



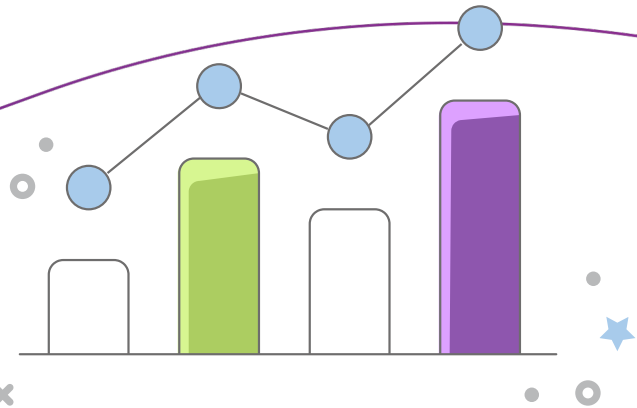
Survey Report 2025

Komprise IT Survey: AI, Data & Enterprise Risk

AI Puts a Shadow on Enterprise IT as Risks Get Real

Table of Contents

- Executive Summary 3
- Key Statistics 4
- Shadow AI & Risk..... 5
- Addressing Shadow AI Threats 6
- Preparing Unstructured Data for AI 8
- Top Tactics for AI Data Preparation 9
- Data Mobility for AI 10
- IT Infrastructure Priorities 11
- Top 5 Takeaways..... 13**
- About Komprise 14



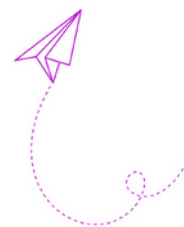
Executive Summary

In mid-2025, enterprises are starting to get real about AI. From prototype to production, there are a lot of steps and plenty of concerns. Enterprise IT organizations are embroiled in the gargantuan task of managing and preparing their vast stores of unstructured data for AI pipelines. Investing in new IT infrastructure to support AI is foundational: the storage, compute and networking technologies for high performance and security. Yet preparing and managing data for AI to support user workflows and governance is equally if not more paramount. Balancing these two priorities effectively can help organizations deliver safe, optimized AI services for employees and customers.



Komprise surveyed 200 IT directors and executives at U.S. enterprise organizations of 1000 employees and larger. The purpose of the survey was to discover how IT teams are preparing their unstructured data for AI and the challenges they are facing. The survey was conducted by a third party in April 2025.

The Komprise IT Survey: AI, Data & Enterprise Risk showed that nearly 80% of organizations have experienced negative data incidences with generative AI—with 13% resulting in financial, customer or reputational damage. **The most common bad outcomes** include false or inaccurate results from queries (46%) and leaking of sensitive data into AI (44%). Many surveys have indicated concern about these risks, but now those concerns are hitting the bottom line.



Other trends identified include an **overwhelming concern about “shadow AI”**: nearly half are “extremely worried” about the security and compliance impact of unauthorized and unsanctioned use of AI tools. Much of this concern today centers on GenAI tools, which are free and widely available on the Internet. IT leaders shared their tactics for dealing with shadow AI, from using data management and AI discovery tools to implementing policies and training.

The survey also discovered that despite an uncertain economic landscape amid tariff price increases, **enterprises are prioritizing developing the right IT infrastructure for AI.**



Key Statistics



Shadow AI Risks

- The vast majority (90%) are concerned about shadow AI from a privacy and security standpoint, with 46% reporting that they are “extremely worried.”
- Most (79%) of IT leaders report that their organization has experienced negative outcomes from sending corporate data to AI, including PII data leakage and inaccurate or false results.
- Most (75%) expect data management technologies to address risks from shadow AI, followed closely by AI discovery and monitoring tools (74%).



Preparing Unstructured Data for AI

- The greatest challenge in preparing unstructured data for AI is finding and moving the right data to locations for AI ingestion (54%) followed by a lack of visibility into data across data storage to identify risks (40%).
- The top tactic for preparing data for AI is classifying sensitive data and using workflow automation to prevent its improper use with AI (73%).
- Nearly all (96.5%) are classifying and tagging unstructured data for AI, with a mix of manual and automated methods for doing so.
- More than half (56%) say that IT is moving data to AI processes for users manually, or with free tools, with 40% saying that users are manually copying data to AI on their own.



