

Generative AI governance essentials

Content management and strong information governance for more trustworthy, secure generative AI



Contents

AI in the service of your organization	3
Generative AI forever changes our relationship with information	3
Risk considerations when implementing GenAI	5
Trustworthy GenAI depends on strong content management and governance	8
Getting grounded with excellent content management	9
OpenText Content Aviator	10
Summary	13



AI in the service of your organization

Generative AI (GenAI) and large language models (LLMs) that can understand human language at scale have forever changed how knowledge workers interact with information. AI is bringing unprecedented productivity in the form of language summarization, making new conceptual connections, drafting content, and visualizing complex data rapidly. We can now interact with impossibly large information stores on the web and within our organizations, as well as more quickly synthesize content based on data from others.

AI can also introduce its own risks. Understanding the foundational requirements for safe and effective AI is an essential first step. Few users fully understand the risks of unrecognized errors in generated content or data, privacy, and sensitive data breaches, regulatory infractions, and more. It is critical that IT and governance leaders address and mitigate against these risks.

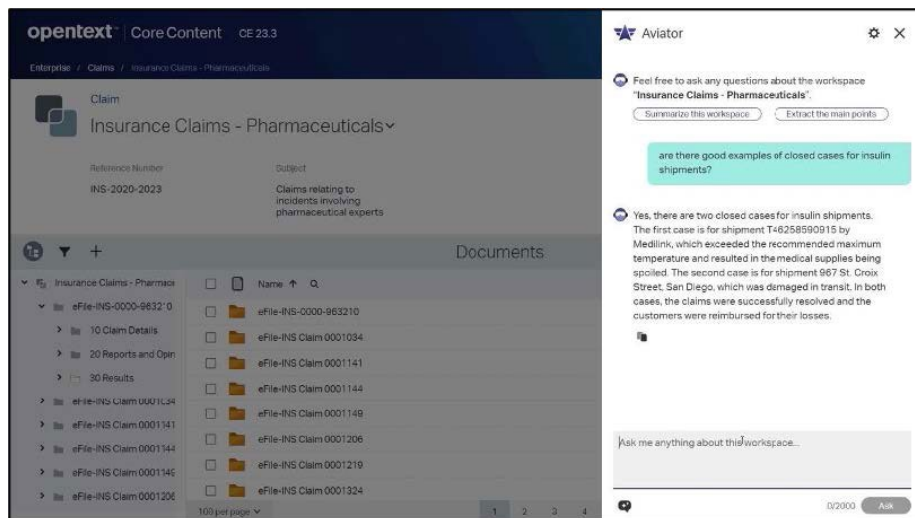
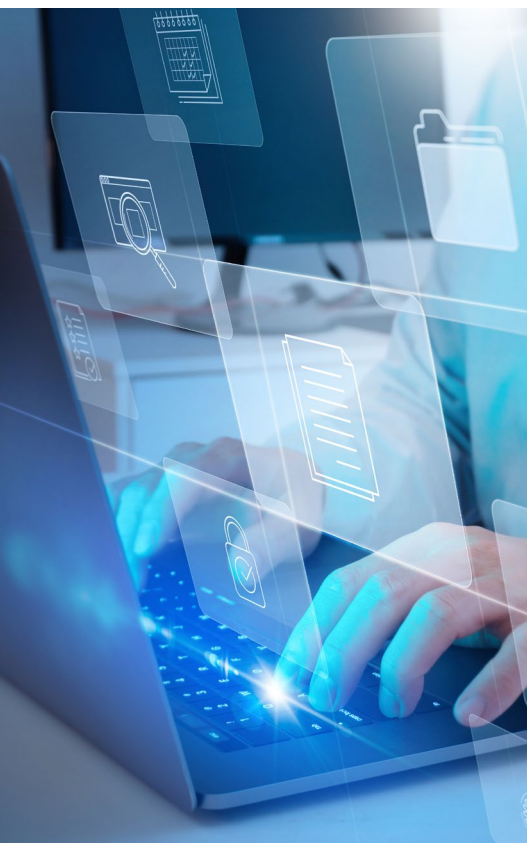
Reimagining how knowledge is discovered and transforming productivity with AI is now within reach if organizations follow core principles of AI governance. This includes managing and securing their data and content, making it available to GenAI, and directing the results.

