carahsoft

opentext™

Four practical ways to automate AI readiness: Accelerating effective GenAI strategies

How automated content analysis, preparation, and protection can build organizational trust and confidence to accelerate AI strategies and implementation

Thank you for your interest in exploring this content.

Carahsoft is the **Trusted Government IT Solutions Provider**® supporting a broad portfolio of industry-leading technologies through NASA SEWP V, CMAS, The Quilt and a wide range of other contract vehicles.

As the **Master Government Aggregator**®, Carahsoft connects government agencies, industry partners, and technology providers to deliver innovative, mission-focused solutions.

In partnership with OpenText Government Solutions, we provide technology solutions that drive modernization, strengthen operations, and ensure compliance with evolving government standards.

To learn more about how Carahsoft can support your technology needs, please visit carahsoft.com

Explore More Resources:
carah.io/opentextresources

Join Events & Webinars:
carah.io/opentextevents

Discover Technology Solutions:
carah.io/opentextoverview

Learn About Procurement:
carah.io/opentextcontracts

Connect With Our Team:
OpenText@Carahsoft.com
888-662-2724

opentext™

WHITE PAPER

Four practical ways to automate AI readiness: Accelerating effective GenAI strategies

How automated content analysis, preparation, and protection can build organizational trust and confidence to accelerate AI strategies and implementation

Contents

Executive summary	3
Systemic barriers to AI readiness	4
The four practical ways to automate AI readiness	6
1: Discover: Understand the content landscape and risk	7
2: Prepare: Provide AI with full content visibility and context	8
3: Govern: Reduce risk through data protection and improved accuracy	10
4: Act: Apply AI in practice and with measurable outcomes	12
Foundational technology for automating and managing AI readiness	14
Conclusion	15
Resource links	15

Did you know?

of organizations lose potential business opportunities due to the inability to access information quickly.

face foundational challenges with information sprawl and inconsistent content quality.

say adopting AI has accelerated their focus on information governance.

Source: *Foundry, MarketPulse Survey for OpenText: GenAI Adoption and Readiness, January 2026*

Most organizations are investing in generative AI, but also struggle as their content isn't AI ready. The majority of data, by volume, is unstructured and spread across dozens of repositories, formats, and work processes. Large volumes of critical information flow through email, shared drives, collaboration tools, and media applications or sit underused in legacy archives. Even when a formal content management and governance system is in place, broader content sprawl and noncompliance can prevent effective GenAI grounding, resulting in less accurate inference.

AI strategies are much less effective when content is disorganized, unprotected, and poorly labeled. When generative AI lacks complete and current information or is overburdened by irrelevant content, it will fail to generate sensible and useful responses. While large language models (LLMs) are trained through an enormous quantity of generalized knowledge, they lack a detailed and informed understanding of business operations. Concerns about security and privacy may lead an organization to avoid using valuable content in AI-enabled processes. But this can result in frustration and distrust in AI and slow or stall the implementation of an AI strategy. When AI readiness is process and context-focused, AI strategies are far more likely to be successful.

This is the true AI readiness gap.

At the same time, boards and regulators are raising expectations for AI governance. Regulations such as the EU Artificial Intelligence Act and a growing number of AI-related regulations are raising a new risk profile and pose a patchwork of compliance requirements. Leaders must show that AI enhances decision-making and processes at scale. AI must not erode the quality of work product, increase task time, or create data leakage or other compliance risks.

This white paper covers how content quality, management, and governance gaps prevent organizations from realizing their full AI-first potential and what steps (and technology) organizations can implement to prepare themselves, including:

- Why today's content systems and workspaces are not designed for AI readiness.
- How the four practical aspects of automating AI readiness provide an operating model
- Why AI governance and information governance are one continuous process, not separate.
- How AI-assisted automation can make information discoverable, usable, governed, and actionable for AI.
- How AI content management solutions can close the AI readiness gap and create an AI trusted zone.