IT Infrastructure Priorities

- Supporting AI initiatives is the top priority for IT infrastructure (68%), followed by 16% saying it is equally important as cost optimization, cybersecurity and core IT upgrades.
- Most IT leaders (45%) express a multi-faced strategy for investing in storage for AI, with equal priority to acquiring AI-ready storage, increasing capacity of existing storage and acquiring data management capabilities for AI.

I. Shadow AI and Risk

We used to talk about shadow IT, which became an issue when cloud computing went mainstream over 10 years ago. IT leaders weren't keen on department heads spinning up cloud instances and apps without proper oversight for security and costs. The notion of central IT architecture and governance standards flew out the window. Over time, IT leaders have worked to determine the proper balance between control and flexibility, as new technologies quickly developed to help business units move faster or smarter. Getting rid of shadow IT isn't possible – nor even desirable.

IT leaders are now talking about shadow AI: a more harrowing threat. Unsanctioned, unmanaged AI can (and has) introduced false information to markets and made errors that threaten a company's reputation and revenues. In the AI age, data privacy is under attack like no other time before.

With AI usage only set to grow as employees discover groundbreaking productivity and competitive benefits from using these tools, IT teams need to instill the right guardrails to protect corporate data and minimize the risks from AI.

Further, GenAI is pervasive and easy to use for any employee at any time. They only need a web browser to access free tools, which don't require fancy skills to use. In a few seconds, an employee can load files and content into a prompt before realizing (if ever) that they have just shared sensitive data, such as customer or financial information, with an open tool on the Internet.

The Komprise survey found that nearly half (46%) of IT leaders are “extremely worried” about the privacy and security risks from shadow AI, while 44% say they are “moderately worried.”

Not only are IT organizations worried, but they are seeing the fallout: 46% have experienced false or inaccurate results from AI which the organization had to reconcile and correct. A few famous incidents include [Amazon's](#)

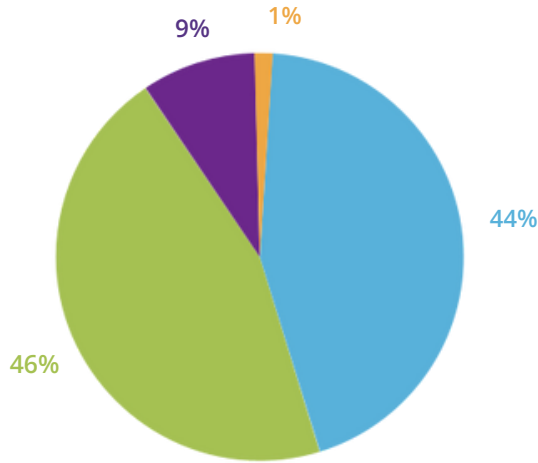
[AI hiring tool](#) which discriminated against female applicants, [McDonald's AI tool](#) that severely botched customer orders and [Chevrolet's chatbot](#) selling cars for a dollar.

Almost as many (44%) have discovered that internal proprietary or PII data was leaked into an AI tool.

Alarmingly, 13% of organizations have experienced financial, customer or reputational damage from these negative data incidents.



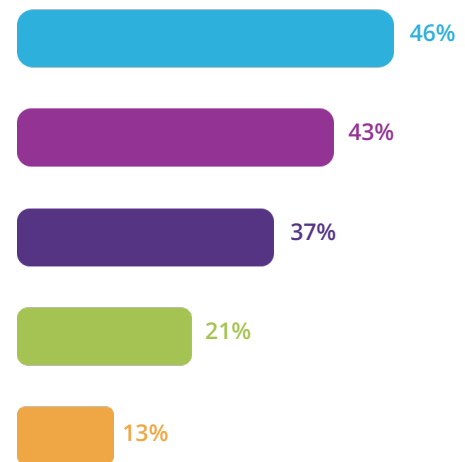
Q: Are you concerned about shadow AI in your business?



- Extremely. We are worried about data privacy and security risks from unsanctioned/unmonitored use of AI.
- Moderately: we know people are using GenAI regularly and trust they will adhere to policies.
- Not really; we fully support AI experimentation.
- Unsure, as we do not know how to fully understand what's happening and our risk.

Q: Have you had any negative data incidences with GenAI?