Generative AI forever changes our relationship with information

Too many search results, too many stray document versions, and too many browser tabs. Knowledge workers are overwhelmed with an unending stream of information and content. It's a drain on productivity and takes our attention away from more productive work. We must read, understand, absorb, and prioritize—often with little guidance—before we can communicate, create, and make critical decisions.








Generative artificial intelligence (GenAI) is an amazing new class of intelligent computing that can make sense of vast volumes of information by understanding human language, pictures, sound, and even video. It accomplishes this impressive feat with the help of AI large language models (LLMs) trained on vast libraries of text, data, and pictures. This gives it the ability to recognize text, speech, and objects across many different languages, cultures, and situations.

When combined with domain-specific information and “retrieval-augmented generation,” GenAI can respond to human requests with striking output in the form of concise abstracts, reformulated content, and translated text, even combining content from multiple documents into a single new work draft. It can produce tables, pictures, audio, and video. These results speed up the creative process and inspire and motivate knowledge workers.



[OpenText™ Content Aviator](#) responds to a typical natural language query

GenAI is immediately applicable in a wide variety of use cases:

-  **Search** for key facts and concepts within workspaces and documents
-  **Summarize** all or part of a large workspace or document
-  **Generate** drafts of emails, memos and other work papers
-  **Translate** responses into the user's native language
-  **Formulate** tables and figures
-  **Locate** reference documents that answer your question
-  **Annotate** graphics with searchable metadata

GenAI is changing our relationship to information retrieval and management along with rapidly emerging capabilities and new use cases. We can now interact with documents, groups of documents, or entire workspaces as if communicating with the author(s). We can gather information using familiar terms and concepts—even if they differ from the exact language in the text—and still navigate to the correct information. We can even use GenAI to search for information and discover entirely new facts and concepts that we didn't even know to look for, all without reading the original information directly.

While GenAI can help us quickly understand content, an important aspect of using it is being able to trust what it produces. That means knowing how to evaluate its output, and having ready access to source materials it is using.

72% of leading organizations say managing data is a top challenge preventing them from scaling up AI use cases

- McKinsey & Company¹

Factual errors, known as “hallucinations” in natural language processing, can result from a poor or overly generalized GenAI implementation. Reducing such faults is an intense area of R&D. One proven method of avoiding hallucinations is to arm GenAI with plenty of factual “grounding data,” dramatically improving overall accuracy. In contrast, incomplete and “noisy” data (repetitive, trivial, or inaccurate) can lead to increased hallucinations. Enhancing GenAI’s accuracy will improve its usefulness and application to further use cases.

Risk considerations when implementing GenAI

Many organizations are excited about harnessing GenAI’s powerful capabilities but may hesitate due to perceived risks. Understanding and [mitigating these risks](#) is essential when developing a GenAI strategy. These risks are often not caused by AI itself but are existing problems that are more easily activated or exploited because AI makes content more accessible.

GenAI excels at consuming, summarizing, and producing unstructured content across many different formats, including text, images, spreadsheets, audio, and video. The primary insight in managing risk for GenAI is that it is an active participant—like any human user—in content-rich processes.

AI governance is a form of information governance. Organizations must bring transparency, predictability, and accountability to GenAI-infused processes, as well as implementation and real-world systems operation. This is why content management is a foundational element of any complete GenAI solution.

While the risks depend significantly on the industry and the user’s job function, there are some common areas of risk that every organization should thoroughly consider and mitigate.

