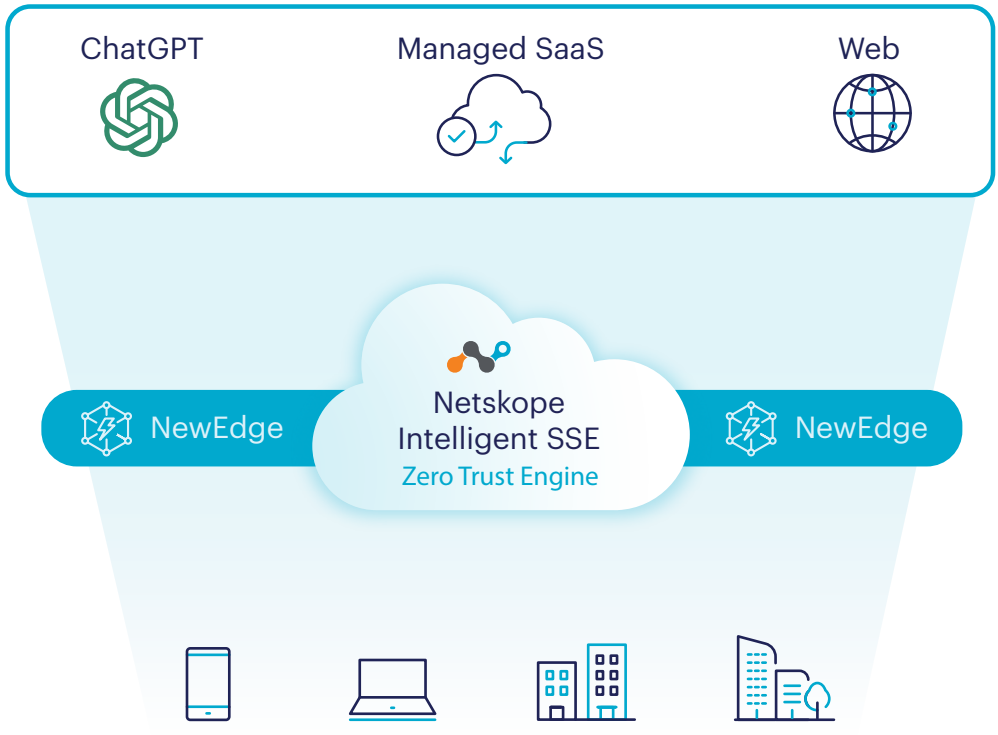
Securing Sensitive Data in the Cloud

Organizations must prioritize robust countermeasures to safeguard the confidentiality and security of sensitive data across managed and unmanaged SaaS applications, and through personal instances and personal accounts. For example, 74% of data theft flows into personal cloud storage instances of popular apps, according to a recent Netskope Cloud Threat Report.

The following key steps are essential for protecting sensitive information, and must be considered core capabilities for modern data protection technology:

1. **Monitoring and Risk Management:** Implement monitoring mechanisms to track the use and potential misuse of risky SaaS applications and instances. Conduct risk assessments regularly to identify vulnerabilities and address them promptly.
2. **Data Minimization and Access Controls:** Limit the exposure of sensitive information through SaaS applications by adopting data minimization strategies. Employ stringent access controls to ensure that only authorized individuals can access and manipulate sensitive data.
3. **Encryption and Data Loss Prevention:** Apply strong encryption techniques to protect data both at rest and in transit. Deploy data loss prevention (DLP) solutions to monitor and prevent accidental data loss or theft.
4. **User Awareness and Training:** Educate employees about the risks associated with AI-based SaaS applications and train them on best practices for handling sensitive data securely. Foster a culture of data protection and emphasize the importance of responsible usage.

As organizations embrace cloud-based services and AI-driven applications like ChatGPT, ensuring the security of sensitive data becomes paramount. By implementing comprehensive data protection measures, including monitoring, access controls, encryption, and user training, organizations can mitigate risks, safeguard confidential information, and maintain compliance in the ever-evolving cloud-enabled world.



