

WHITE PAPER

'Fit for AI'

Transforming Unstructured Data Chaos into AI Order

2024

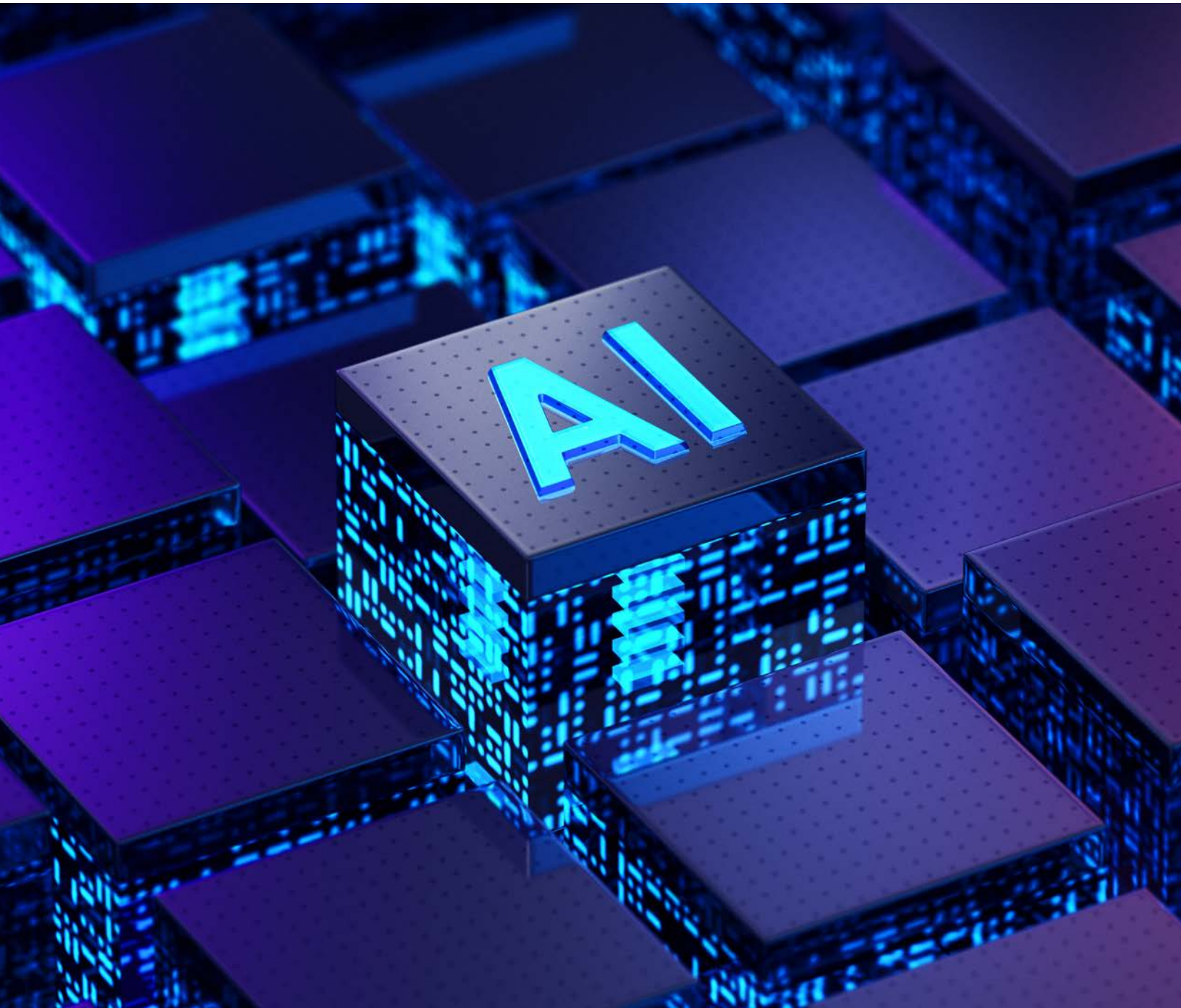




Table of Contents

3	Introduction
5	Introducing The Nasuni ‘Fit for AI’ Value Pyramid
6	Assess Unstructured Data Silos
7	Consolidate Data Silos
8	Assess Security
9	Curate Data
10	Unifying File Access Through Edge Devices for Streamlined AI Workflows
11	Understand New and Changing Regulations
13	AI Next Steps
13	Questions to Ask When Evaluating AI Solutions
16	Next Steps in Getting ‘Fit for AI’



