DATA CENTURY

Gen-AI Readiness: Unlocking Unstructured Data

2026 PLAYBOOK

Table
Of Contents

Introduction

Most enterprises are sitting on a goldmine of unstructured content (policies, PPTs, PDFs, emails, knowledge articles, call notes, transcripts, images, and operational logs) yet only a small fraction is searchable, trustworthy, or re-usable. Historically, this information was viewed as a compliance and security liability rather than a strategic asset.

GenAI fundamentally changes the equation. With the right foundations, unstructured data can shift from a cost center to a compound asset that accelerates revenue, shortens cycle times, strengthens risk and compliance, and improves decision-making across the enterprise.

This playbook outlines

- What has changed in the market
- High-value use cases across financial services, professional services, and healthcare
- A pragmatic checklist for building a production-grade GenAI stack for unstructured data, grounded in proven best practices across ingestion, retrieval, generation, evaluation, and governance

DATA
CENTURY

Market Context

From Protect ≫ to **Monetize**

Unstructured data dominates the enterprise

A**n estimated 80-90% of enterprise data is unstructured,** yet most of it has traditionally been governed under risk frameworks (DLP, retention, encryption, rights management) rather than engineered for insight, reuse, or automation.

80%- 90%

Why this moment is different

Trust & Scale

The following advancements make it feasible to trust and scale AI on unstructured data with far greater confidence.

- GenAI (for reasoning and summarization)
- AI-Augmented Search and Retrieval (for grounding models in authoritative content)
- Modern observability + evaluation pipelines (for measuring accuracy, drift, and relevance)

Why it matters

Organizations can now extract structured value from documents and flows that were previously opaque – credit files, KYC documents, RCSA narratives, policies, SOPs, trade finance documents, regulatory updates, analytics knowledge bases, and more.

A New Operating Model

Unstructured data is shifting from an unmanaged repository ("dump and search") to a managed product that has increased in value with the advent of GenAI, with:

- SLAs for freshness and quality
- Automated enrichment and tagging
- Usage-aware ranking
- Version awareness and archival
- Evaluation and grounding standards
- Auditable lineage

This shift enables organizations to unlock monetization opportunities across revenue, operations, risk, compliance, and client experience.

Monetizing Unstructured Data

Representative Use Cases

The use cases in this section show where GenAI can deliver real, near-term impact by unlocking the value trapped in unstructured data. They highlight the specific pain points (i.e., manual reviews, fragmented knowledge, slow decision cycles) that GenAI can meaningfully reduce, while strengthening accuracy, compliance, and client experience.

Together, these examples outline a clear path for where GenAI can be deployed first, where returns are most tangible, and where organizations can build lasting capability.

Across three industries

Financial Services

Professional Services

Healthcare

Financial Services

Client Onboarding

Beyond KYC

Banks process large volumes of onboarding documents (IDs, corporate registries, ownership charts, financial statements) many of which are received as PDFs or scans. GenAI and layout-aware parsing models can automatically extract key fields, cross-validate them against internal systems, and identify inconsistencies. Process-mining analytics can highlight bottlenecks in onboarding flows, while SLA prediction models identify cases likely to breach timelines.

Impact

Faster activation, fewer manual reviews, reduced error rates, and higher-quality client data from day one.

Examples of Unstructured Data:

- Onboarding forms
- Scanned IDs
- Incorporation documents
- Credit reports
- Ultimate Beneficial Owner (UBO) declarations
- Internal notes
- Client deliverables

Payments STP Optimization

Wire and ACH processing remains burdened by exceptions due to incomplete fields, formatting inconsistencies, incorrect codes, or missing beneficiary data. AI-enabled pre-validation engines can predict high-risk transactions before they fail, enrich missing fields, and recommend fixes. Over time, self-learning feedback loops improve both accuracy and speed, reducing intervention load for payment operations teams.