- Yes, we have had inaccurate and false information from AI that we had to correct and reconcile.
- Yes, we have had proprietary or PII data that was leaked into an AI tool and we found out about it later.
- Yes, we had an issue with using copyrighted data in a derivative work that an employee created.
- No.
- Yes we have had one of the above negative outcomes and it resulted in financial, customer or reputational damage to our company.



Addressing Shadow AI Threats

AI is a new type of opportunity and threat-- one that can propagate quickly across the enterprise. Nearly 60% of employees intentionally use AI at work with a third using it weekly or daily, according to an April 2025 global study of 32,000 workers, [sponsored by KPMG](#).

So how do you go about managing the risks without suppressing the positive, innovative benefits of AI in the workforce? Rather than outlawing AI, IT leaders are deploying multiple tools and processes to prevent negative outcomes. IT and security organizations may choose to restrict certain AI tools or use cases as well as which data sets are allowed for AI model training and inferencing.

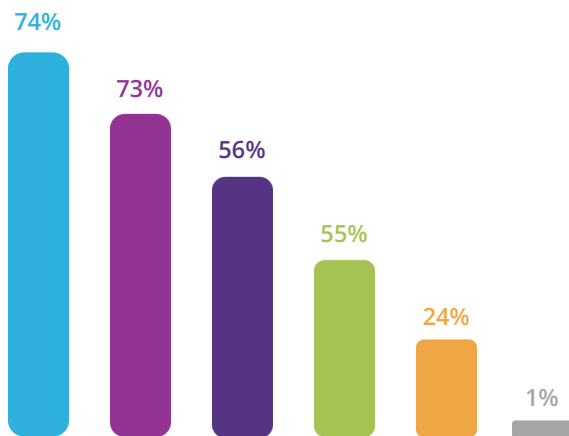
Most (74-75%) are expecting data management and/or AI discovery and monitoring tools to combat shadow AI risks. Meanwhile, a little over half (55-56%) are using traditional security tools such as access management and data loss prevention (DLP) tools and implementing new policies and training. Let's take a closer look:

- **Data management systems for AI:** An unstructured data management solution can index data across all storage so that IT users can create a query to discover sensitive data (such as IP or PII), classify it with tags, and move it to a secure location. This ensures that sensitive data is not accessible for AI ingestion. These solutions also allow IT to monitor AI data workflows for sensitive data and conduct investigations on data flows into AI projects which generate negative or false outcomes.
- **AI discovery & monitoring:** These solutions deliver complementary functionality to unstructured data management, by tracking which AI tools are being used in the organization and by whom. They can block usage, if the tool or data set is not in compliance with company rules and regulations. Some can detect which data sets were shared with AI or anonymize data sets containing personal data.

Most (74-75%) are expecting data management and/or AI discovery and monitoring tools to combat shadow AI risks.

AI risk is complicated, with countless, easy to use free tools now available to employees across the internet. Managing shadow AI while still allowing employees to use AI tools that enhance their work requires IT to orchestrate and support several technologies, policies and processes. The goal is to quickly detect and thwart employee actions that endanger sensitive data and corporate secrets.

Q: What are you doing about shadow AI, if anything?



- We are using data management to audit and govern data workflows and prevent sensitive data leakage.
- We are using AI discovery and monitoring tools to analyze improper user behavior.
- We are using DLP, access management and/or data management tools to find and manage sensitive data.
- We are focusing on internal policies, education and training rather than restrictive methods.
- We have a team evaluating solutions and haven't officially implemented controls or guidelines yet.
- No actions now.

II. Preparing Unstructured Data for AI

Unstructured data is the lifeblood of AI. For AI to be successful and relevant to an organization, it needs the right unstructured data at the right time. IT's role here is to deliver **unstructured data visibility, classification and segmentation** so that data scientists, analysts and others can find the data they need and send it to the optimal location for AI analysis.

Yet unstructured data is large, diverse and unwieldy. Most enterprises have dozens of petabytes of this data across diverse file types and sizes. It is onerous to search across it and analyze and move it efficiently to AI.