- **AI model safety** is an industry-wide concern meant to avoid outputs that may be offensive, harassing, dangerous, or malicious. Safety is a major research area for LLM producers. Choosing a model from a source committed to AI safety is critical for most use cases.
- **Accurate grounding data** is crucial for precise GenAI responses. If the data isn’t well-curated, GenAI may produce inaccurate answers or “hallucinations.” Moreover, if grounding data is poorly interpreted or noisy, it can adversely affect accuracy.
- **Keeping content secure** is a core tenet of any GenAI system. Grounding data is often the organization’s most sensitive content, so it must be protected. Security should be applied to each user’s permissions without restricting its use to those who need it. The grounding data, once processed, must also be kept secure to ensure the integrity of GenAI responses.
- **Privacy protection** is critical in virtually all environments. It should not be possible to exploit GenAI to expose personally identifiable information (PII), financial data, or protected health information.

¹ McKinsey & Company, The Data Dividend: Fueling Generative AI, 2023

EU Artificial Intelligence Act high risk categories

High risk AI systems pursuant to Article 6 of the EU Artificial Intelligence Act include the following use cases:

- **Biometric identification** and categorization of natural persons
- Management and operation of **critical infrastructure**
- **Education** and vocational training
- **Employment**, workers management, and access to self-employment
- Access to **essential private services and public services and benefits**
- **Law enforcement** that may interfere with people's fundamental rights
- **Migration, asylum, and border control** management
- Administration of **justice and democratic processes**



- **Bias and fairness** can unconsciously enter AI processes. LLMs and grounding data can both introduce unexpected bias. User training, careful use case management, and oversight of grounding data can reduce biases and prevent negative impacts and regulatory infractions.
- **Regulatory control and legislation**, such as the EU AI Act, are emerging to govern the use of GenAI, the grounding data used, and AI data processing. Organizations must carefully match their use cases and controls to permissible applications. For example, in some jurisdictions, biometrics such as facial recognition require explicit permission from the subject.
- **Explainability and transparency** are crucial in model selection, data grounding, use case integration, and permitted uses. Users need training on GenAI utilization, training data rights, output authenticity, and allowed/prohibited uses. Organizations must also be able to assure customers, auditors and regulators of safety measures.

Emerging regulations for artificial intelligence

AI regulation is evolving rapidly, as are AI ethics and acceptable use standards in many industries and use cases. Organizations must carefully consider how AI will affect compliance with laws, regulations, and organizational standards. Reconciling these many issues can be complex. For large international organizations, regulation could affect different parts of the business in novel ways depending on the location of the data, the customers, or the public policy implications of how it is used.

As of 2024, regulations are emerging in most developed countries or jurisdictions, including:

The EU AI Act

This legislation was approved on May 21, 2024, and will go into effect in the second half of 2024 or 2025. It defines four categories of AI risk and creates rules for each category. It emphasizes transparency, accountability, and oversight and is particularly concerned with AI use cases that affect critical infrastructure, systems, employment, and law enforcement.

US Executive Order

This order applies to US executive branch agencies controlled by the President of the United States. It aims to ensure the safe, secure, and trustworthy development and use of AI through guidelines, reporting requirements, risk assessments, and other measures across critical areas like cybersecurity and biosecurity.

The Canadian Directive on Automated Decision-Making, issued in 2019

Sets out rules and procedures for using automated decision systems in the federal government. The directive mandates the assessment and mitigation of risks and biases associated with ADS, as well as the monitoring and evaluation of their performance and outcomes.



Primary areas of AI regulation

While each country has different priorities for how AI should be regulated, there are several key themes, many of which mirror industry guidelines and best practices.