Strong guardrails must be put in place to mitigate against malicious actors trying to trick AI systems to clear payments that shouldn't otherwise be cleared.

Impact

- 25–35 percentage-point uplift in Straight-Through-Processing (STP)
- Reduction in repair queues
- Improved settlement timelines
- Material operational cost savings.

Examples of Unstructured Data

- Payment instruction PDFs & email attachments
- Free-text remittance information
- Bank-provided and client-provided wire forms
- Client onboarding documents that influence payment metadata

RCSA Modernization

Enterprise risk teams spend substantial time drafting RCSA narratives, updating evidence, and documenting control effectiveness. GenAI can automate the creation of risk descriptions, rationales, and mitigations by synthesizing authoritative content from policies, historical assessments, issue logs, and regulatory requirements. Automated evidence retrieval further reduces manual effort and inconsistency.

Impact

- Standardized assessments
- Reduced cycle times
- More defensible control rationales
- Improved audit readiness.

Examples of Unstructured Data

- Control narratives & risk descriptions
- Internal audit reports and observations
- Issue logs and remediation plans
- Documents containing control test results
- Policy and procedure manuals
- Compliance guidance & regulatory memos
- Historical RCSA submissions across PDF, PPT, and Excel formats, etc.

Regulatory Document Intelligence

Financial regulators issue frequent updates, guidelines, and interpretations that must be understood, categorized, and mapped to existing policies and controls. GenAI can parse regulatory texts, extract obligations, highlight gaps, and surface relevant internal documents. Version-aware retrieval ensures that teams always reference the newest authoritative content.

Impact

- Faster regulatory absorption
- Enhanced defensibility
- Reduced compliance risk.

Examples of Unstructured Data

- Regulatory texts and supervisory notices
- Interpretive letters and regulator FAQs
- Internal legal assessments and compliance interpretation
- Policy manuals and control libraries
- Evidence documents submitted during exams or reviews
- Industry guidance or association bulletins, etc.

Professional Services & Knowledge-Heavy Functions

GenAI Knowledge Management

Professional services teams rely on large volumes of proposals, methodologies, deliverables, research, and client documentation. A retrieval-augmented assistant tool can immediately surface relevant content, synthesize answers with citations, and summarize long materials into actionable guidance. This type of tooling can also be used to help teams understand who in the organizations are a good fit for specific opportunities by connecting knowledge management systems to talents systems. Domain lexicons help the system understand firm-specific terminology, while LLM-as-judge scoring ensures factual accuracy and policy alignment.

Impact

- Productivity uplift
- Accelerated onboarding
- Reduced rework
- Better reuse of institutional knowledge

Examples of Unstructured Data

- Past deliverables (PowerPoints, reports, PDFs)
- Internal methodologies and frameworks
- Sales proposals and RFP responses
- Research memos and industry analyses
- Knowledge articles
- SME interview notes & transcripts
- Working papers or appendices attached to client engagements

Healthcare Analytics & Operations

Conversational Analytics Interface

Healthcare operations teams often wait days or weeks for analytics teams to produce updated dashboards or reports. GenAI-powered conversational analytics turns natural-language questions into structured SQL queries, executes them under strict governance, and returns answers as charts, tables, or narratives with links to source data.

Impact

- Substantial reduction in analytics turnaround times
- Improved decision-making,
- Increased capacity for data teams—without compromising security or trust.

Examples of Unstructured Data

- Lab and diagnostic reports
- Operational dashboards exported as images or PDFs
- Quality and safety review documents
- Narrative sections of claims or utilization reports

How to Start

The GenAI Readiness & Build Checklist

This checklist sequences the actions needed to build a scalable GenAI capability for an illustrative use case in Professional Services with unstructured data.

It is intentionally pragmatic and execution-focused, giving leaders a practical roadmap to build momentum quickly while establishing the foundations for scale.

01

Align the Enterprise

Define business outcomes and select 3–5 high-impact use cases with measurable KPIs.