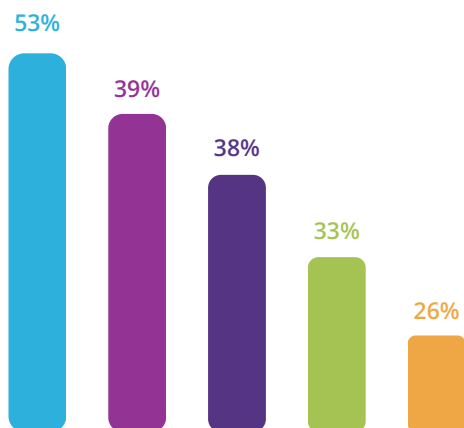
If they send too much data, AI processing gets too expensive.
If they send too little, the results will be suboptimal and even inaccurate. If employees send sensitive, restricted data to their AI projects, you're now looking at public access to company secrets, and potential compliance violations and lawsuits.

Top Challenges

The top challenge in preparing unstructured data for AI, indicated by 55%, is **quickly finding and moving the right unstructured data to locations where AI lives**. Secondary challenges include a lack of visibility across data stores to understand and identify risks, and segmenting and classifying data.

Given the overall immaturity of AI and the technologies that support it for an enterprise deployment, it is not surprising that **more than 30% lack internal agreement on the right strategy** for data management and governance.

Q: What is your greatest challenge in preparing unstructured data for AI?



- It's difficult to quickly find and move the right data to locations where AI lives based on departmental requests.
- We don't have full visibility into our data across storage and clouds so that we can see what we have and fully identify risks.
- We don't have easy ways to classify and segment our data for efficient user search.
- We have internal disagreement on how to approach data management and governance for AI.
- We don't have adequate tools to prevent and monitor sensitive and proprietary data leakage to AI.

Top Tactics for AI Data Preparation

Enterprise IT organizations seek easier, automated ways to prepare data for AI. Manual search and metadata enrichment/tagging across billions of files to classify and organize data is not viable.

IT's job number-one is to protect sensitive data, with **the majority (73%) looking to use workflow automation tools to classify sensitive data and prevent its improper use with AI.**

The second leading tactic for AI data preparation is automated scanning and classification, to bring needed structure to unstructured data. In some cases, unstructured data management technologies include these capabilities. Integrations with AI tools can deliver rapid data classification across large data sets by cracking open files, searching for keywords and creating a curated set of data.

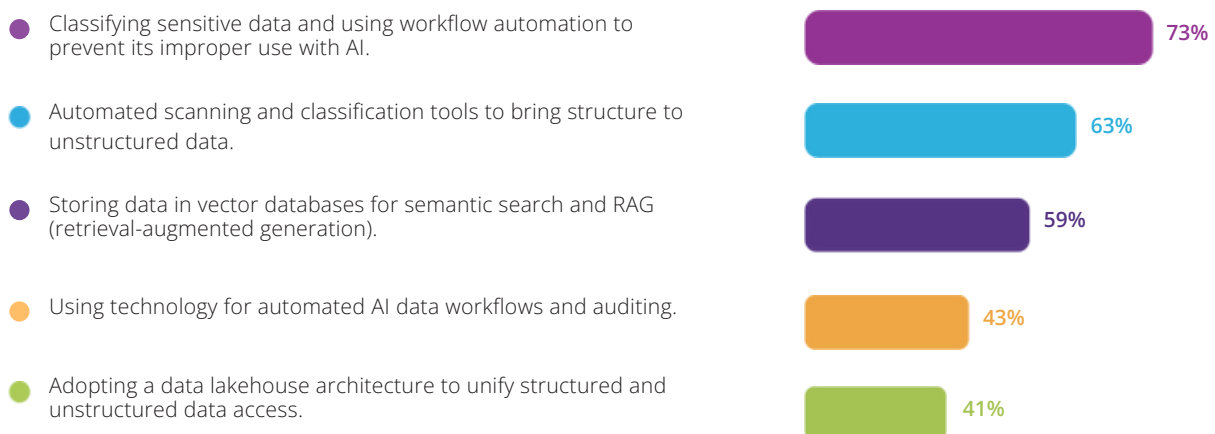
The top tactic for preparing data for AI is classifying sensitive data and using workflow automation to prevent its improper use with AI (73%).

Unstructured data management solutions with tagging capabilities can then take that AI output and automatically apply the appropriate tags to the files. This way, a researcher could use the unstructured data management solution to search on keywords and locate all the related files across distributed file systems without the assistance of IT.