Systemic barriers to AI readiness

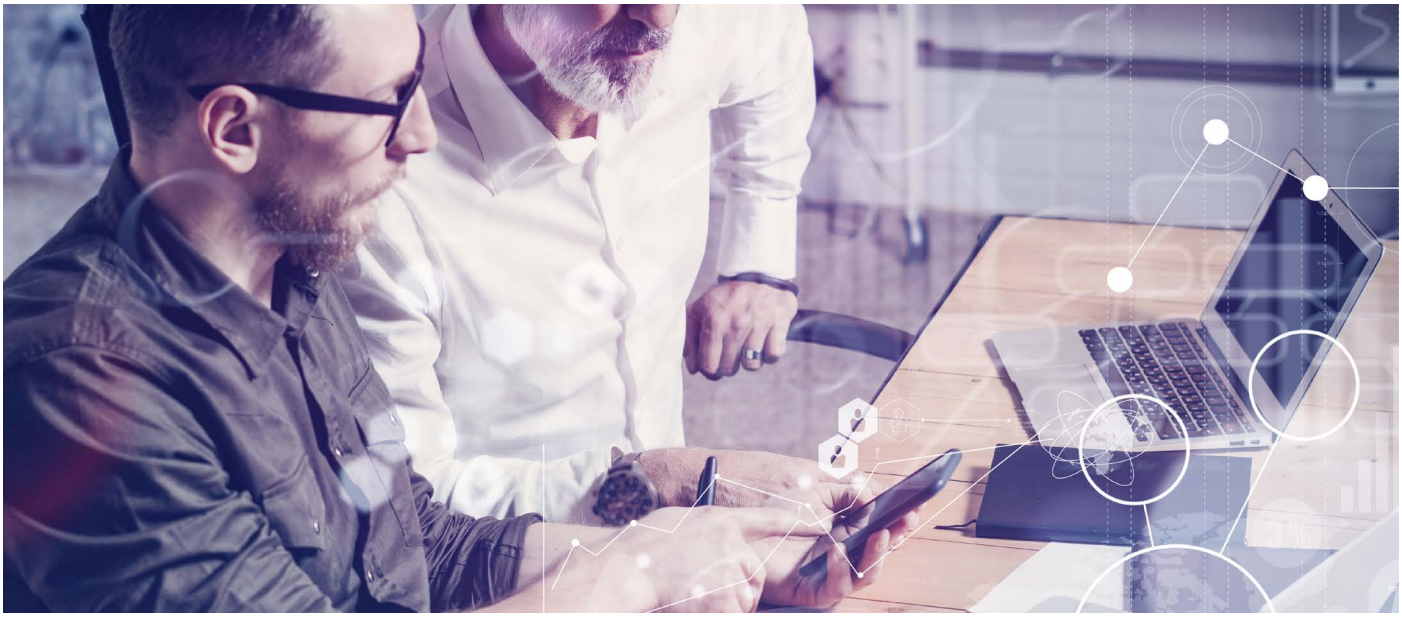
Most organizations believe their content management systems, collaboration tools, and workspaces provide a strong foundation for AI. While these tools appear structured, governed, and searchable, how users manage and govern their information creates obstacles for AI integration and introduces compliance risks. Not all content is in a usable format or adequately protected for AI access. An AI readiness program aims to address these issues and foster an environment where an AI strategy can succeed.

Silos, content sprawl, and shadow IT

In most organizations, silos, content sprawl, and shadow IT are common. As a result, important source material and current decision-making data are scattered across email, team collaboration apps, shared drives, document repositories, archives, and line-of-business applications. Even shadow AI—the extracurricular use of public AI—creates regulatory issues, data leaks, and even incorrect inference and decision-making. Without active, automated oversight and organization, safe and informed GenAI at scale becomes difficult or impossible.

Inaccessible or opaque content

Large volumes of unstructured content are stored in formats that AI cannot interpret—audio, video, scanned images, legacy documents—or in systems that AI cannot access. Even if your IT is aware of this data, it creates an AI “dark zone” that sabotages the effectiveness of AI implementations. Even organizations with mature content management programs often have substantial amounts of hidden or missing content that semantic search may overlook, making it unavailable for retrieval-augmented generation (RAG). If AI cannot see the information, it cannot consume or be grounded on it.



