International Journal of Advanced Research & Higher Studies (IJARHS)

DOI: https://doi.org/10.70818/ijarhs.v04i03.2025.250315

Business Intelligence in the Age of Enterprise AI: Assessing Readiness for Unstructured Data Utilization

Abul Fattah Mohammad Al Mohaimin[1], Mushfiqur Rahman[2], Dr. Rubaiyat -E-Mehnaz[3], Kaium Siddik Anando[4], Sazzad Hossain, PhD[5]

[1]CEO, Link & Win. LLC, Data and AI Leader, MIT Certificated AI Executive Leader, Google Certificated Professional Machine Learning Engineer, USA

[2]Department of Computer Science and Engineering, Bangladesh University of Professional, Dhaka, Bangladesh

[3]MBBS, MRCP (UK), CCD (Birdem) MACP (USA), Former Junior Consultant, Birdem general hospital, Dhaka, Bangladesh

[4]Sr Consultant, Link&win. LLC, Former Consultant, Ernst & Young LLP. (EY), Dhaka, Bangladesh

[5]Professor, System Management & Information Security Department, Faculty of Artificial Intelligence & Digital Technologies, Samarkand State University named after Sharof Rashidov (Uzbekistan)

ABSTRACT

The rapid advancement of enterprise artificial intelligence (AI), particularly generative AI, has fundamentally reshaped the scope and expectations of business intelligence (BI). While organizations increasingly invest in advanced AI technologies, many initiatives fail to achieve sustainable value. A critical yet often overlooked factor is organizational readiness to manage and utilize unstructured data, which constitutes approximately 97% of enterprise data assets. Traditional BI architectures, optimized for structured data, are ill-equipped to support AI systems that rely on natural language, documents, images, and contextual knowledge. This article examines unstructured data readiness as a foundational requirement for AI-enabled BI. Drawing upon enterprise transformation experience and a detailed healthcare case study, the paper proposes a five-dimensional readiness framework encompassing data discovery, data quality and metadata, access and security architecture, infrastructure and processing capabilities, and retrieval-augmented generation (RAG) readiness. The findings demonstrate that enterprises addressing these dimensions systematically can achieve measurable productivity gains, enhanced decision-making, and faster return on investment. The article concludes with a practical readiness roadmap, common implementation pitfalls, and a business case framework to guide strategic adoption of AI-driven business intelligence.

Keyword: Business Intelligence, Enterprise AI, Unstructured Data, AI Readiness, Data Governance, Retrieval-Augmented Generation

1. Introduction

Across global enterprises, the adoption of artificial intelligence has accelerated dramatically. Executives approve generative AI initiatives with the expectation of transforming decision-making, enhancing productivity, and unlocking competitive advantage. Yet, despite strong executive sponsorship and increasing technology budgets, many AI initiatives stall or fail to scale beyond pilot deployments. The root cause is rarely the sophistication of AI models

themselves. Instead, failure often stems from inadequate readiness to manage and exploit unstructured data the dominant form of enterprise information[1].

Traditional business intelligence evolved around structured data sources such as relational databases, transactional systems, and enterprise data warehouses. These systems support predefined queries, dashboards, and reports that rely on normalized schemas and well-governed data models. However, generative AI and modern machine learning systems operate on fundamentally different inputs. Large language models (LLMs) process text, documents, images, and contextual narratives rather than SQL tables. Consequently, the most valuable organizational knowledge strategic plans, clinical notes, customer communications, policies, and product documentation resides outside traditional BI infrastructures[2].

Industry research indicates that more than 90% of data and AI leaders have increased investment in data management due to AI initiatives[3]. This shift reflects growing recognition that AI success depends less on algorithms and more on the quality, accessibility, and governance of underlying data assets. In this context, unstructured data readiness emerges as a strategic determinant of AI-driven business intelligence.

This article addresses the following research objective:

- To develop and operationalize a structured framework for assessing enterprise readiness to leverage unstructured data for AI-enabled business intelligence.

2. Literature Review

2.1 Business Intelligence Evolution

Business intelligence has historically focused on structured data analytics, emphasizing reporting, performance measurement, and decision support[4]. Early BI systems relied on extract-transform-load (ETL) pipelines feeding centralized data warehouses. These systems excelled at descriptive and diagnostic analytics but were limited in handling ambiguity, natural language, and contextual reasoning.

2.2 AI and Data Management

Recent research highlights the growing importance of AI governance, explainability, and ethical data usage[5]. However, much of the literature assumes the availability of clean, structured datasets. As enterprises shift toward generative AI, this assumption no longer holds. Studies increasingly recognize unstructured data as both a critical asset and a major bottleneck[6].

2.3 Retrieval-Augmented Generation

Retrieval-augmented generation has emerged as a dominant pattern for enterprise AI deployment. Rather than fine-tuning models on proprietary data which is costly, risky, and difficult to govern RAG systems retrieve relevant documents and inject them into model prompts[7]. This approach depends heavily on robust unstructured data pipelines, metadata, and vector search capabilities.

