

Say YES to AI



Without Putting the Business at Risk

Deploy Normalyze Data Security Posture Management to take advantage of generative AI applications and Large Language Models with the confidence that your critical data remains secure.

The AI attack surface

As organizations increasingly adopt AI, the attack surface expands, introducing new and complex risks that security teams must address. According to a McKinsey Global Survey from 2023, 50% of organizations reported using AI in at least one business function, up from just 20% in 2017. This rapid adoption, while driving innovation and productivity, also brings significant security challenges.

Key Risks Introduced by AI:

- Unintended Data Exposure:** AI models like LLMs and copilots, such as Microsoft Copilot, can inadvertently expose sensitive data if proper controls are not in place. These tools often use any data they encounter, increasing the risk of public exposure and data breaches.
- Access Governance Challenges:** Many organizations lack robust access controls, allowing internal users broader access to AI tools than necessary. This can lead to unauthorized access to sensitive data, increasing the risk of breaches.
- Regulatory Compliance Risks:** AI models are not inherently designed to adhere to stringent regulations such as GDPR, HIPAA, and CCPA. This non-compliance can result in severe legal and financial consequences, especially regarding data usage, protection, and auditability.
- Lack of Audit and Traceability:** AI models often operate as "black boxes," making it difficult to ensure compliance with regulations requiring decision-making transparency and audit trails. This opacity can lead to challenges in tracing data usage and ensuring that sensitive information is handled correctly.

» These risks, combined with the pressure to rapidly deploy AI solutions across organizations, can lead to significant security gaps if not addressed proactively.

The foundation for an AI strategy

To securely run AI within their organizations, security teams need to start with a complete picture of their critical data, including where it is, what it is, when it is used, which users and resources use it, and why.

- Identify and label data** Labeling data, for example with Microsoft Information Protection (MIP) labels, ensures that AI models and systems are aware of the sensitivity of the data they handle, allowing for appropriate handling and processing and enforcement of policies. MIP labels provide an audit trail and ensure data is handled according to legal and organizational standards.
- Implement access governance** Ensuring that only authorized users and systems can access the data feeding into custom models and Retrieval-Augmented Generation (RAGs) helps protect sensitive information from unauthorized access, reducing the risk of data breaches. It also helps reduce the amount of sensitive or unnecessary information exposed to AI models, thereby limiting potential attack surfaces.
- Conduct real-time sensitivity analysis** Preventing sensitive data from being inadvertently included in model training or inference processes reduces the risk of data misuse or accidental exposure, which could lead to reputational damage or legal consequences. It also provides full governance and visibility into data usage.

»

Learn more at: normalyze.ai

Deploy AI securely with Normalyze

A recent Gartner survey found that 48% of organizations expressed concerns about data privacy and security when deploying AI, primarily due to the risks of unintended data exposure through LLMs and AI copilots.

Security teams typically need to address at least one of the three different scenarios depending on how they are deploying AI in their organization: AI Copilots, custom LLM and AI applications, and Retrieval-Augmented Generation models. Below is a quick overview of each, the risks associated and how Normalyze DSPM solves those challenges.

AI Copilots

Overview:

- » AI copilots like Microsoft Copilot, Google Gemini, and GitHub Copilot introduce new risks by indexing any shared files within an organization, potentially exposing sensitive data. Organizations must address accidental data sharing, proper identification of sensitive data, and classification of AI-generated outputs.

Key Risks:



Accidental Data Sharing: AI copilots might inadvertently index files containing sensitive information that are shared organization-wide, increasing the risk of unauthorized access.



Output Classification: AI copilots can generate outputs that might expose sensitive information if not properly classified.



Sensitive Data Identification: Organizations may not fully identify and classify sensitive data before deploying AI copilots, leading to potential data leaks.

Normalyze Solution:



Accidental Sharing Protection: Normalyze identifies and alerts security teams about files containing sensitive information that are being shared, allowing for access restrictions to prevent indexing by AI copilots.



Sensitive Data Identification: Normalyze scans and classifies datastores to ensure sensitive data is discovered before AI copilots are rolled out. Organizations can also manually vet curated sites to minimize unintended data leakage.



Sensitive Data Labeling: Normalyze applies Microsoft Information Protection (MIP) labels, enforcing protection policies like encryption and access controls. These labels ensure that Copilot follows these policies, preventing unauthorized access or sharing.