General precautions and security recommendations with the use of Generative AI applications

Using AI models like ChatGPT in a business setting offers significant advantages in terms of productivity, efficiency, and innovation. However, ensuring data privacy and security is crucial when utilizing such AI models. Here are some enhanced best practices for security teams and employees to protect company's data:

1. **Local Deployment:** Whenever possible, deploy AI models locally on your company's machines. This eliminates the need for data to leave your company's network, reducing the risk of data leakage.
2. **Data Anonymization:** Instruct corporate users to spend some time anonymizing or pseudonymizing sensitive data before utilizing it in AI models. This involves replacing identifiable data with artificial identifiers. Even if leaked, the data would be useless without the original identifiers.
3. **Data Encryption:** Whenever possible, implement encryption both at rest and in transit for the most confidential corporate data. This ensures that even if the data is exposed, it remains unreadable without a decryption key.
4. **Strict Access Control:** Utilize robust access control mechanisms to corporate resources and data repositories to restrict interaction with AI models and the associated data.
5. **Audit Trails:** Maintain detailed audit logs of all activities related to data handling and AI model operations. These logs aid in identifying suspicious activities and serve as a reference for future investigations.
6. **Data Minimization:** Train all employees to adhere to the principle of using the minimum amount of data necessary for effective functioning of the AI model. By limiting data exposure, the potential impact of a breach can be reduced.
7. **Regular Updates and Patches:** Stay vigilant in keeping local software up to date with the latest patches and updates. This safeguards against known vulnerabilities.
8. **Third-party Audits and Certifications:** Choose AI services from providers who have undergone rigorous third-party audits and possess certifications such as ISO 27001, SOC 2, and GDPR compliance.
9. **Data Usage Policy:** Establish clear policies on data handling and usage within your organization. Ensure that employees are well-informed about these policies and understand the significance of data security.
10. **Data Backup:** Regularly perform data backups to ensure the ability to restore data in the event of any loss or compromise.
11. **Constant Review:** It is always advisable to review the most current usage policies and terms of service of any AI tool to understand how they utilize data sent via the API to improve their models.

How Netskope secures sensitive data in the use of Generative AI applications

Netskope is a market leader in cloud security and data protection, with over a decade of experience, offering the broadest visibility and the finest control over thousands of new SaaS applications like ChatGPT. Netskope offers SkopeAI for GenAI, a security solution that specifically addresses the use of Generative AI apps like OpenAI ChatGPT, Bing AI, Google Bard and many more. Here are some core technology capabilities that Netskope offers to information security teams to protect sensitive data, with emphasis on how these capabilities are easily leveraged for securing ChatGPT and other generative AI tools:

Application access control

1. It all starts with visibility. Netskope provides automated tools for security teams to continuously monitor what applications (such as ChatGPT) corporate users attempt to access, how, when, from where, with what frequency etc. It is essential to understand the different levels of risk that each application poses to the organization and have the ability to granularly define access control policies in real-time based on categorizations and security conditions that may change over time.
    - As an example, security teams would greatly benefit from gaining insights into the extensive range of applications used by their corporate employees. With many thousands of new apps available, the ability to filter and categorize them by name, usage, or category (such as ChatGPT, social, collaboration, file repositories, and more) is essential. Moreover, it is important for security teams to understand the risk level, compliance standards, activities, and usage details of each app.

2. While applications that are more explicitly malicious should be blocked, when it comes to access control, oftentimes the responsibility of the use of applications like ChatGPT should be given to the users, tolerating and not necessarily stopping activities that may make sense to a subset of business groups or to the majority of them. At the same time security teams have the responsibility to make employees aware of applications and activities that are deemed risky. This can mainly be accomplished through real-time alerts and automated coaching workflows, involving the user in the access decisions after acknowledging the risk. Netskope provides flexible security options to control access to generative AI-based SaaS applications like ChatGPT and to automatically protect sensitive data.
    - Examples of access control policies include real-time coaching workflows that are triggered every time users open ChatGPT, such as customizable warning popups offering guidance about the responsible use of the application, the potential risk associated and a request of acknowledgement or justification.

