Deep Analysis

Custom Research

AI: Moving from Readiness to Responsible Implementation

By: Alan Pelz-Sharpe & Matt Mullen

opentext™

This report was commissioned by OpenText and prepared by Deep Analysis.

Executive Summary

The era of general AI "readiness" planning is over; it's time for focused action to implement AI technology responsibly in organizations. The key to operational AI success lies not in using all your data, but in strategically activating the right subsets of your unstructured data (e.g., documents, emails, manuals) for specific, high-value use cases.

To do this, you must move beyond checklists and take three practical steps:

1. Target precisely: Identify and select only the unstructured data collections essential to your priority AI initiatives.

2. Understand the data: Assess the suitability of individual items and entire collections of unstructured data for AI use.

3. Build in protection from the start: Design for veracity, privacy, and security to ensure the AI's outputs are trustworthy, compliant, and secure.

The conclusion is a clear call to action: organizations must shift from planning to responsibly activating their data. This white paper outlines a 90-day plan and roadmap to convene stakeholders, design a secure data pipeline, and establish governing policies, turning strategic data into a competitive AI advantage.

It's Time to Move from Readiness to Action

In the four years since generative AI first emerged from technology laboratories and moved into our organizations, organizations have focused on preparing themselves for using AI technology. Initially, this concerned large language models (LLMs) and their proxies, like ChatGPT. Then attention shifted to productivity uses of the technology, in the form of assistive tools like Copilot. Recently, the focus has been on embedding the technology into operational workflows as agents.

This shift from chat to agentic has had common underlying themes. In 2024, Deep Analysis partnered with AIIM to produce an infographic with questions designed to help organizations evaluate progress in their preparation for using AI. Two years on, those questions remain prescient for many organizations still cautiously evaluating their readiness to implement AI.

Those questions run the gamut of the operational planning areas that organizations have grappled with over the last four years. They cover line-of-business use of the technology and determining how well it might work. Where might the people and skills to manage it come from? Finally, what are the implications for managing our data? Any AI use will depend on that data. This white paper focuses on that question.

Broad checklists function well at a high level, but they're not designed to help organizations take actions beyond determining readiness. In order to move towards actively using AI, we need to move from "readiness to *plan*" to "readiness to *act*."

In this white paper, we discuss active steps for moving your AI preparations from whiteboard ideation to project planning and using AI in your organization.

We focus on providing practical steps for using the *unstructured data* your organization already manages to power your use of AI, and pragmatic ways to ensure you're not overwhelmed by the scale of the tasks.

Deep Analysis-AIIM Questions to Evaluate Progress in Preparing for AI Technologies

1. Have you identified the processes you think AI will improve?
2. Can you list the tasks that AI will power within those processes and map them to possible AI solutions?
3. Have you identified and quality controlled the existing pools of knowledge you'll feed into the AI?
4. Do you know which people and skills you'll need to build and operate your AI solution?
5. Have decision-makers at your organization agreed on which measurements to use to judge the AI's success?

Download the infographic here.

Not all unstructured data is the same: find the bits you need

Unstructured data fills organizations: it's all the information they receive, produce, and organize that doesn't fit neatly in rows and columns. Emails, product manuals, documentation, schematics, marketing collateral, and operating procedures: indeed, unstructured data in files spread far and wide makes up the bulk of your organization's knowledge base.

This can feel overwhelming. Organizations with active information management practices utilizing suitable technology to manage this data – such as effective enterprise content management systems – occupy a fortunate position. Those without can still take steps to improve their situation, as examined later in this paper.

In both cases, however, what's important is that you won't attempt to use all of the unstructured data you identify. Instead, you will focus on activating *only* those sets of unstructured data that help to power the specific, appropriate use cases identified within your organization.

From there, you will examine this unstructured data to understand its characteristics, so that you're using it in the right places at the right time, ensuring that it's always up to date and relevant to the context AI requires.