Poorly organized and governed content

AI's reliability depends on the quality of the grounding data. Outdated files, duplicate documents, missing metadata, and inconsistent tagging lead models to guess meaning instead of relying on clear, accurate, and complete data. This can cause hallucinations, off-topic responses, and decreased trust. Poor organization and metadata can also lead to grounding and inference biases that can be difficult to detect and mitigate.

Insufficient content security and data leak protection controls

When security policies are unevenly or poorly managed across repositories, it becomes difficult or impossible to control how AI processes PII, sensitive, proprietary, or other regulated information. Compliance with data sovereignty and jurisdictional AI regulations can be difficult or impossible. Permissions-aware AI assures that the AI is operating under the same rules and authorities as your users. Without consistent, policy-based security measures aligned with user permissions and privileges, AI could process content, increasing the risk of data leaks for the organization.

Fully realizing the potential of an AI strategy.

While GenAI chatbots and localized AI assistants can increase productivity, the full potential of AI within an organization might be limited if these systemic barriers aren't addressed. Only a small portion of the overall content is accessible at any given time; full context and supporting grounding information are often missing, and security risks are exposed. All of these factors decrease the accuracy, usefulness, and trust in AI.

To address this, organizations should establish an AI readiness framework that supports continuous improvement and implement an automated AI content pipeline that prepares machine-readable, protected content, and generates a contextual knowledge graph to realize the full potential of an AI strategy.

What generative AI changes (and what it does not)

What changes	What does not change
• Reduced time finding key information and facts	• Need for organized, secure, and well-governed content
• Speed and scale of content generation	• Regulatory and privacy obligations
• How decisions are made and documented	• Accountability for business outcomes and removing bias

The practical ways of automating AI readiness

AI readiness is not a single project. It is an ongoing practice that shapes how information is discovered, prepared, governed, and used. The following four practical ways help organizations identify where their content systems fall short and what must change to support AI responsibly and at scale.

These four ways create a durable foundation for responsible, scalable AI adoption.

Discover

Create visibility across all content

"Discover" helps to build a more complete, curated repository on which AI can ground its responses and decisions.

Prepare

Transform content into AI-ready intelligence

"Prepare" addresses the structure and quality of content to prepare it for AI use.

Govern

Embed policy and protection into the content layer

"Govern" ensures AI operates safely and ties AI readiness to organizational trust and accountability.

Act

Apply AI in the context of real work

"Act" moves AI from pilots into daily workflows.

McKinsey notes that companies have traditionally focused on structured data, which represents only about 10% of available data; the other 90% is unstructured content such as emails, chats, documents, and media.[1]

1. Discover:

Understand the content landscape and risk

AI readiness begins with a clear view of the information the organization is curating for AI. A large share of valuable institutional knowledge lies in relatively unmanaged and siloed repositories; understanding what content is there, how much of it needs to be incorporated in the AI strategy for grounding, and what should be ignored or removed for compliance reasons can help to improve AI strategy outcomes.

The problem: If AI can't see it, it's not part of your AI strategy

Ad hoc work processes, shadow IT, and content sprawl often hide valuable data that could be used in grounding. In day-to-day work, without assistance, users may store important information only within email, team chats, and unmanaged share drives, resulting in a fragmented environment where critical knowledge is lost.

The fragmented landscape leads to knowledge gaps that hinder otherwise successful AI projects.

What AI-ready discovery looks like

This discovery approach can dramatically expand the available grounding, remove extraneous information, and better organize the information around meaningful business context.

A discovery approach that supports AI readiness should:

- ✔ Connect to content sources quickly to identify and onboard valuable content.
- ✔ Identify which legacy data is likely to have the most impact for onboarding to a content management platform.
- ✔ Provide a federated search that can group content across silos.
- ✔ Normalize and structure target content and metadata consistently.
- ✔ Classify content by subject, type, format, and sensitivity.
- ✔ Eliminate redundant, obsolete, and trivial (ROT) content.
- ✔ Allow for conceptual and process-oriented views and visualizations of content.