Advanced detection and safeguarding of sensitive data

Users make mistakes and may negligently put sensitive data at risk. While access to ChatGPT can be granted, it is paramount to limit the upload and posting of highly sensitive data through ChatGPT directly and indirectly and across other potentially risky data exposure vectors in the cloud. This can only be accomplished through Netskope's modern data loss prevention (DLP) techniques and advanced cloud security controls. With Netskope's data loss prevention (DLP), powered by ML and AI models, thousands of file types, personally identifiable information, intellectual property (IP), financial records and other sensitive data are confidently identified and automatically protected from unwanted and non-compliant exposure. Netskope detects and secures sensitive data in-motion, at-rest and in-use and through every possible user connection, in the office, in the datacenter, at home and on the road.

1. First, Netskope advanced DLP is able to automatically identify flows of sensitive data, and categorize sensitive posts with the highest level of precision.

   Accuracy ensures that the system protects every piece of information that is sensitive in any structured and unstructured format including images, screenshots, compressed files, clipboards, chat messages etc. It is also a fundamental aspect to ensure that only sensitive data is detected, and not harmless queries and safe tasks through the chatbot. This is achieved automatically through a comprehensive set of data detection technologies and advanced classification algorithms that include both manually defined data discovery rules and automated detection engines such as deep learning techniques, natural language processing (NLP), sentiment and semantic analysis. Deep learning and NLP leverage supervised and unsupervised machine learning for complex tasks, very similar in the approach to the ones used by generative AI models.
2. Netskope DLP provides artificial intelligence (AI) and machine learning (ML) -based image classification, together with optical character recognition (OCR), with the ability to automatically recognize sensitive files and document types based on multiple identifiable characteristics. These models also leverage Convolutional Neural Network (CNN) algorithms and YOLOv5 vision AI to analyze visual imagery. Precisely, these techniques allow the system to automatically detect electronic images of passports, driver's licenses, photo IDs, tax forms, medical cards, source code, social security cards, credit/debit cards, resumes, NDA documents, patents, M&A and checks, just to mention a few, with higher accuracy and performance, even when such images and documents are partially corrupted, crumpled, blurry and generally not clearly sharp.
3. Netskope DLP also gives the ability to build custom ML-based classifiers. With the Train Your Own Classifier (TYOC) functionality, fundamentally based on supervised ML, organizations train the system to learn to identify new unique data sets in the form of PII-free irreversible ML features.
4. It is important to ensure the utmost protection of proprietary mission-critical documents preventing any unauthorized exfiltration or duplication. File and document fingerprinting techniques can be employed to index entire documents and identify precise or partial copies of the information they contain. Particularly, Netskope DLP can examine semantic deep learning embeddings of sequences of words in the documents. It then encodes the embeddings to numeric vectors and then calculates cosine similarities. By detecting similarities in content across different environments and transmission channels, these techniques enhance the ability to identify and prevent unauthorized dissemination.

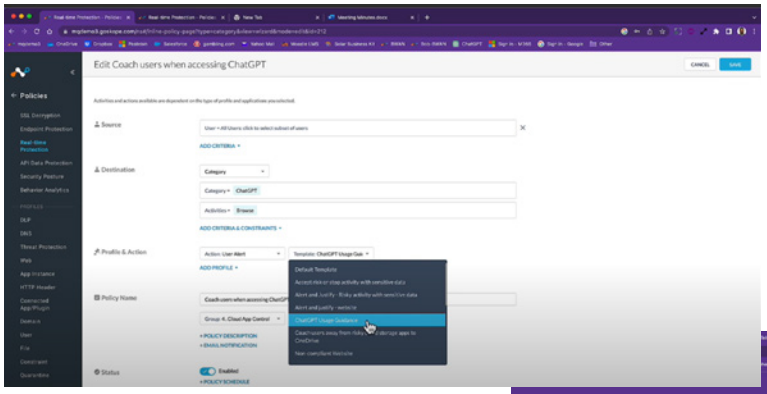
Real-time data protection and automatic user coaching