Key Takeaway: *You don't need to deal with all your data (unstructured or otherwise), just those subsets that are relevant to the AI task at hand.*

What to understand about items of unstructured data

First, "unstructured" does not mean "uniform." These datasets all lack formal structure, but they can differ profoundly in their format, content, and management needs.

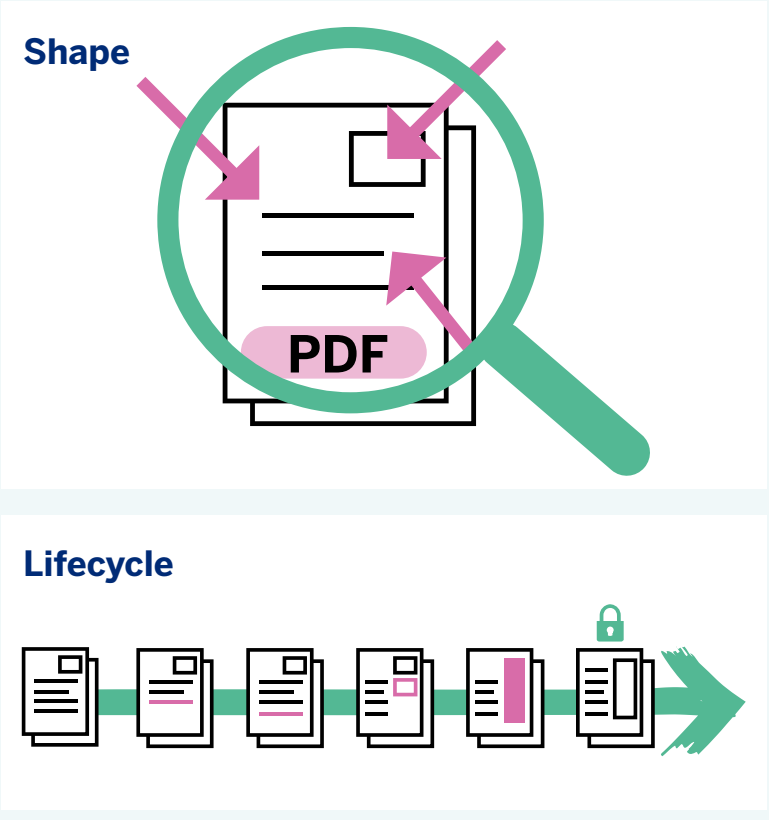
Unstructured data has both a shape and a lifecycle:

- Shape: the unstructured data's content, length, and format.
- Lifecycle: how the unstructured data is created, used, and stored.

For example, a product knowledge base article has a shape (i.e., a handful of headings, a typical length of 500 words, and a storage format such as PDF) and a lifecycle (created internally by a product manager, reviewed in

Figure 1

Unstructured Data Shape vs. Lifecycle



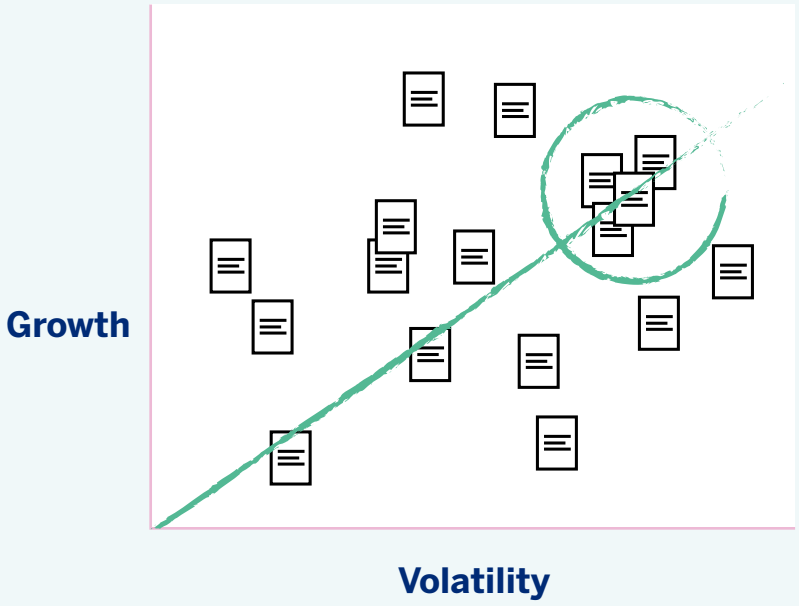
parallel with a product, eventually archived as the product goes out of circulation). Compare that to an operating manual's shape (hundreds of headings, 10,000 words, stored as a PDF) and lifecycle (created externally by a third-party supplier, reviewed and updated by an external supplier, eventually archived). While both overlap and might be used by similar audiences, they differ significantly in their shape and lifecycle.