Output Classification: Normalyze classifies AI-generated outputs to ensure that no malicious prompts expose sensitive information.

Custom LLMs and AI Applications

Overview:

- » Custom LLMs and AI applications built on platforms like AWS Bedrock and Azure OpenAI present significant challenges in managing sensitive data, including the risks of data ingestion, post-leak management, and ensuring secure application communication.

Key Risks:



Sensitive Data Ingestion: There's a risk of sensitive data being inadvertently included in LLM training or inference processes.



Application Communication: It's crucial to identify which applications are communicating with models or RAGs that may have been compromised with sensitive data.



Post-Leak Management: Managing the fallout if sensitive data has already been leaked into the LLM is a significant challenge.

Normalize Solution:



Data Source Scanning: Normalize scans and identifies data sources feeding into custom LLMs, ensuring that sensitive data is not included in the models.



Cloud-Based AI Security: Normalize secures cloud-based AI deployments in AWS Bedrock and Azure OpenAI by detecting any sensitive data being fed into the foundational or custom models.



Risk Detection and Alerts: Automated detection of sensitive data in repositories, coupled with alerts, helps prevent data breaches.



Unsanctioned AI Tool Detection: Normalize can also detect unsanctioned AI tools being used across the business, providing further control over AI deployments.

Recommendations for mitigating sensitive data exposure

- **Monitor Training Data and RAG Data:** Regularly audit and review training data sources for sensitive information, using Normalize to scan and protect these sources.
- **Track Data Lineage:** Maintain detailed records of the origin and usage of training data, implementing tools to trace data lineage from datasets to deployed models and applications.
- **Monitor Model Inputs and Outputs:** Implement logging and monitoring for all model inputs and outputs to scan for sensitive data.
- **Implement Data Anonymization:** Anonymize sensitive data before using it for training.
- **Access Controls and Permissions:** Restrict access to sensitive data, ensuring that only authorized personnel have the necessary permissions.

»



Learn more at: normalize.ai

RAG (Retrieval-Augmented Generation) Models


Overview:

- » RAG models reference external data sources without directly ingesting data, creating unique challenges in data security, particularly around controlling access to sensitive external data.

Key Risks:

-  **External Data Reference:** Because it is easier to add data sources via RAG than it is to add training data and because those data sources can be large, RAG models may be more likely to access sensitive external data without proper vetting, leading to potential security risks.
-  **Data Protection:** Ensuring that only appropriate data is accessed by RAG models is crucial to minimizing exposure.

Normalize Solution:

-  **Data Protection in RAG Models:** Normalize identifies sensitive data in external data sources so security teams can control what data is made available to RAG models. At runtime Normalize APIs check RAG datastores, ensuring only appropriate data is accessed and preventing unauthorized exposure during AI queries.
-  **User Access Governance:** Enforcement of least privilege access on datastores reduces the risk of unauthorized internal users making datastores containing sensitive data available to RAG models.

Recommendations for mitigating sensitive data exposure

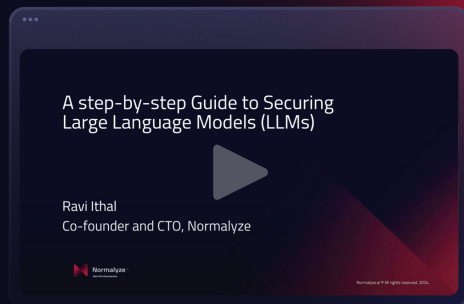
- **Monitor RAG Data:** Regularly audit and review external data sources for sensitive information, using Normalize to scan and protect these sources.
- **Track Data Lineage:** Maintain detailed records of the origin and usage of RAG data, implementing tools to trace data lineage from datasets to deployed models and applications.

Learn more about Normalize

Our Step-by-Step Guide to Securing LLMs provides you with the essential strategies and best practices to deliver on your strategic AI initiatives while safeguarding your data with continuous scanning, classification, labeling and controls.

[Watch now](#)

A Step-by-Step Guide to Securing LLMs



Copyright © 2024 Normalize, Inc. Normalize, the Normalize logo, Pioneers of Data Security Posture Management are properties of Normalize, Inc. One-Pass Scanner™, Data Risk Navigator™, Data Access Graphs™ and DataValuator™ are trademarks of Normalize. All rights reserved. All other trademarks and copyrights are the property of their respective owners. DSAI0924

»

Learn more at: normalize.ai