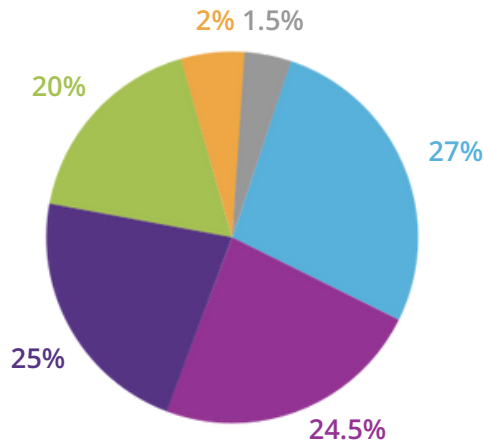
The survey showed equal interest in data management and AI approaches for data classification via metadata enrichment, with 25% using a data management system and 25% using an AI tool.

Third, nearly 60% of participants said they would store data in vector databases for semantic search and retrieval augmented generation (RAG). Vector databases allow organizations to convert file data in formats that capture meaning rather than just keywords, making this a useful strategy for search engines, chatbots, and recommendation systems.

Q: What tactics are you considering and employing for AI data preparation?



Q: Are you tagging and classifying your data for AI?



- We are doing some tagging and classifying of data, but it is mostly manual.
- We have an automated data management solution for data classification.
- We are using AI to enrich metadata for data classification.
- We are using a combination of AI tools and data management tools for data classification.
- We are not doing this today but looking for the right technology.
- We don't have a need for this right now.

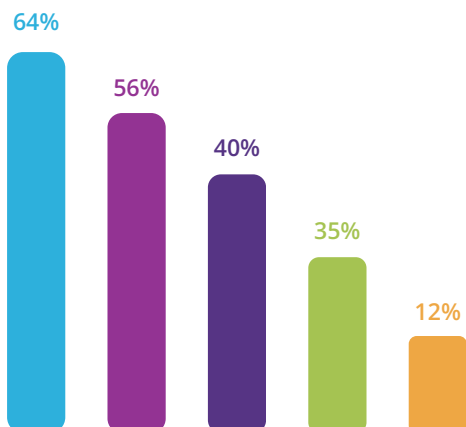
Data Mobility for AI

Once the data has been tagged, classified and segmented, organizations need efficient ways to move the data to AI pipelines. IT teams are using or considering one or more methods such as manually copying their data, free tools, and using automated data management tools for these tasks. **Yet an automated data management solution is the most-common preference today, indicated by 64%.**

Automated unstructured data workflow technologies can streamline the process of curating and moving the right data from storage to locations for use in AI, such as to a cloud data lakehouse, with proper governance. Data from storage to locations for use in AI, such as to a cloud data lakehouse, with proper governance. This technology can index data across hybrid storage, identify and confine sensitive data and execute policy-based automated tagging of data sets to help users search for the exact data they need.

An automated workflow could search for data tagged with "MRI," "glioma" and "female", copy the data to the cloud, and then repeat the process as new data enters the organization. Unstructured data workflow solutions include dashboards to monitor workflows in progress and investigate data sets used and by whom in a specific project, if needed.

Q: If your users want to move data to an AI process, how do they do it?



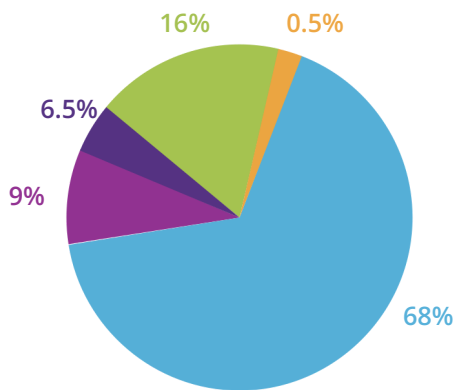
- We use a data management workflow solution that automates AI ingest and keeps an audit trail for data governance.
- They ask IT and we move it manually or via free tools.
- Each user manually copies data they need to ingest to AI.
- They use free tools to copy the data.
- We are still investigating the best options to accomplish this.