With this foundation in place, organizations can quickly see:

- Which content is already suitable for AI use.
- Which content—and therefore which AI use cases—requires further review.
- Where risk lies and additional governance controls are needed.

Discovery turns potential AI grounding content from a hidden, fragmented, flat corpus into a navigable, governable resource, ensuring AI never operates in the dark and that the organization always maintains visibility of its knowledge.

1 McKinsey & Company, *Charting a path to the data- and AI-driven enterprise of 2030,* September 5, 2024

As per Accenture, 48% of organizations say they lack enough high-quality data to operationalize generative AI initiatives.[2]

2. Prepare:

Provide AI with full content visibility and context

The next challenge is preparing content so AI can interpret it accurately and consistently. Most organizations still rely on manual or over-simplified tagging, basic folder structures, and simple templates to organize information. These methods may support human search, but they do not yield a navigable knowledge graph that can inform and guide RAG-based grounding to generate insightful responses. As a result, accessible but non-contextualized content can lead to disappointing GenAI inference.

The problem: Unstructured content that AI cannot reliably use

A large share of enterprise information sits in formats that AI cannot easily analyze. Scanned documents may lack machine-readable text used for semantic search. Audio and video files rarely have transcripts, time codes, or speaker labels. Emails and PDFs may contain important insights, but no consistent metadata to describe them. And across repositories, few documents carry clear labels for people, organizations, topics, or events.

This creates several issues when generative AI attempts to work with unstructured content:

- AI must infer structure from scratch each time it interacts with a file.
- More computing resources are required to understand the content, yet the answers are often less accurate.
- AI governance rules, such as restrictions based on content type, sensitivity, or business context, become harder to enforce.

Without preparation, AI is forced to guess. And when AI guesses, reliability drops and risk increases.

Metadata and text extraction technologies:

- Classification
- OCR
- Speech recognition
- Facial recognition
- Objection detection
- Named entity recognition
- Semantic tagging
- Geolocation

Content enrichment technologies

Preparing content for AI involves transforming messy, inconsistent information into structured, machine-readable context. An AI-ready preparation layer uses multimodal AI and analytics to enrich content so both humans and AI can understand it.

This includes:

- Intelligent classification to identify document type and related ontology.
- OCR to extract text from scans and images.
- Speech-to-text to generate time-aligned transcripts for audio and video.
- Named entity extraction to identify names, organizations, locations, events, and topics.
- Semantic tagging and clustering to reveal how content relates across projects, cases, or time.

2 Accenture, *Data readiness in the age of generative AI*, 2024.

- Facial recognition and trained object detection in video and images to identify people, physical items, and situational state.
- Geolocation, which can help to identify jurisdiction, AI governance policy, and data sovereignty compliance requirements.

These capabilities convert unstructured data into enriched content that AI can use confidently. They provide rich labeling necessary for a more complete context and to determine which information is appropriate for which AI use cases.

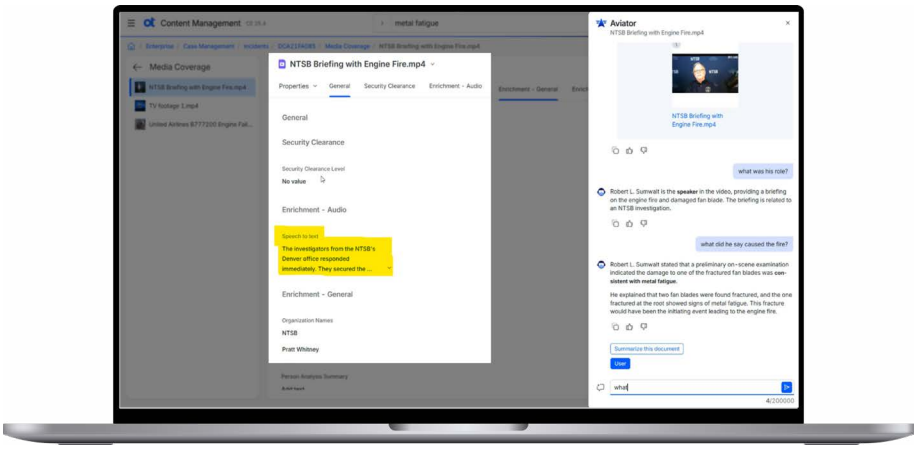
Why preparation matters for long-term AI readiness