It's also important to note that AI accesses unstructured data in an atomized form such as snippets of information and data rather than the entire "shape." This means that how we've decided to store and manage unstructured data is not particularly important for its use by AI.

Key Takeaway: *Understanding how useful unstructured data might be for your AI use case requires understanding how it is created and the form it takes.*

Figure 2

Unstructured Data Sets' Potential Sweet Spot for AI Use: High Volatility + High Growth



What to understand about collections of unstructured data

Now we zoom out from individual pieces of unstructured data to look at entire collections of pieces when combined, and how the two differ.

Think of your organization's unstructured data as a distinct set of collections, each with a similar internal shape and lifecycle. These collections grow and change at different rates. For example, consider the product knowledge base. While we understand the format and lifecycle of individual articles, we know far less about the collection's overall growth and volatility.

We know that an individual knowledge base item might be updated during its lifecycle, but how often on average does that take place? That rate tells you the volatility of the collection.

A simpler yet equally important metric is the total size of the collection and its typical growth rate across a given timeframe.

In combination, the volatility and growth figures shed light on how the collection might have to be maintained when used in combination with AI, and also its potential value in that AI use case. A rapidly growing and volatile collection likely draws more regular interrogation and produces more errors, both of which are strong factors in support of its use with an AI-driven solution that is less likely to produce errors and quicker to correct them.

Formal content management systems (CMS) often provide these sorts of metrics for unstructured data, helping to accelerate the process of understanding the suitability of unstructured data for use with AI in organizations already using a CMS.

Key Takeaway: *Knowing how often an unstructured data set is updated significantly helps with decision-making about its suitability for high-value AI use cases.*

What to consider about unstructured data states

After looking at individual pieces and collections of unstructured data, we now turn to their *states*. Remember that the lifecycle of an individual piece might include many stages, from creation through updates to eventual archival (or disposal). Now we determine how that lifecycle can be expressed for collections, so that the right sets of unstructured content become available to AI in the required context.

What do we mean by context? Surely, we'll only want the latest, approved versions of anything. But remember that AI use cases can be both internal- and external-facing – even potentially both – so AI needs information and awareness about the specific relevant unstructured data states for each. For example, when an employee asks about a product vs. when a customer does, that's context.

Think about these example states that could apply to a piece of unstructured data:

- **In preparation:** being assembled
- **In motion:** being peer reviewed
- **At rest:** available for all to view

These examples are simplified, and most well-organized information management practices will have a broader range of states that can be applied. Remember, though, we're talking about all unstructured data, not just documents. It's important to think of a unified way to describe states across the range of unstructured data required to meet your chosen use cases.

Also, remember that unstructured data is not static. When made accessible to an AI product that is delivering context-sensitive experiences, information must be in its most up-to-date state (and, critically, not offered in contexts that should be forbidden).