III. IT Infrastructure Priorities

The year 2025 has been marked by significant economic, political and business uncertainty. On-and-off U.S. tariff policies have disrupted supply chains and resulted in higher prices on many products-- including those designed for the data center. Despite this unpredictability in pricing and product sourcing, organizations are moving full steam ahead on their investments for AI.

This finding suggests that IT leaders and business executives are bullish on AI, given its potential to boost internal productivity, operational excellence and product development and enrich customer relationships. No one wants to be left out of the AI arms race. Building the infrastructure to support and protect data and applications is a critical undertaking.

Q: Where is infrastructure IT's priority on supporting AI initiatives in the business?



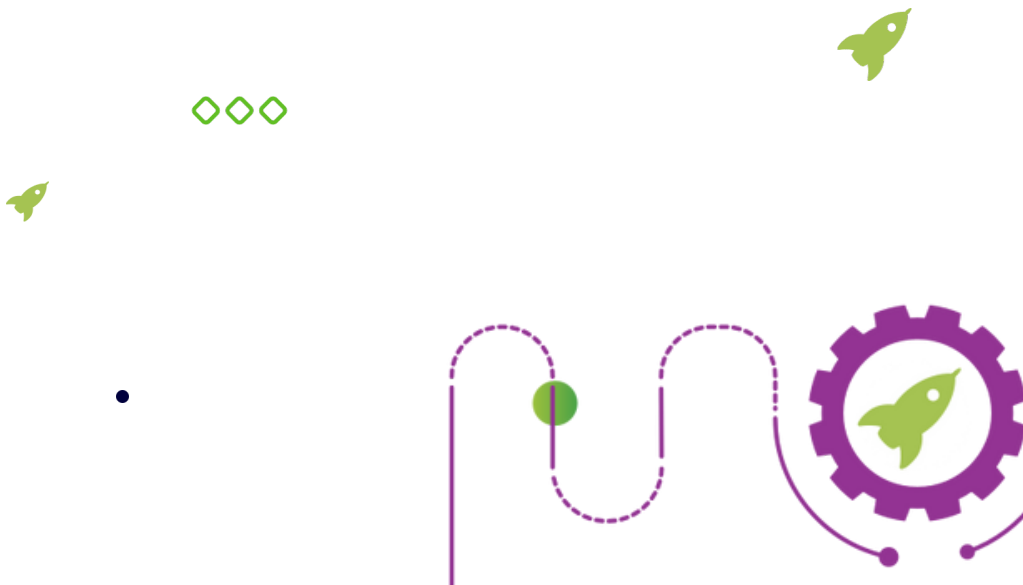
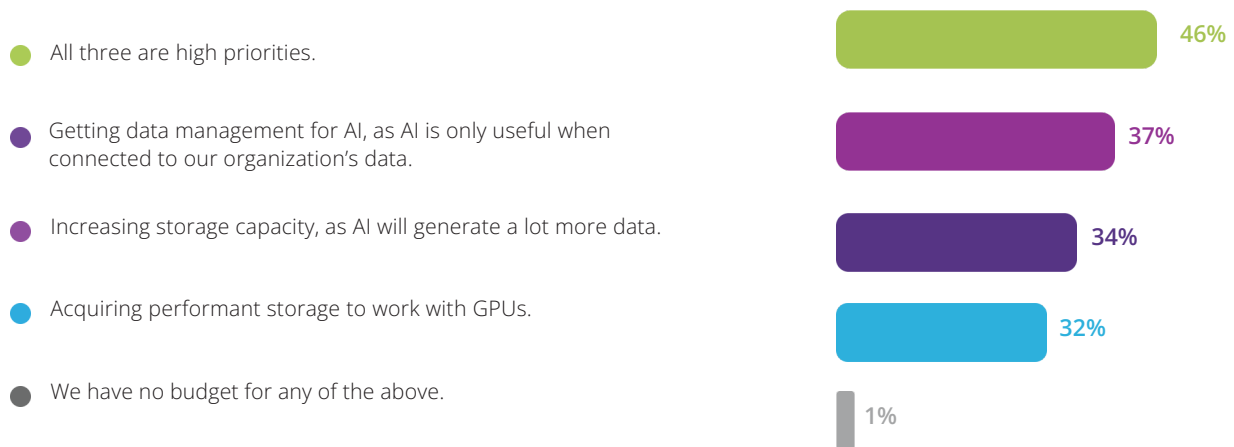
- A top priority.
- Right after cybersecurity.
- Cost optimization is the top priority, AI is second.
- It is equally important as security, cost management and core IT infrastructure upgrades.
- It is not a high priority now.