In the ever-evolving landscape of enterprise technology, artificial intelligence (AI) stands out as a force, poised to reshape how organizations operate. As AI and machine learning continue to advance, the potential for transformative innovations has become evident, from highly accurate predictive analytics to human-like conversational interfaces. For IT leaders navigating this paradigm shift, it is crucial to recognize that the foundation for a successful AI strategy lies in a comprehensive data strategy.

Why Read This?

This piece is a vital read for IT leaders aiming to understand the pivotal role a robust data strategy plays in unleashing the full potential of AI. Whether you approach this as a Nasuni-sponsored piece or an educational guide, the central message remains clear: it is critical to prioritize your data strategy in the era of AI.

Key Questions Answered:

- Why is a data strategy fundamental for implementing a successful AI strategy?
- What challenges do enterprises face with unstructured data, and how can AI transform this landscape?
- How can IT leaders consolidate scattered storage silos to create a single-source-of-truth for AI systems?
- What benefits can organizations unlock by making unstructured data accessible to AI systems?
- Why is Nasuni a strategic choice for building the foundation of an AI-driven data strategy?

Essential to have a robust data strategy for successful AI strategy.

Enterprises are struggling as their data today is not ready.

Introduction

Artificial intelligence (AI) represents an unprecedented disruptive and transformative technological shift which will have a large impact on the enterprise organizations. There is no doubt that new AI advancements are the catalyst for enterprises to look at their unstructured data in a new light.

The Time is Now to Prioritize Data Strategy

Despite the undeniable potential of AI, enterprises today find themselves unprepared to fully leverage its capabilities due to a lack of prioritized data strategy. The urgency to embrace data strategy in tandem with AI implementation stems from the current landscape, where data remains largely underutilized and unstructured. Without a solid foundation in data management and strategy, customers risk missing out on the transformative benefits AI promises to deliver.

The perception of data has shifted from a burdensome cost to an asset, prompting chief data officers to become asset managers. Unstructured data often becomes “dark” or unused, and over time becomes more and more difficult to find. It may be combined with other types of data in a data lake, or spread out across many locations and departments.

Unstructured data, often overlooked and stored in scattered silos, holds untapped potential for businesses. Organizations today have data assets in different formats that have hidden value still remains highly underutilized.

Diverse file data types across industries containing valuable insights

Industries	File types
Oil and Gas	Seismic data
Media	Post-production assets, graphics, videos, and images
Government	Forensic, surveillance
Manufacturing	Quality data, inventory and logistics, design files
Construction	AutoCAD drawings, floor plans and engineering designs
Finance	Transaction records, market data, portfolio holdings
Legal	Case files, legal briefings, compliance documents, contracts

Consolidating dispersed unstructured data repositories into a single, accessible source is crucial for laying the groundwork for an effective AI strategy. This consolidation not only reduces storage costs and duplication but also addresses compliance and security concerns, paving the way for actionable insights and improved data management. By establishing a unified foundation through consolidation, organizations can unlock the transformative power of AI and machine learning, turning data from a liability into an asset driving analytics-based decision-making.