Despite growing recognition of these trends, a comprehensive readiness framework integrating technical, organizational, and governance dimensions remains underdeveloped. This study seeks to fill that gap.

3. Conceptual Framework: Unstructured Data Readiness

3.1 Rationale for the Framework

As enterprises increasingly adopt generative and analytical AI systems, it has become evident that traditional business intelligence maturity models are insufficient. Most existing BI and analytics frameworks focus on structured data pipelines, reporting maturity, and decision-support capabilities. However, enterprise AI systems particularly large language models and multimodal models derive their value primarily from unstructured data such as documents, emails, images, audio, and video.

Unstructured data readiness represents the organization's ability to systematically discover, govern, process, and operationalize such data for AI-driven insight generation. Without this readiness, AI initiatives suffer from poor output quality, security risks, lack of user trust, and limited scalability. Therefore, a dedicated conceptual framework is required to evaluate readiness across technical, organizational, and governance dimensions.

The proposed framework conceptualizes unstructured data readiness as a multi-dimensional construct that directly enables AI-driven business intelligence outcomes. It emphasizes that readiness is not achieved through a single technology investment but through the coordinated development of foundational capabilities.

3.2 Overview of the Five-Dimensional Framework

The Unstructured Data Readiness Framework consists of five interdependent dimensions:

- Data Discovery and Cataloging
- Data Quality and Metadata
- Access and Security Architecture
- Infrastructure and Processing Capabilities
- Retrieval-Augmented Generation (RAG) Readiness

These dimensions collectively support AI-Driven Business Intelligence, while being governed by cross-cutting mechanisms such as data governance, compliance, and human-in-the-loop oversight.

The framework assumes that weaknesses in any single dimension can significantly reduce overall AI effectiveness, even if other dimensions are mature. Thus, readiness should be assessed holistically rather than through isolated technical evaluations.

3.3 Dimension 1: Data Discovery and Cataloging

Data discovery and cataloging form the foundation of unstructured data readiness. Organizations cannot leverage data assets they cannot see or understand. Unlike structured systems, unstructured data is often scattered across collaboration platforms, personal drives, email systems, legacy archives, and departmental repositories.

- This dimension focuses on answering fundamental questions:
- Where does unstructured data reside?
- What types of content exist (documents, images, audio, etc.)?
- What is the volume, growth rate, and usage pattern?
- Which data assets have the highest business and analytical value?

Effective discovery enables organizations to move from anecdotal assumptions to evidence-based prioritization. Cataloging tools and metadata inventories transform opaque data landscapes into manageable assets, enabling downstream quality, security, and AI processing initiatives.

3.4 Dimension 2: Data Quality and Metadata

Data quality in unstructured environments extends beyond accuracy and completeness. It encompasses accessibility, semantic coherence, and contextual richness. AI systems are particularly sensitive to poor-quality inputs, as irrelevant or low-confidence content can propagate errors across generated outputs.

This dimension evaluates:

- Extractability of content (e.g., OCR confidence for scanned documents)
- Completeness and consistency of metadata (author, date, version, document type)
- Semantic relevance and noise levels
- Document integrity and format usability

Metadata plays a critical enabling role by providing contextual signals that enhance retrieval, ranking, and filtering in AI systems. Organizations with mature metadata practices achieve significantly higher AI relevance and trustworthiness.

3.5 Dimension 3: Access and Security Architecture

Unstructured data frequently contains mixed-sensitivity information within a single asset, making traditional role-based access control insufficient. AI systems further complicate this challenge by acting as intermediaries that retrieve, summarize, and recombine information.

This dimension addresses:

- Automated content classification and labeling
- Attribute-based and context-aware access control
- Dynamic redaction of sensitive elements
- Auditability and traceability of AI access

Security maturity ensures that AI systems can safely consume enterprise data without violating regulatory requirements or internal policies. In this framework, security is not treated as a constraint on AI innovation but as an enabling condition for responsible deployment.

3.6 Dimension 4: Infrastructure and Processing Capabilities

AI-driven business intelligence places unique demands on infrastructure. Traditional BI architectures optimized for batch reporting and structured queries cannot efficiently handle large-scale document ingestion, real-time updates, or vector-based retrieval.

This dimension evaluates the organization's ability to:

- Store unstructured data cost-effectively at scale
- Process content through extraction, transformation, and enrichment pipelines
- Support real-time and event-driven processing
- Optimize computational costs for AI workloads

Infrastructure maturity enables scalability, performance, and sustainability, ensuring that AI solutions remain viable as data volumes and usage grow.

3.7 Dimension 5: Retrieval-Augmented Generation (RAG) Readiness

RAG readiness represents the highest level of unstructured data maturity. It reflects the organization's ability to operationalize unstructured data directly within AI applications.