Retrieval-Augmented Generation (RAG)

A common technical solution to help AI applications with context is retrieval-augmented generation (RAG), an add-on to generative AI systems to help provide context to prompts. This was created to address the situation that foundation LLMs don't know the specifics of an organization and need additional information about precise operational context. Populating RAG systems with organization-specific unstructured content has become common in contemporary AI systems development as a result. You can learn more about this in our two-part series on agents:

AI Agents: What They Are, How They Work, and Where Organizations Might Best Use Them

AI Agents Part 2: The Evolution from Single Systems to Complex Process Orchestration

This leads us to a paramount concern: protecting the integrity, privacy, and security of data exposed to AI systems. This protection is not just a technical requirement but a cornerstone of trust, compliance, and operational reliability. The move from "readiness to act" to "action in production" hinges on the ability to manage three critical protective dimensions: veracity, privacy, and security.

Key Takeaway: *Knowing which unstructured data, in which state, is suitable for which audience helps you plan the data's segmentation for the AI use cases you're developing for internal and external audiences.*

How to protect veracity

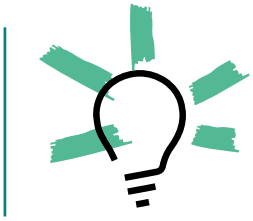
The greatest risk to an AI initiative is not technical failure but the propagation of misinformation. When AI agents generate answers or make decisions based on your unstructured data, you must have absolute confidence in the source material's accuracy and relevance. This is the principle of veracity.

Unstructured data is often a living record. A knowledge base article may be outdated, a draft procedure may be under review, and a customer email may contain unverified claims. If an AI system retrieves and presents this information as fact, it erodes user trust and can cause significant operational or reputational damage.

Your data management processes must enforce a chain of custody for truth. This requires the following capabilities:

- **State-aware retrieval:** Your RAG or AI retrieval systems must be explicitly configured to access data in the correct state (e.g., "approved," "published," "current fiscal year"). They must exclude drafts, archived versions, or deprecated policies unless specifically queried in a historical context.
- **Provenance and citation:** Any output from an AI system must be traceable back to its source documents. This is non-negotiable for auditability, user verification, and continuous improvement of your knowledge sources.
- **ROT awareness:** Understand how to identify unstructured data that is redundant, obsolete and transitory (ROT) and exclude it from active AI retrieval pools. For example, a product manual for a discontinued line or a policy superseded two years ago should trigger an alert for review or automatic deprecation.

Key Takeaway: *Have a plan to protect the veracity of what an AI application pulls from your unstructured data to ensure that it remains trusted and relevant to its audience.*



The greatest risk to an AI initiative is not technical failure but the propagation of misinformation.

How to protect privacy

Sensitive and personal data such as personally identifiable information (PII), intellectual property, and confidential business intelligence reside primarily in unstructured data. An AI system that broadly ingests this data without controls creates a massive compliance and liability risk.

Emails, customer support tickets, HR documents, and contract repositories are filled with sensitive data. Aggregating this into an AI's context window can lead to unintended data leakage, violations of GDPR/CCPA, and exposure of competitive secrets.

You must implement privacy-by-design at the data layer, before the AI ever sees it.

- **Classification and redaction:** Use automated tools to scan and classify unstructured data for PII (names, account numbers, addresses) and sensitive keywords. Create "clean" versions of documents where this data is redacted or tokenized for use in AI contexts that don't require full details.
- **Audience-based data segregation:** Your data architecture must support siloing. Customer-facing AI agents should only have access to knowledge bases approved for external viewing, not internal strategy memos. Internal HR assistants should access policy documents but not individual employee files without strict, logged authorization.
- **Consent and purpose limitation:** Map your data collections to the legal basis for their processing. Ensure that using data in an AI training or inference context aligns with its original collection purpose or has the necessary consents.

How to protect security

Activating data for AI creates new attack surfaces as the data becomes a high-value target for extraction and a potential vector for prompt injection or data poisoning attacks. Controlling access and preventing misuse are paramount concerns.