Enrichment strengthens both the quality and the safety of AI outcomes:

- ✔ Content becomes easier for humans and AI systems to interpret
- ✔ Retrieval augmented generation (RAG) architectures can use knowledge graphs and indicators such as content date to better inform context
- ✔ Governance decisions, such as redaction, retention, or risk classification, become automated and consistent
- ✔ Humans remain in the loop to validate low-confidence predictions, reducing the impact of errors

Preparation is not a one-time cleanup effort. New content enters the organization every day. AI readiness depends on continuous enrichment pipelines that run in the background, keeping the knowledge base fresh, accurate, and aligned with the organization's evolving governance requirements.

Turning unstructured information into structured intelligence is what enables AI to support real work - safely, efficiently, and at scale.



GenAI assistant (OpenText Content Aviator) giving accurate responses for an audio file by having access to the transcription as part of the audio file's metadata.

McKinsey notes that even among "high-performing" organizations (those attributing a meaningful share of EBIT to Gen AI), 70% report difficulties developing processes for data governance and integrating data into AI models quickly.[3]

3. Govern:

Reduce risk through data protection and improved accuracy

Governance is often treated as a policy exercise that is well documented, well defined, and well intended. Most organizations already maintain extensive frameworks that outline how information should be handled. But policy alone does not create AI governance. To support AI readiness, governance must move from documents and guidelines into the content itself, where it can influence how AI accesses, interprets, and uses information.

The problem: Inconsistent or manually applied policies across information

Enterprises typically have strong governance foundations:

- Privacy and data protection policies
- Obsolete information, draft versions, or expired "convenience copies"
- Lack of observability for content completeness, compliance, and usage
- Acceptable-use guidelines for AI

The challenge isn't the lack of a governance policy, it's applying governance consistently across all unstructured content. Sensitive information is scattered across documents, emails, transcripts, and legacy repositories. Not every system enforces the same security or retention rules. Organizations can't rely on users to manually and consistently apply metadata, organize, and follow policies for everything, or even a portion of content. Users also often forget to delete expired drafts, old versions, and convenience copies, which can affect the accuracy of AI responses.

From an AI governance standpoint, this introduces several serious vulnerabilities. AI relies entirely on the governance in place and depends on automated enforcement of policies and policy-based guardrails. AI might base its answers on outdated or non-compliant content. It could reveal personal data or confidential information in its responses. It might even depend on content that should have been restricted, remediated, or deleted under policy. Weak guardrails lead to AI amplifying non-compliance issues and elevated risk.

What AI-ready governance looks like

AI readiness requires governance that is active, automated, and tightly coupled to the content itself. An AI-ready governance program identifies risks, enforces policies, and proactively prevents unsafe content from being processed by AI on behalf of unauthorized persons.

This includes:

✔ Automatic detection and labeling of PII and sensitive data in documents, emails, transcripts, and other formats. Labeling also helps to measure exposure to potential bias or ethical issues that can inform future guardrails and mitigations for compliance.

3 McKinsey & Company, *Charting a path to the data- and AI-driven enterprise of 2030,* September 5, 2024

- ✔ Classification of content to ensure proper retention policies are applied, eventually removing convenience copies and expired or obsolete content.
- ✔ Content management with strong, well-governed native AI capabilities to create an integrated “AI trusted zone” for content and to deliver AI assistance directly to users, agents, and through AI interactions with related integrated products and services
- ✔ Observability in the form of an overview of ontologies, measures of completeness, and auditability of usage, and biases
- ✔ Policy enforcement by role and user before AI is applied, or provide AI with safe alternatives, such as redacted content that removes PII or other sensitive information

These capabilities put governance into practice rather than leaving it as an aspirational standard.