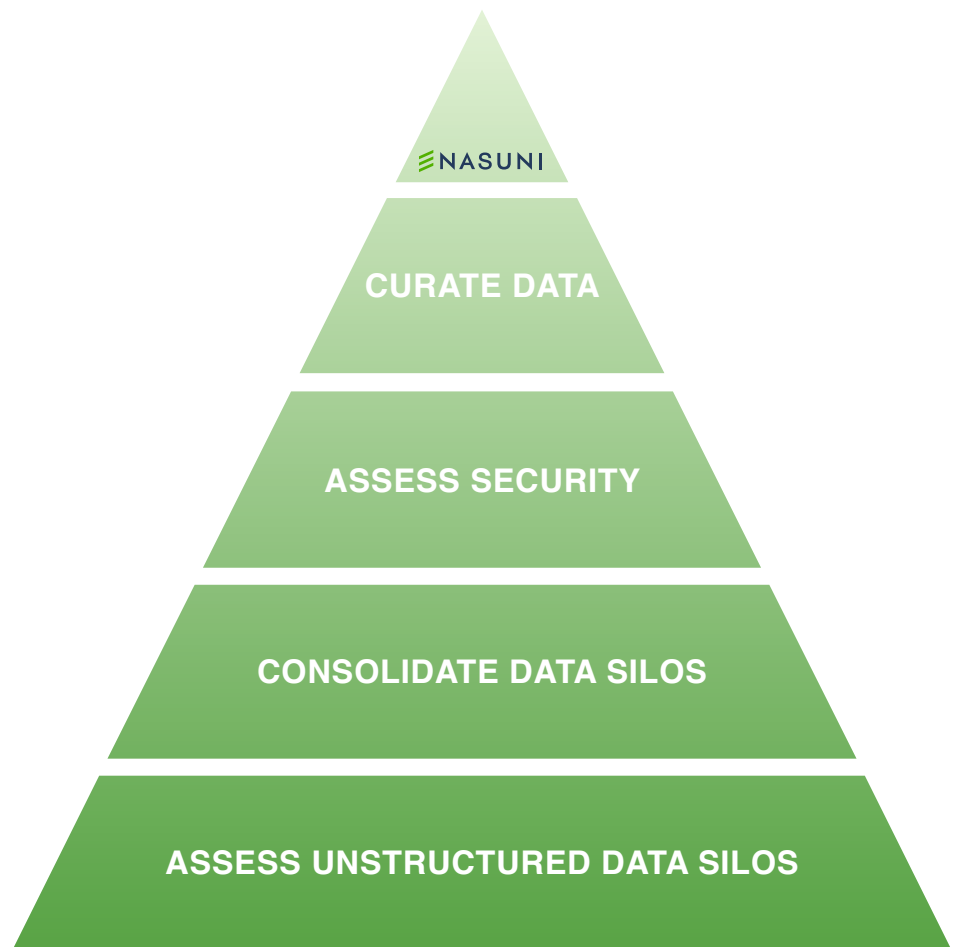
Introducing The Nasuni 'Fit for AI' Value Pyramid

Unified data lights the path to AI. Scattered, dark silos starve insights. Following the 'Fit for AI' framework can help transform file data fragmentation into file data consolidation, powering enterprise intelligence to:

- **Assess** file data silos to map alignment, costs, value, and risks.
- **Rationalize** file storage silos into one single-source-of-truth.
- **Secure** and protect the consolidated data.
- **Curate** and assess the data prior to using it with AI services.

Following the framework and working with Nasuni can decrease the opacity of fragmented data to data transparency, and data disorder into curated fuel-driving analytics and AI at scale.

Nasuni 'Fit for AI' Framework



Assess Unstructured Data Silos

Assessing where corporate file assets reside today inside of an organization represents a crucial first step on the journey to a single-source-of-truth repository that can be used in AI pipelines.

As part of the initial assessment of the storage silos, Nasuni will create a business value assessment to calculate the business value of consolidating file silos laid out in 4 categories:

- 1 IT Capital Costs:** Compares the cost of the existing file silos to the costs of file data unification with Nasuni.
- 2 IT Operational Costs:** Compares the IT time and resources required to operate the file-based silos to the operational costs once the data is unified with Nasuni.
- 3 Business Productivity and Revenue:** Compares the cost of workforce inefficiencies and revenue impact by not unifying the data to the Nasuni single-source-of-truth state.
- 4 Business Continuity:** Compares the risk to consolidated business continuity with existing file infrastructure silos.

Nasuni professional services can work with you to analyse your current file data storage environment and provide an evaluation of space usage and plan the migration accordingly to ensure a smooth and seamless process.

Built-in reporting with customizable data exporting delivers shareable documentation across teams on the volume and composition of data assets ready for migration and consolidation initiatives.