How do you prevent a malicious actor from querying a seemingly benign customer service chatbot to extract the full text of internal financial reports via cleverly engineered prompts?

Security must shift to envelop the entire AI data pipeline.

- **Role-based access control (RBAC) for AI:** Just as users have permissions, so should your AI agents. Define service principles for your AI applications with strict, minimal-access rights to specific data collections.
- **Input/output sanitization and monitoring:** Deploy guardrails that monitor prompts and generated responses for suspicious patterns that indicate data extraction attempts, prompt injection, or the generation of harmful content.
- **Immutable audit trails:** Log every interaction – the prompt, the data snippets retrieved, and the response generated. This trail is essential for forensic security analysis, regulatory compliance, and tuning system behavior.

Key Takeaway: *Protection is not a post-deployment add-on. Veracity, privacy, and security filters must be designed into the data selection and preparation workflow from the start. The data you choose to activate must pass through these three gates before it is deemed fit for AI consumption.*

Call to Action: From Readiness to Responsibility

The four-year period of "readiness for readiness" has served its purpose. The checklist phase is over. The foundational questions about processes, tasks, and skills have been asked. Moving valuable AI projects from the whiteboard into the workflow now takes center stage.

Plan of action

This paper has provided a pragmatic path forward, focusing on the most critical and complex fuel for AI: your unstructured data. The journey is not about boiling the ocean, but about intelligent activation:

1. Target with precision: Identify and activate only the specific collections of unstructured data that power your prioritized use cases.
2. Understand the terrain: Analyze the shape, lifecycle, volatility, and state of these collections to architect how they will flow into your AI systems.
3. Build in protection: Design for veracity, privacy, and security from the first step of data selection. These are your non-negotiable filters.
4. Govern with clarity: Establish the policies that will make this practice repeatable, scalable, and safe.

Figure 3
Your 90-Day Action Plan

Weeks 1-4:
Convene & Select

Gather the owners of your top-priority AI use case and the stewards of its required data. Map the specific, bounded collections of unstructured data involved. Deliverable: A one-page "Data Activation Charter" for the pilot.

Weeks 5-8:
Analyze & Design

Profile the selected collections for shape, lifecycle, and volatility. Work with IT and security architecture to design the retrieval pipeline, incorporating state management, access controls, and privacy filters. Deliverable: A technical design document for the secure AI data feed.

Weeks 9-12:
Prototype & Draft Policy

Build a minimal, secure pipeline to feed this data into a test environment. In parallel, draft the first iteration of the AI Data Supply Chain Policy and Context & Audience Policy using this pilot as the model. Deliverable: A working prototype and two draft policies for leadership review.

The era of AI-assisted operations is not coming; it is here. The organizations that will lead are not those with the most data, but those that can most responsibly, securely, and effectively activate the right data at the right time. The mandate for the CIO or CAIO in your organization is clear: move beyond planning and orchestrate AI's activation. Start now, start focused, and build the foundation of data intelligence that will power your AI advantage for years to come.

The readiness phase is over. It's time for responsible action.

About OpenText

OpenText is a leading Cloud and AI company that provides organizations around the world with a comprehensive suite of Business AI, Business Clouds, and Business Technology. The company helps organizations grow, innovate, and become more efficient and effective, in a trusted and secure way, through information management.

OpenText Content Cloud delivers intelligent content management that helps organizations access, share, and apply structured and unstructured information faster and with confidence. As a market leader with a full suite of content services, from capture and IDP to information governance and data archiving, OpenText Content Cloud fuels AI-driven work at scale.

opentext™

About Deep Analysis

Deep Analysis is an advisory firm that helps technology vendors, buyers, and investors understand and address the challenges of innovative and disruptive technologies in the enterprise software marketplace. The firm's work is built on decades of experience advising and consulting to global technology firms large and small.

Deep Analysis

02.26 | 248-000169-001