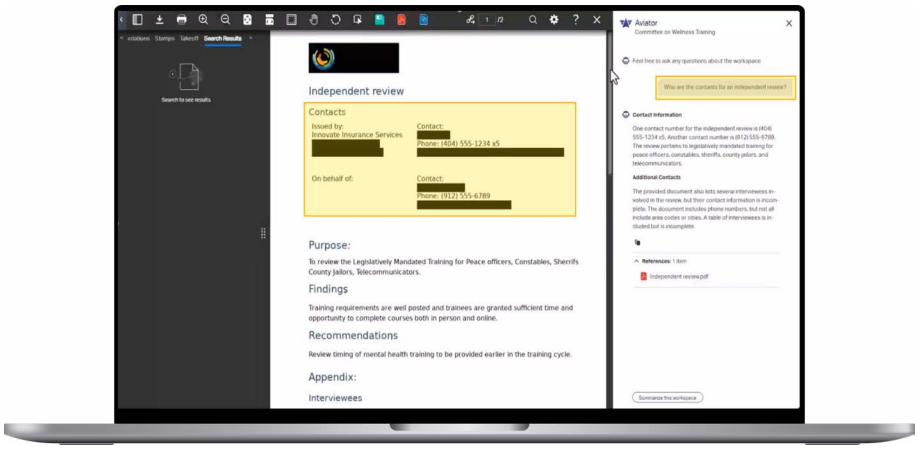
What this enables for AI governance

When governance is embedded at the content level, policies become enforceable and measurable:

- Compliance and risk are evaluated and identified continuously, so that policy can be applied ahead of any AI processing
- Redaction applies equally to human access and AI retrieval, ensuring models never see sensitive information
- Governance actions are logged and auditable, supporting compliance and operational transparency

This alignment ensures that AI readiness and AI governance move together. As AI scales across the organization, the guardrails scale with it.

By embedding policy in the content itself, organizations reduce risk, build trust, and create the necessary conditions for responsible AI adoption. Governance becomes a living capability, one that protects information, supports compliance, and ensures AI acts within safe and well-defined boundaries.



GenAI assistant (OpenText Content Aviator) giving compliant responses for a document where PII has been redacted.

4. Act:

Apply AI in practice and with measurable outcomes

Even when organizations make progress on discovery, preparation, and governance, many still struggle to achieve the full impact of AI. In many standalone pilots, generative AI is used with a limited, curated set of content, and users interact through a standalone chat interface with less exposure to risk. These pilots can showcase AI's capabilities and highlight its potential, but they often stall due to poor performance, governance concerns, and difficulty scaling.

The problem: AI that lives outside of daily work

When AI is applied within an information silo without regard for work process context, its effectiveness is limited. Responses may be engaging, but they are not fully informed as to the specific customer, incident, contract, or case context that the user is managing. Context may also include a combination of these, such as the customer and the case together. As with users, AI doesn't work as well within such a silo, leading to faltering or poorly performing AI strategies.

Common limitations include:

- AI sits outside the applications and workspaces where work truly happens.
- Responses are not grounded in the content associated with a specific case or process.
- Users cannot turn insights into governed actions within content systems.
- Feedback is not captured, so enrichment and governance do not improve over time.
- AI capabilities across various systems are disconnected, limiting context and actionability

From an AI readiness and AI governance standpoint, these gaps create additional challenges. Organizations cannot easily control which content an assistant uses in each scenario. Human-in-the-loop review becomes harder to enforce for high-risk use cases. And most importantly, leaders struggle to connect AI activity to measurable business value.

What AI-ready action looks like

In an AI-ready organization, AI does not sit on the sidelines. It operates in context, embedded directly within the workspaces and processes employees rely on daily. This ensures that AI is grounded in the right information, constrained by governance rules, and able to produce insights that matter in the moment.

AI in context means it is naturally integrated directly within business workspaces that represent common business processes such as cases, projects, or accounts. It also becomes content-centric, focused on the user's cross-document view of claims, investigations, onboarding, or incident response. This workspace-centric approach ensures that AI draws on relevant, managed information to deliver valuable insights and actions.