Identify and align business, data, technology, risk, compliance, and security leaders.

Establish trust requirements: grounding expectations, acceptable hallucination risk, citation behavior, and audit needs.

02

Get Data Ready: Data Lifecycle Management and Protections

Confirm data ownership, classification, and lineage for unstructured sources (documents, slides, PDFs, email attachments, shared drive content).

Align sensitivity labels, encryption standards, and data-handling requirements with enterprise security and privacy policies. Review retention schedules and archival rules to ensure only appropriate content is ingested into AI workflows.

Establish "safe zones" or controlled environments for GenAI experimentation (with monitored access and audit logs).

Coordinate with Information Security, Privacy, and Data Governance teams to validate readiness and reduce regulatory risk before expanding to production use cases.

03

Establish AI Foundations

- Implement automated evaluations for fidelity, relevance, and safety – using production and synthetic datasets.
- Track experiments with versioned corpora, embeddings, and prompts for full reproducibility.
- Implement governance controls: sensitivity labels, access enforcement, audit trails, and model approval workflows.

04

Fix the Content Supply Chain

- Build telemetry across the RAG stack (latency, hit/miss, similarity, fallback logic, token usage) to track against the KPIs defined in Step A "Align the Enterprise".
- Implement automated ingestion pipelines: cleansing, PII removal, tagging, duplicate detection, and exception routing.
- Preserve structure with layout-aware parsing for PPTs, tables, diagrams, and PDFs to maximize grounding accuracy.
- Apply metadata discipline: maintain helpful filters; use domain lexicons; incorporate recency.
- Deploy curation workflows: relevance gating, aging and archival, content-gap analysis, and usage-aware ranking.

05

Build the Retrieval & Generation Experience

- Use hybrid search (lexical + vector) with principled boosts and deterministic fallbacks.
- Build AI-powered research assistants with citation-first answers, structured summaries, and deep-linking to authoritative content.
- Implement guardrails: prompt scoping, grounding requirements, structured logging, and cost/rate controls.

06

Deliver Flagship Use Cases

- **Illustrative Use Case:** Knowledge assistant in Professional Services with evaluation-driven improvement loops.
- Operate with a rhythm: backlog → experiment → evaluate → ship → observe → retrain.
- Publish monthly reports on answer quality, retrieval recall, latency, and usage.

07

Operate & Improve

- **Maintain Service Level Objectives (SLOs) for answer quality and latency with clear runbooks for incident response.**
- **Monitor drift across models, prompts, embeddings, and corpora with automated alerts.**
- **Control Total Cost of Ownership (TCO) through caching, prompt optimization, and retrieval tuning.**

DATA CENTURY

Closing Thoughts & Where to Go Next

GenAI gives enterprises the ability to convert unstructured data into a high-leverage strategic asset. With strong foundations (evaluation, observability, content engineering, and governance) organizations can move quickly and safely.

Organizations that succeed:

Get their data ready

Establish lifecycle management and protection controls so that unstructured data can be safely exposed to GenAI systems without regulatory or privacy risk. This step converts data from a compliance liability into an innovation-ready asset.

Build telemetry & automated evaluations

Implement real-time monitoring of retrieval quality, answer accuracy, and cost metrics. This ensures teams can measure and continually improve AI performance, rather than relying on anecdotal feedback.

Stabilize the content supply chain

Standardize ingestion, tagging, and curation workflows. This creates a single, trusted corpus for all downstream AI use cases.

Deliver one flagship use case per quarter

Produce iteratively to experiment, evaluate, and retrain. This builds organizational confidence while demonstrating tangible ROI.

Data Century works with clients to provide evaluation frameworks, dashboard templates, and a 90-day "get ready" plan tailored to their environment and risk posture.

DATA CENTURY

Contact Us

161 Bay Street, 27th Floor
Toronto, ON, Canada
M5J 2S1

info@datacentury.ca
www.datacentury.ca