Consolidate Data Silos

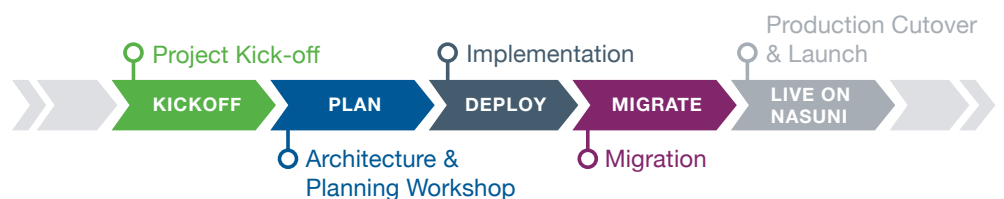
Part of your journey to consolidating disparate file data stores includes the migration of the file data to Nasuni. Nasuni's data migration services are the pathway to an implementation that eliminates risk and accelerates value. Based on hundreds of successful migrations and deployments, which often involve consolidation of existing file silos, Nasuni professional services blends the right application of technologies, consulting services, and project scope to ensure fast and painless migrations built on a solid foundation.

Following our methodology, we begin the process of transferring our expertise to our customers by educating our customers on architectural best practices with the goal of translating customer operational requirements into an operational deployment.

During the process our team makes recommendations for improving the migration consolidating by leveraging our broad experience with more than 800 customers deployed and thousands of successful migration projects.

Migration Workflow

Nasuni is here to help your organization every step of the way.



On average, our customers tell us they spend **60% less time** managing their infrastructure once they have moved their distributed file data to Nasuni.

Nasuni's data migrations can consolidate data silos in the tens of terabytes to global, multinational companies with petabytes of data, hundreds of locations, and thousands of users. In each case, we work with our customer to determine the best path to consolidation that results in an architecture that will provide the single-source-of-truth for AI services, and will support their business processes into the future.

This serves up immediate operational benefits. On average, our customers tell us they spend 60% less time managing their infrastructure once they have moved their distributed file data to Nasuni. This is brought about by consolidating file silos, backup software, sync tools, disaster redundancy, etc. into one, simplified Nasuni Management Console. With this step completed, your infrastructure teams should now have additional time and resources to apply to experimenting with and familiarizing themselves with AI processes.

Assess Security

Threats such as ransomware pose danger not only to day-to-day business operations, but to the carefully constructed file data assets powering AI and analytics innovations. As malicious attacks grow more sophisticated, organizations must prioritize multi-layered protection around information stores that are being used to fuel productivity efficiencies, cost reductions, and competitive advantages.

Even with the most advanced machine learning models, bad data guarantees bad insights. Few investments sting as acutely as having sensitive datasets encrypted and held hostage due to security gaps stopping daily business in its tracks. Reconstructing lost or compromised institutional knowledge can be expensive and time critical.

Up to 80% of a company's data is in file shares. Nasuni provides built-in protection capabilities to ensure file resilience.



Upholding continuity against malicious threats represents a non-negotiable starting point for reliable AI and business operations. While consolidation to access controlled Nasuni file data lakes limits attack surfaces, additional safeguards remain essential.

The Nasuni File Data Platform offers a full complement of ransomware services that helps protect and recover file data from ransomware attacks. Detection begins at the network edge where users are located, notifying IT teams of suspicious file patterns, malicious file extensions, and ransom notes across the entire organization.

Mitigation policies reduce business impact before an attack can spread. Files that were impacted can quickly be recovered, bringing affected users back online fast and comprehensive audit logs and incident reports keep detailed records of threat events.

Curate Data

As established, an effective AI strategy relies on consolidated, well-governed data foundations. Attempting to analyze fragmented, low-quality data spread across siloed systems can result in costly integrations, faulty insights, and poor return on investment.

Once corporate file data has been assessed, consolidated, and secured with Nasuni, the next step is data curation.

By leveraging specialized data intelligence tools, organizations can thoroughly analyze the composition and utilization of unstructured file data trapped in nested network folders. Interactive usage dashboards can quantify storage consumption down to the department and file type while identifying infrequently accessed subsets earmarked for archival.

Dashboard displaying activities across the edge appliances, volumes and users. Activity dashboards also draw insights from embedded metadata such as file owners, update timestamps, and other existing taxonomies to guide policy decisions.