Some of the most common areas of concern include:

- **Data protection and privacy**
Protecting personal identity, privacy, healthcare, and financial information is critical and a principle pillar of AI governance frameworks and regulations.
- **Bias, discrimination, and AI ethics**
AI training can inadvertently amplify societal bias and inequity. Providers are working to mitigate these risks, but careful control of bias in training and outputs is essential.
- **Biometric markers, including facial recognition**
In some jurisdictions, facial recognition and other passive bio-identifiers are prohibited without explicit subject opt-in. Noncompliance with these requirements leads to fines or civil liability. It is crucial to understand in which jurisdiction such AI can be used and to carefully record and match who has opted in.
- **Content ownership/copyright**
AI systems can quickly consume, learn, and replicate content as a whole or in part. Organizations can be liable if they unknowingly publish a copyrighted work through AI. Likewise, the Copyright Office of the US Library of Congress has stated that AI output without additional human authorship doesn't enjoy the protection of copyright.²
- **Accountability**
Organizations and users must remain accountable for their use of AI in both decision-making and content generation. In the most critical decisions, human users, operators, and providers are collectively accountable for AI output.
- **Critical infrastructure**
Protecting critical infrastructure, such as the electric grid, nuclear power generation, public transportation, emergency and first responder services, and disaster response, is a crucial area of concern. AI systems that could affect these services must be well documented, and risk mitigation procedures must be implemented.

² U.S. Copyright Office, Library of Congress. (2023). Copyright registration guidance: Works containing material generated by artificial intelligence. Federal Register, 88(FR 16190), 16190-16194.

Arming GenAI with your most trusted data, responsibly and correctly curated, is the most important factor for trustworthy GenAI in your control

Trustworthy GenAI depends on strong content management and governance

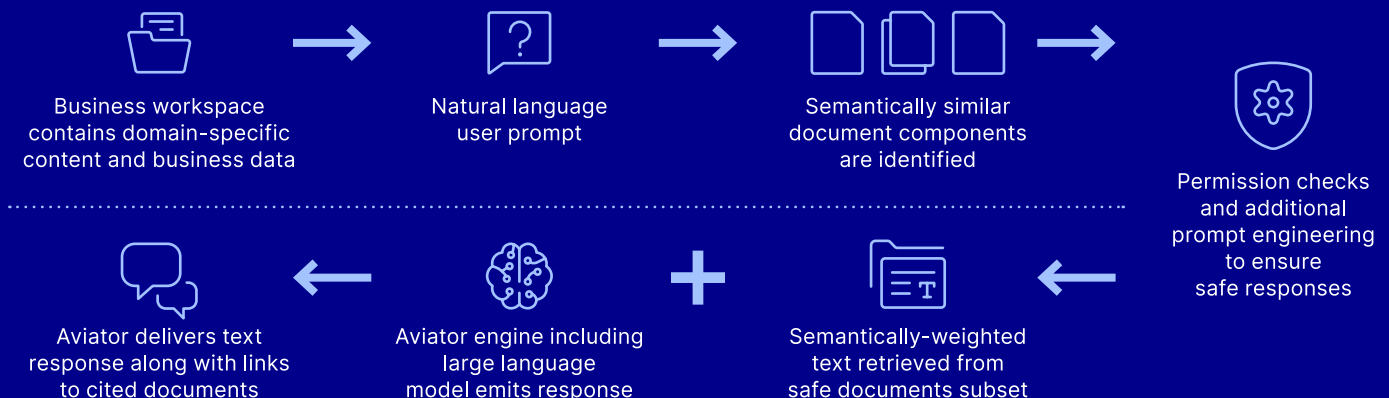
GenAI can help us be more productive while remaining a trustworthy, intelligent assistant during our workday and processes. To be effective, GenAI should provide us with high-quality, accurate responses to our requests along with proof points or citations for verification. Occasionally, however, inaccurate responses occur and can appear reasonable.

Several factors affect GenAI accuracy, and most of them are within your control if you have the right toolset behind you as you go:

- **GenAI model selection**
Choosing the correct base model is vital as it understands language and semantics across all supported languages. Its ability to produce accurate responses depends on key inputs like the model size, context window size, training data selection and accuracy, and other factors such as “human in the loop/HITL” training.
- **Prompt engineering**
Receiving good responses depends on the quality of the request or “prompt” the AI is fed. Prompt engineering involves crafting clear instructions to help the model infer the desired result. It is also essential for AI safety, filtering out potentially harmful or offensive results. In internal GenAI processes, prompt engineering enhances user prompts with necessary instructions.
- **Trustworthy grounding data**
GenAI responses almost always draw from the grounding data you provide. If that data is inaccurate, outdated, or full of noisy or irrelevant data, your response will suffer.