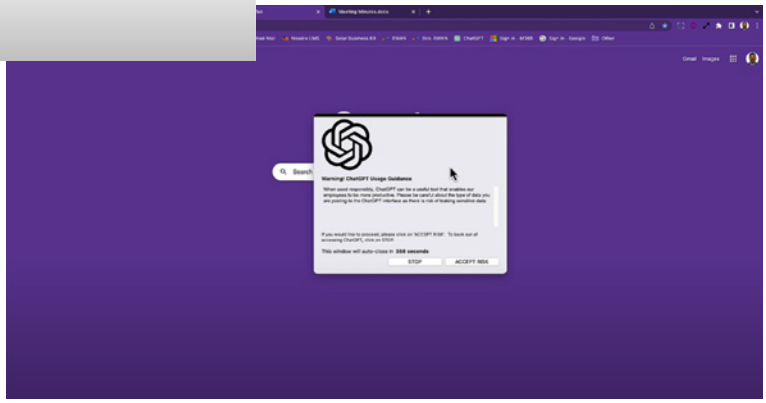
1. Netskope DLP offers several enforcement options to stop and limit the upload and posting of highly sensitive data through ChatGPT. This real-time enforcement applies to every user connection, ensuring data protection in the modern hybrid work environment where corporate users connect from the office, home, and while on the road. For example, in the case of ChatGPT and other generative AI apps, in addition to selectively stopping uploads and posts of sensitive information, visual coaching messages can be automated in real-time to provide guidance about data posting violations, inform the user on corporate security policies and allow to mitigate repeated risky behavior over time taking away the burden from security response teams.

   Netskope DLP is natively integrated to the comprehensive Netskope Security Service Edge (SSE) solution, and is continually aware of users behavior, geolocation, security postures, device risks, application risks and reputations, personal application instances etc. This allows DLP to automatically adapt to the evolving risk context and tailor the security response to different situations.

2. With generative AI apps, protecting data uploads may not be just enough. Developers, for example, can now integrate ChatGPT and other models into their apps and products through API, or ChatGPT derivates (e.g. AutoGPT) to their workflows. So a developer could link a proprietary source code, an entire database in the cloud, an online 365 excel sheet or provide full access to an application. Only monitoring sensitive data inline through traditional firewalls and DLP will miss these egress routes when the data is already in the cloud and not flowing inline. Netskope provides a comprehensive data protection solution for SaaS applications that discovers and protects sensitive data inline and in the cloud. The solution help selectively prevent sensitive data from being transferred to the cloud, and safeguards against unauthorized cross-application access to sensitive data already in the cloud. Additionally, Netskope provides visibility into cloud-to-cloud integrations for risk assessment and mitigation.

   As new capabilities and app ecosystems are developed, this approach provides the best and most comprehensive data protection to safeguard confidential information, source code on developer apps, customer databases on Salesforce.com and much more, limiting or preventing sensitive data exposure to untrusted ecosystem apps, including generative AI-based apps.





Other Netskope controls and final considerations

- Netskope also allows robust access control mechanisms based on zero trust principles to corporate data repositories to restrict interaction with AI models and the associated data. This significantly mitigates the risk of internal threats.
- Another important security measure is the ability of identifying malicious user behavior, repeated offenders behavioral anomalies. Netskope integrated user and entity behavior analytics UEBA is a component of the Netskope security platform that focuses on analyzing user and entity behavior to detect and mitigate potential security threats. UEBA solutions utilize advanced analytics and machine learning algorithms to monitor user activities, network traffic, and data access patterns to identify anomalous or suspicious behavior. Netskope UEBA specifically aims to provide insights into user behaviors, such as their interaction with cloud applications, data transfers, login activities, and data access permissions. By analyzing these behavioral patterns, Netskope UEBA helps organizations identify insider threats, compromised accounts, data exfiltration attempts, and other security risks.
- In addition to the above-described use cases, Netskope offers a comprehensive set of AI and ML-based security capabilities in the platform, which include:
  - Advanced ML models for malware detection, complementing more traditional signatures, heuristics methods, and sandboxing techniques.
  - Anti-phishing and URL filtering, featuring automated URLF signature generation, DGA detection, fast flux domain detection, web content filtering, and categorization.
  - IoT Security, providing IoT device classification and identification, dynamic device grouping, and anomaly detection.
  - WAN access anomaly detection.
  - Workflow automation, app health monitoring, cloud autoscaling, and adaptive incident prioritization.

To learn more visit:

www.netskope.com/skopeai

www.netskope.com/solutions/netskope-for-chatgpt-and-generative-ai

www.netskope.com/products/security-service-edge