This is why Nasuni includes a data intelligence solution as an integral part of its file data platform. It delivers an indispensable magnifying lens into the consolidated file data set allowing data stewards to curate file data stored in Nasuni based on total awareness rather than intuition.

Additionally, as Nasuni integrates with a company's existing identity management systems this makes it straightforward to highlight group permissions and access control lists to inform consolidated security protocols.

Modern search technologies such as Azure AI Search also facilitate smarter data curation with powerful indexing capabilities. Solutions such as Azure AI Search enable organizations to efficiently structure and prepare content for actionable insights. Its integrated AI enrichment automatically extracts text, labels images, and generates captions to unlock insights from complex unstructured data. The intelligence to tag metadata, translate text, and process sentiments alongside robust search features allow companies to filter, refine, and validate the curated dataset to guarantee quality and usability downstream. These serverless search services can handle time-consuming data processing at scale, contributing to a curation-ready dataset ready to be used with AI.

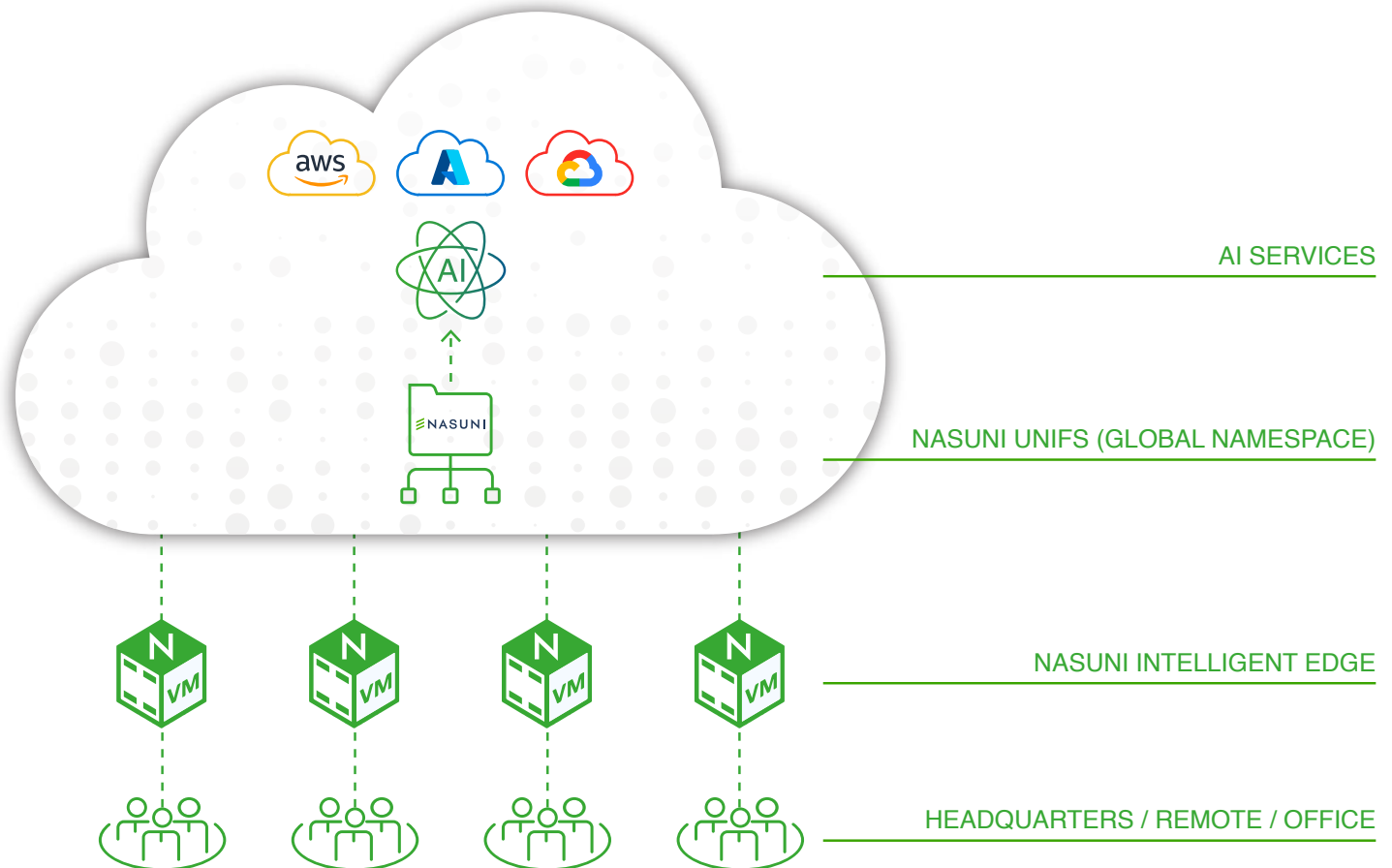
Unifying File Access Through Edge Devices for Streamlined AI Workflows