How OpenText™ Content Aviator processes a query

Relevant and user-safe content from business workspace is combined along with large language model to generate response



Great GenAI needs great content management

Getting grounded with excellent content management

When a prompt is submitted, either

- The prompt is processed to extract key concepts and combined with user-provided content such as a document or an image
- Content is searched using the keywords and other hints and relevant content is retrieved.

This data is combined with one or more models and the prompt to produce a result. This process is called retrieval-augmented generation, and it's the basis of virtually every GenAI implementation available.

To achieve success with GenAI, we want to start with the most relevant content available to the user. To formulate this candidate list of content, we must answer several key questions:

- **Does the user have access to the content?**
Retrieval-augmented generation should only have access to content approved for the user through access control and system permissions. This is critical for security.
- **Is it the current version of the content?**
Older drafts may have non-factual data, other mistakes, or concerns that would be inappropriate. Content sprawl can litter the grounding data with outdated content.
- **Is it approved and permissible content for GenAI?**
Some content may contain trivial or obsolete information, be prohibited by policy or statute, or be out of context.
- **Does it meet the prompt's restrictions?**
In some more advanced cases, the prompt itself may specify the types or domains of content that are appropriate for the prompt. For example, when answering tax filing questions, we may want to look only at filing documents rather than the entire tax file.

[Great GenAI, then, needs great content management.](#) All aspects of content management come into play when developing a source of acceptable grounding data for AI. These include:

- **Access controls** to determine who can see what content.
- **Security markings** protect whole classifications of groups of content, such as confidential or proprietary information.
- **Version control** to select the most recent content automatically.
- **Classification and other metadata** to identify suitable document types.
- **Retention management** to prevent expired or obsolete content from showing up.
- **Business context** to tie the content to a relevant context, such as a project.



The degree to which a content management system can perform these tasks goes beyond having these individual features. Understanding the relationship between the content, policies, workflows, and other automated systems will help assure better control and predictability in how GenAI performs.

For example, while providing excellent document collaboration tools is important, it's also important to properly classify and contextualize the content developed during collaboration through integration with platforms such as Microsoft® Teams or line-of-business applications such as SAP®, Salesforce® or Microsoft Dynamics 365®.

A fully integrated desktop and process for the user enhances every aspect of content management, including access control, security, versioning, classification, retention, and context. Make sure you are making the most of content management in your organization!

The basics of GenAI “context”

One of the key elements of GenAI and LLMs is the “context window.” This refers to the amount of text that can be processed simultaneously. This typically includes the following, across an entire conversation or multiple chat prompts:

- **Prompt text**
This is the user’s submitted prompt, in addition to templated additions to the prompt that might represent organizational instructions on acceptable use, such as filtering out PII and avoiding NSFW content.
- **Source material (grounding data)**
Any grounding data retrieved by the GenAI implementation counts towards the context window.
- **The entire response**
Any response text is appended to the context window, so follow-up prompts form a whole, contextual back-and-forth chat.

The available context window is measured in “tokens,” equivalent to approximately $\frac{3}{4}$ of a word per token. Various models can handle anywhere from 128K to 1M tokens per conversation and this is increasing as GenAI advances. Depending on the content, 1M tokens may represent between 1,500-5,000 pages of text across a single GenAI conversation.

While it might seem like more is better, “casting a wide net” doesn’t necessarily improve results. What we actually want is the right context. It does no good to include all versions of a document, for example as this will either create noise that results in a confused or inaccurate response.

OpenText Content Aviator

OpenText Content Aviator is the resident GenAI for all OpenText content management systems, including OpenText™ Core Content Management, OpenText™ Content Management (formerly OpenText™ Extended ECM), OpenText™ Documentum™ Content Management (formerly OpenText™ Documentum™), OpenText™ Core Archive for SAP® Solutions, and other repositories.