When AI is embedded this way, assistants can:

- ✔ Answer questions based on the content in the specific workspace or case a user is working on.
- ✔ Provide citations and supporting evidence so answers can be verified.
- ✔ Respect legal holds, retention rules, access controls, and geographic restrictions.
- ✔ Trigger governed actions, such as filing a document, updating metadata, launching a workflow, directly in the content system

This shift turns AI from a passive tool into an active participant in business processes.

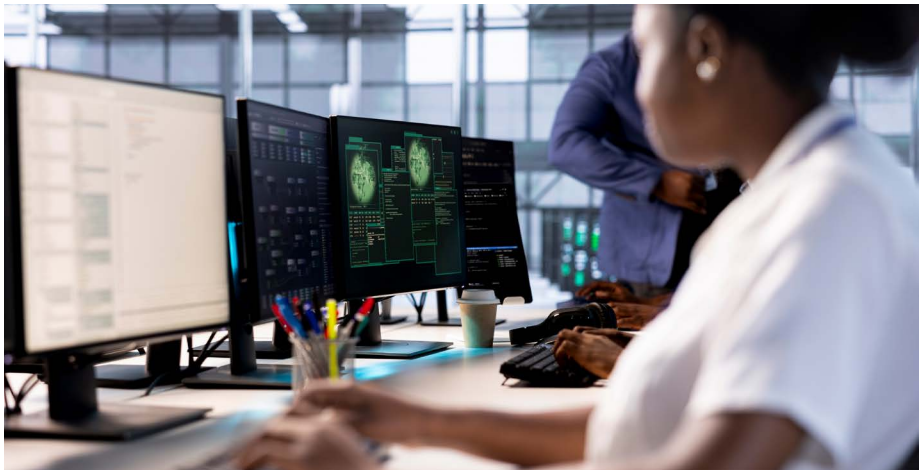
Why context makes AI readiness and AI governance sustainable

AI in context does not rely solely on generative AI. Instead, it uses a hybrid approach that applies the right technique for each task:

- Machine learning and analytics for classification and preparation of content
- Metadata that ties content, application data, and work processes together to form a knowledge graph, mapping the full relational context
- Generative AI for natural language understanding and generation, and agentic AI within policy-driven governance guardrails

This balance controls cost, improves accuracy, and ensures AI stays aligned with enterprise governance requirements. Models focus only on relevant, prepared, governed content, reducing risk and improving performance.

By putting AI where work happens, and grounding it in structured, governed content, organizations create the conditions for real transformation. AI readiness becomes actionable, and AI governance becomes practical and enforceable. Most importantly, AI delivers outcomes that employees can trust and leaders can measure.



OpenText's AI content management approach provides a scalable technology foundation that addresses these pillars:

- OpenText™ Knowledge Discovery provides AI-driven content analysis, metadata enrichment, classification, and sensitive data protection on an ongoing basis, so content remains accessible, in context, and safe
- OpenText content management systems provide information governance and serve as a contextual hub through the organizing principle of the business workspace
- OpenText™ Content Aviator™ provides embedded AI with native, integrated governance guardrails and AI in context