Carrying out AI workflows at the network edge aligns with key emerging best practices. As AI models grow more complex with deeper learning capabilities, shuffling large datasets from local endpoints to a centralized cloud location can introduce latency, bandwidth congestion, and security risks.

Nasuni solves this using edge appliances that eliminate cloud latency for files that are required to be local. Nasuni continuously captures changes across all edge locations and synchronizes the changes to a global file system.

This allows applications to take advantage of cached local data, and is especially helpful when there is lots of “local data gravity”. Nasuni’s intelligent edge solutions are multi-protocol, allowing for access and file sharing across many different applications.

Organizations can thereby build smarter AI systems leveraging consolidated live data and gravity-centric cached data. This empowers more dynamic machine learning based on the most contextually relevant assets available at the edge rather than limited stale copies. Unified data access and processing close to where data lives and gets utilized amplifies the capabilities of AI innovation.



Understand New and Changing Regulations

The rapid evolution of AI technologies can be difficult enough to track, but now organizations must make sense of new governmental regulations and restrictions. Any large organization with legal and security departments is going to have to align with the nascent and rapidly evolving governmental directives. So, it is essential that enterprise technology leaders track these initiatives and their impact on their operations.

The recent [U.S. Executive Order](#), the [U.K. Bletchley Declaration](#), and the [E.U. AI Act](#) differ in various ways, but all attempt to address the risks associated with this evolving technology.

All three of the frameworks identify bias, security vulnerabilities, and unintended consequences as issues that necessitate oversight. They prioritize risk mitigation, but suggest different approaches.

Unsurprisingly, all three emphasize coordination between nations and the private sector. Which players in the private sector drive this conversation is critical.

All three frameworks claim that they intend to balance safety and ethics while supporting innovation. All three generally avoid setting hard limits on AI development itself.

While the focus seems to be on high-risk applications, all three recognize that lower-risk AI systems also warrant transparency.



Care should be taken that **AI does not lead** to companies' breaching existing data regulations.

The E.U. AI Act is currently the most detailed, with a focus on implementing a binding set of rules; the others consist of broader principles at this stage. This is unsurprising given how the EU has implemented strict privacy initiatives, such as the GDPR, in recent years.

The U.S. Executive Order is clearly more focused on security; the EU on fundamental rights; the U.K. declaration at this stage is less obvious.

The U.K. declaration and EU Act both advocate the creation of central regulators, unlike the U.S. Executive Order.

Ultimately these frameworks are at an early stage (with the EU AI Act the first to make it into legislation) as they seek to catch up with the speed of AI developments, but they will continue to evolve. Moving forward, as technology leaders implement AI solutions in their enterprise, they should continue tracking and developing regulation to ensure their organization is compliant across multiple nations and regions. This is particularly important if the enterprise is going to use AI within products they create and sell, but also critical with regards to classified data that is made available to AI solutions and should not leave the organization and therefore breach existing regulatory frameworks such as HIPAA, GDPR, CCPA, etc..



As technology leaders implement AI solutions in their enterprise, they should **continue tracking and developing regulations** to ensure their organization is compliant across multiple nations and regions.

AI Next Steps

Questions to Ask When Evaluating AI Solutions

AI presents a set of complex models, applications and types of AI that require detailed evaluations. In parallel to the steps in the Nasuni 'Fit for AI' value pyramid, companies should strive to clearly identify high value use cases that AI could address within their organization and then consider the following five broad questions as part of their AI evaluations.