It combines a world-class large language model—Google’s Gemini™—with a richly integrated user experience and services to give users instant access to GenAI chat responses and other services. Content Aviator honors all content security and depends solely on the grounding data in the customer’s content management system. This provides safe, trustworthy GenAI capability to every content management user.

How Content Aviator determines context

Content Aviator takes a user-centered approach to determining context and starts by inferring the appropriate scope from what the user is currently viewing. This might be any of the following:

Current document	Business workspace(s)	Search results	Business process
If the user is viewing content, immediate and integrated Q&A is displayed	If the user selects one or more workspaces, matching documents from that workspace are used	Depending on the application and context, search results can be used as grounding context	When Aviator is embedded in an action or workflow, the context is determined by that process

This is the starting point. In some instances, such as one or more business workspaces, the prompt is used to determine a smaller set of relevant documents, giving better overall refinement of the grounding context. If no grounding data can be identified, Content Aviator will let the user know that it cannot make an inference or provide a sensible response.

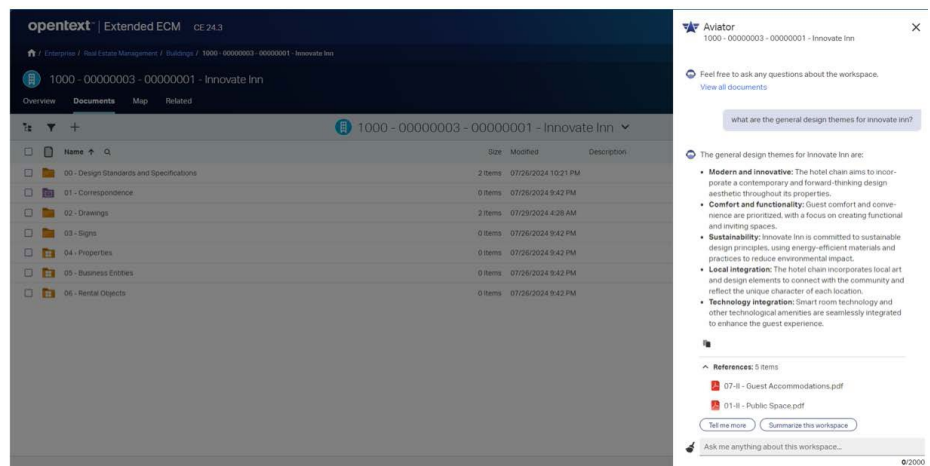
The power of the OpenText “business workspace”

A primary advantage of content management is the ability to organize and control information through the addition of containers and metadata. However, simple folders and tagging document types are not enough. So many elements must be coordinated, from access control to retention. Ultimately, this must all fit together in a sensible business process. Given the volume of information that must be organized, automation is essential. What is the context of the content within your process?

OpenText connects content to process through the power of the business workspace. Connected content combines documents with line-of-business application data in one place, with a familiar and easy-to-use user interface.

A business workspace is an easily accessible, familiar, and consistent way to manage content for all your critical content. For instance, if a customer calls about a missed shipment, you can find information about them without leaving your familiar Microsoft® Office 365® interface. Customer data is drawn from a CRM system, and the account team information is available from your Microsoft Teams app. This allows you to identify the sales account team member responsible for supporting or selling to the customer.

Trustworthy AI responses are grounded in content managed by the business workspace. Policies, content labeling, and access control are all controlled from centralized governance administration.



The business workspace centralizes governance, security, and classification to provide a point of consistency for safe grounding data

All these integrations, with Microsoft 365, CRM, ERP, HR, and other systems, drive important labeling and organization of content and metadata. This is done through a library of common business scenarios attached to business rules, security requirements, records policies, and more.

The business workspace also unifies how teams work together, assuring that users who spend a lot of time in various applications all have a unified view and organization of content. They always have the latest version of the document at their fingertips because the business workspace is integrated and embedded within applications, such as Microsoft Teams, Outlook, SAP, Salesforce, and Microsoft Dynamics.