Exactly how enterprises are planning to create this infrastructure, however, is neither standard nor straightforward. The strategy will differ across organizations depending upon goals, budget, IT architecture and internal expertise.

The majority (46%) are placing near equal emphasis on:

- Acquiring high-performing storage that can work with GPUs;
- Adding capacity to existing storage to handle growth in data generated from AI, and;
- Using unstructured data management technologies to prepare and move corporate data to AI platforms.

Q: What is your top priority storage investment for AI?



Top 5 Takeaways



Generative AI is Hitting the Bottom Line

AI mania has turned into the daily integration of GenAI in the workforce. The dangers we've been hearing about are now becoming real. Most organizations have experienced negative data incidences with generative AI—with some resulting in financial, customer or reputational damage. The most common bad outcomes include false or inaccurate results from queries and leaking of sensitive data into AI. Unless IT can get this under control, executives will shut down initiatives that can deliver measurable competitive advantage.



Shadow AI Risks Requires New Tools

Most IT leaders are worried—half of them extremely so—about the privacy and security risks from shadow AI. The risk occurs when sensitive data is leaked to commercial AI tools, exposing PII and company secrets. To prevent adverse outcomes of GenAI, IT will invest in new technologies: data management to avoid sensitive data leakage and enable proper data classification and AI discovery tools which track AI apps and usage inside a company.



Data Classification and Segmentation is a Top AI Strategy

AI needs unstructured data, yet this data must be precisely curated for accuracy, cost and security requirements. Data classification is challenging for unstructured data because filesystems retain only basic system metadata. IT leaders need adequate tools for users to find the right unstructured data along with visibility across data stores to identify risks and classify data. IT leaders will look for automation to enrich metadata with relevant keywords for user search and to exclude sensitive data from AI workflows.



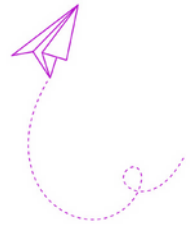
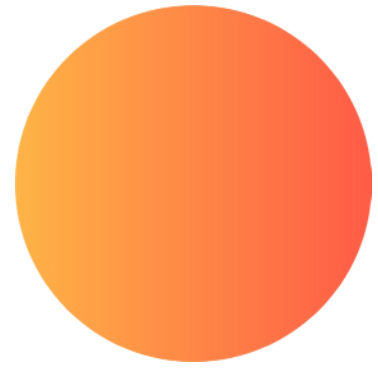
Automation for AI Data Pipelines Takes Off

Unstructured data is large and unwieldy in enterprises. An automated data management solution is the leading preference for moving data into AI pipelines. These solutions streamline the process of curating and moving the right data from storage to locations for use in AI, such as to a cloud data lakehouse. This technology can index data across hybrid storage, execute policy-based automated tagging of data sets to help users search for the exact data they need and address governance needs with auditing.



IT Teams Double Down on AI Infrastructure

IT organizations are moving full steam ahead on infrastructure investments for AI—ahead of cybersecurity and cost optimization. Even amid tariff-related price increases and a push to conserve spend, IT leaders know they must support AI projects with speed, performance and security in mind. This will entail an equal focus on procuring adequate AI-ready data storage and a robust data management solution to prepare data for ingestion with proper guardrails for security, privacy and compliance.



About Komprise

Komprise powers the connection between unstructured data management and AI. Komprise Intelligent Data Management delivers a single platform to easily analyze, migrate, transparently tier and manage the lifecycle of petabytes of file and object data across hybrid environments. With Komprise, enterprise IT gains full visibility across silos to optimize storage, backup, ransomware and cloud costs. Komprise Smart Data Workflows and the Komprise Global File Index unlock unstructured data insights and access for AI.

Learn more at www.komprise.com



Komprise, Inc.
1901 S. Bascom Ave. Suite 500
Campbell, CA 95008
United States

For more information:
Call: 1-888-995-0290
Email: info@komprise.com
Visit: komprise.com

For media requests email
marketing@komprise.com