Natural Language Processing

Data classification, translation, and data extraction

Machine Learning

Algorithms that can learn from data to make predictions/decisions without explicit programming

Vision AI

Image/video recognition, and machine vision

Generative AI

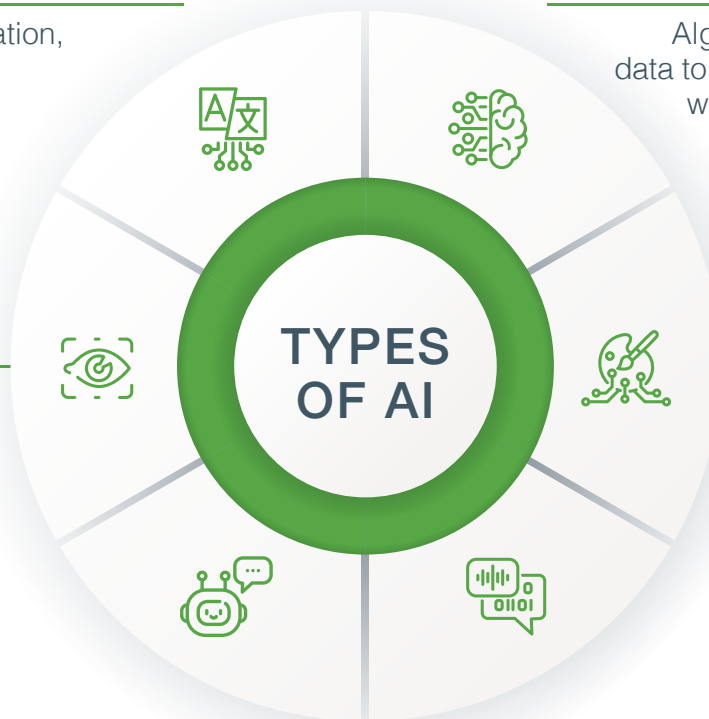
Creates realistic artifacts such as images, text, audio and video

Conversational AI

Chatbot prompt based interaction

Speech Based AI

Text to speech, speech to text



1 Is there a clear understanding of the use cases and the data?

No amount of advanced algorithms or model architecture can compensate for inadequate training data. Getting the right data at scale to build, evaluate and maintain AI solutions should be the absolute priority – before any modeling occurs. That’s why seasoned AI practitioners always stress that it’s all about data.

Deeply understanding the use case is also core to understanding the requirements of a large language model. Is it chatbot interaction? Is it generative? Is multimodal required?

Some of the common use cases tend to be AI Search, Generative AI, and model fine-tuning. Understanding the use case and mapping out the requirements, participants, permissions, governance, and lifecycle early on is key to success.

Common AI use cases with unstructured file data.



Search / AI Search

Searching data sets for keywords and/or prompt-based interactions



Generative AI

Proposal Generations, RFP Generations, etc



Model Fine Tuning

Augmenting foundational models with corporate data

2 What level of implementation support will be required?

It is critical to consider what internal skills will be needed to implement and leverage the solution. The AI vendor should be able to address this question in detail. It will be helpful to find out what sort of qualified technical support contacts they will provide, along with the expected availability and responsiveness of these teams.

There are many other technical, regulatory, economic, and organizational factors to consider when looking at a LLM-based tool or AI solution – these five questions are a starting point. Request a detailed set of possible cost projections and demand extensive projections that include potential unforeseen costs such as operational risk insurance or cyber insurance. Consider partnering with an existing, trusted solutions provider or AI consultant to help navigate the evaluation.

Choosing an AI solution is a complex task for any enterprise, and IT leaders should not be discouraged by the complexity. These are interesting and occasionally confusing technological times, but LLMs and AI solutions are going to transform how we extract value from data, and perhaps how we do business in general, IT leaders should take the initiative to train themselves and their teams in identifying solutions that will optimize their organization’s benefits while safeguarding critical assets.



Finding out whether the solution has been evaluated for bias and understanding how it has been adapted or modified to mitigate the effects of biased data is **crucial**.