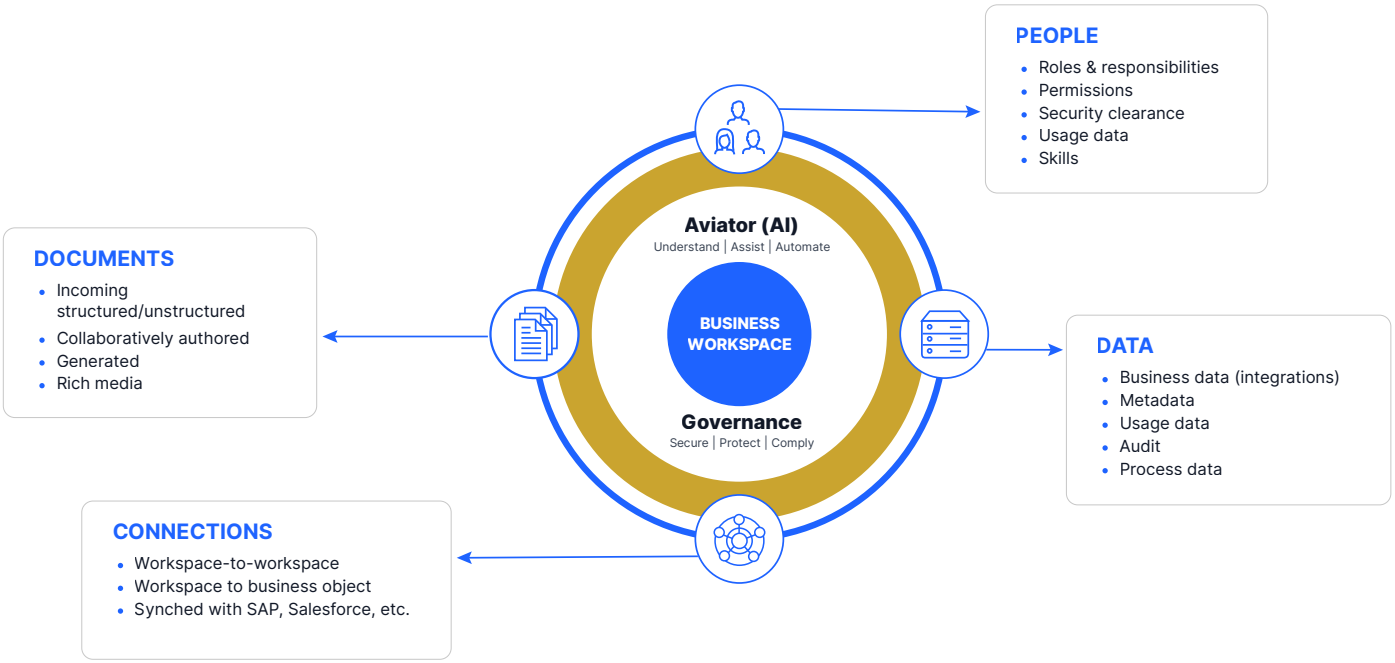
Foundational technology for automating and managing AI readiness

Organizations need a holistic and integrated approach to build their AI strategy. This involves embracing all four practical aspects of automating AI readiness discussed in this paper as part of a comprehensive AI content management program.

A unified path to AI readiness and AI governance

By integrating content management, multimodal enrichment, strong governance, and embedded AI, OpenText provides a practical route to AI readiness. Content is organized, governed, and protected. AI is more grounded and contextually relevant, delivering safer and more meaningful results. Users gain powerful new ways to work with information without leaving the ECM systems they rely on daily.

This alignment of AI content management creates a modern content architecture where AI readiness and AI governance advance together, strengthening as the organization grows.



AI content management helps bridge the four ways of automating AI readiness centered on the business workspace

Resources

Assess your AI readiness ›

Learn more about AI governance and AI readiness with OpenText ›

Explore OpenText Content Aviator for secure AI content assistance ›

Discover OpenText Knowledge Discovery for AI ready content enrichment ›

Read the full AI readiness research ›

Blog: Secure, trusted and ethical AI ›

As experimentation and planning around AI transitions into a mandate for material results, successful adoption will require a strong technology foundation and willingness to look at how content drives business processes. Organizations that can successfully orchestrate the discovery, preparation, and governance of their unstructured data will mitigate the risks of hallucinations and data leaks and unlock a new level of insight and productivity.

AI readiness is an ongoing policy, analysis, and technology-driven discipline that thrives on continuous improvement. It ensures that AI strategies remain resilient, compliant, and deeply impactful. The path to a trusted, AI-driven future requires a foundation built on clarity and control. Crucially, this control relies on a robust human-in-the-loop (HITL) framework, where organized content allows subject matter experts to validate how AI is grounded and governed, and maintain oversight. With a comprehensive approach to AI content management, organizations can move forward with confidence—transforming their vast information archives from a liability of sprawl into their most significant competitive advantage.

Copyright © 2026 Open Text • 05.26 | 262-000212-001P