The business workspace helps to reduce content sprawl and confusion, automatically applying content security and metadata consistently through automation. It is the central hub from which information governance policies, security, and classification of information are driven. The business workspace is integral to Content Aviator as it forms the central role in providing GenAI context, focusing the natural language interaction on the topic at hand, improving response quality and keeping content secure.

Additional automation tools to help keep GenAI secure and accurate and keep users productive

Content management incorporates automation capabilities across the entire information lifecycle. The following content services prepare and control content in a way that directly improves the GenAI experience:

- **Content analytics/file analysis**
Scanning content for PII, financial data, healthcare data, and other risky content is essential. Content analytics can quickly scan your legacy content to determine where the risk lies, let you review and tag it, and then move it into a secure content management repository, complete with appropriate metadata. GenAI then proactively knows which documents to avoid to prevent data leaks or privacy concerns ahead of time.

Resources

[OpenText Content Aviator >](#)

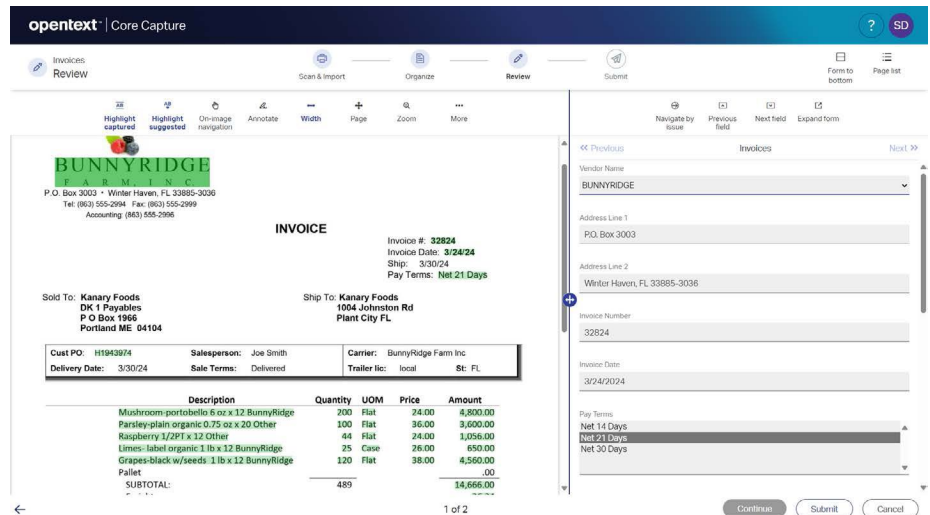
[Content Aviator Interactive Tour >](#)

[Trustworthy, secure GenAI starts with strong governance >](#)

[Mitigating Risk with Voltage Fusion Data Security Platform >](#)

- **Advanced records management**

Data privacy regulations often require that personally identifiable content is removed as soon as possible (except when under legal hold). Expired content should be disposed of to avoid unnecessary disclosure and to keep noisy, legacy content from interfering with everyday operations.



Capture technologies can extract key metadata from incoming forms and other documents, automatically labeling content

- **Intelligent document processing**

Accurate and detailed metadata labeling has never been more critical. When capturing large volumes of external documents, making sense of those documents within your business processes is vital. As you incorporate GenAI into your user's workday and active processes, this additional metadata—intelligently recognized within handwritten forms, images, and other specialized document types—can be a real help in identifying grounding data for GenAI.

Summary

GenAI transforms how knowledge workers relate to information. It gives users instant command of large bodies of information to produce summaries, extract quick facts and insight, and draft and create new work product. GenAI also introduces a variety of new areas of risk, all of which result from the critical nature of how content is created, managed, and consumed in everyday work.

GenAI is meant to be easy to use, but it must remain accurate, safe, and secure while honoring privacy. Organizations need to implement a AI governance framework of accountability, transparency, and oversight. The core toolset in providing AI governance for the organization is a foundation of strong, governed content.