3 How does the AI solution manage privacy and security?

A firm data foundation is key to any AI strategy and is a critical enterprise asset. Any tools deployed within organizations need to secure and protect the data deployed to it. IT leaders should start by evaluating the public, private, and open-source LLMs available and identify the variation that best suits their company's privacy and security needs. Then find out how the model leverages customer data, personal information, and other sensitive data.

For a commercial public model, it is essential to verify the protection of your company's data against inadvertent exposure. Evaluate the contract or terms & conditions agreements to guarantee that organization's data remains secure and does not leak into the foundational model behind the tool.

As a data owner or steward, the focus should be on ensuring that users utilizing these tools can only do so with data for which they have explicit permission. Preserving role-based access control is imperative to safeguard this critical asset.

Next, make sure that the security department is involved in establishing policies for how data can be used in these processes. If the solution does not meet your organization's data privacy and security standards, continue looking for alternative options.

4 Has the AI solution been evaluated for bias?

If the LLM or AI tool does satisfy data security requirements, the next step is to dig into the question of bias. Finding out whether the solution has been evaluated for bias and understanding how it has been adapted or modified to mitigate the effects of biased data is crucial. This cannot be based on trust, due to the potential negative impact of a biased AI tool on your brand and business. The AI vendor should be able to prove that the curation of the training data and/or any related model adaptations reliably mitigate bias.

5 Is the AI solution equipped to address regulatory compliance?

The AI vendor needs to be ready to adapt to new and changing AI regulations. This is not easy, as the regulations vary between regions, countries and even states. How will the AI vendor comply? Who validates compliance? Does the vendor have standards backing? These and other compliance-related questions are especially important for multi-national corporations that need to adhere to varying regulations around the world. Inquire about how the vendor stays on top of all these concerns and how their approach and any future changes will be communicated to the company's legal and security departments. See the section on 'Understanding New and Changing Regulations' for more information.



6 Is the AI solution designed to improve over time?

LLMs are trained on a particular corpus of data, but businesses are not static. It is important not to rely on an AI tool that can't learn from feedback and improve over time as data changes. Ask the solution provider how their model adapts over time and how they ensure that the data which feeds their model remains current. What is their strategy for keeping the solution tuned?

Also, be aware of how the solution handles so-called 'dirty data', both on the part of the enterprise and the model. Dirty data pollutes or degrades the performance of the model and can produce inaccurate results. If the AI vendor suggests that concerns around dirty data can be circumvented through prompt engineering, be wary. This indicates that the bad data still resides inside the model, and using the wrong prompt may generate a potentially toxic result. Such a solution is not an enterprise-ready tool.

Next Steps in Getting 'Fit for AI'

Nasuni's 'Fit for AI' framework is a process for preparing your unstructured file data for use in AI use cases.

As data grows and AI continues to advance, the insights extractable from data will become far more impactful to the success of a business. Data centricity breaks down data silos. Whether between departments, teams, or stored in legacy systems, siloed data severely limits strategic value. Data-centric models that promote a global namespace, such as Nasuni, promote capturing, connecting, and using data enterprise-wide, whether that data is on-cloud, or at the edge.

Going forward it's important to review the framework and understand how your unstructured data strategy is impacting your ability to take advantage of AI scenarios. Nasuni can help. We know it's not possible to predict every AI use case, so we will work with you to develop a strategy that future-proofs your unstructured data storage, security, and edge access and offers your organization maximum flexibility to take advantage of the power of AI. Call us today.



Contact us to learn more about how your organization can power your AI solutions with Nasuni's File Data Platform. [Learn more @ nasuni.com](https://www.nasuni.com)



ABOUT NASUNI CORPORATION

Nasuni is the leading hybrid cloud storage solution that powers business growth with effortless scalability, built-in security, and fast edge performance using a unique cloud-native architecture. The Nasuni File Data Platform delivers operational excellence by consolidating NAS and backup, eliminating data silos, and making management easy and flexible without changes to apps or workflows. Its built-in security offers proactive defense and rapid recovery, lowering organization's risk from the detrimental effects of ransomware attacks and other disasters. Synchronized access to file data everywhere ensures user productivity by supporting remote and hybrid work. For more information, visit www.nasuni.com.