This dimension includes:

- Effective chunking strategies aligned with content semantics
- Appropriate embedding model selection (general or domain-specific)
- Optimized vector search and hybrid retrieval mechanisms

- Continuous evaluation and feedback loops

RAG readiness ensures that AI outputs are grounded in authoritative enterprise knowledge, improving accuracy, explainability, and user trust.

3.8 Cross-Cutting Governance and Human Oversight

While presented as discrete dimensions, the framework recognizes governance and human oversight as cross-cutting enablers. Policies for data ownership, compliance review, ethical AI use, and feedback mechanisms ensure alignment between technical capabilities and organizational values.

Human-in-the-loop mechanisms reinforce accountability, particularly in high-stakes domains such as healthcare, finance, and legal decision-making.

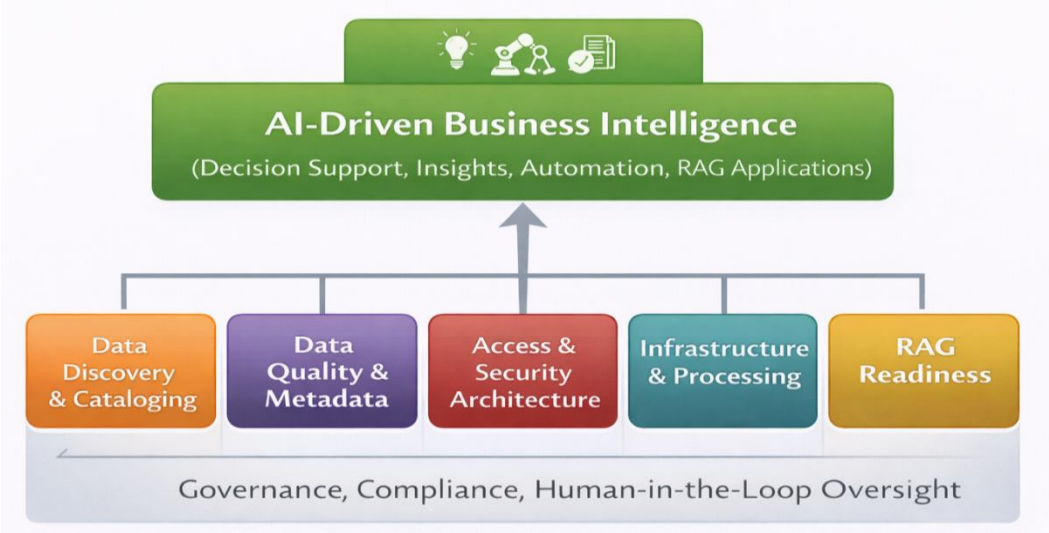

Figure 1. Conceptual Framework of Unstructured Data Readiness

Figure 1 presents a layered conceptual framework illustrating how unstructured data readiness enables AI-driven business intelligence. The five foundational dimensions—data discovery and cataloging, data quality and metadata, access and security architecture, infrastructure and processing capabilities, and retrieval-augmented generation readiness—form the structural base of the model. These dimensions collectively support AI-driven BI outcomes such as intelligent decision support, knowledge retrieval, automation, and contextual insight generation. Governance, compliance, and human-in-the-loop oversight span all layers of the framework, emphasizing that unstructured data readiness is both a technical and organizational capability. The framework highlights the interdependence of dimensions, demonstrating that deficiencies in any foundational layer can undermine AI performance, trust, and scalability.

4. Methodology

4.1 Research Design

This study adopts a **qualitative, practice-based research design** to examine enterprise readiness for unstructured data utilization within AI-driven business intelligence environments. Qualitative methodologies are particularly appropriate for investigating complex socio-technical phenomena where organizational processes, governance mechanisms, and contextual factors play a decisive role[1].

Practice-based research emphasizes insights derived from real-world implementations rather than controlled experimental conditions, allowing the study to capture how enterprise AI readiness evolves through actual deployment experiences[2]. This design supports theory development through empirical observation and aligns with prior research on digital transformation and enterprise analytics maturity[3,4].

4.2 Data Sources

To ensure analytical depth and triangulation, the study integrates three complementary qualitative data sources.

4.2.1 Enterprise AI Readiness Assessments

Enterprise AI readiness assessments were conducted across multiple industries, including healthcare, financial services, manufacturing, and professional services. Readiness assessments are widely recognized as effective instruments for evaluating organizational preparedness for advanced analytics and AI adoption[5].

Each assessment examined:

- Availability and discoverability of unstructured data assets
- Data governance and security controls
- Analytical infrastructure and tooling maturity
- Organizational capabilities for AI-enabled BI

These assessments enabled cross-industry comparison and identification of recurring readiness gaps.

4.2.2 Healthcare Case Study

A detailed healthcare sector case study was selected due to the sector's high reliance on unstructured data (e.g., clinical notes, medical imaging, reports) and strict regulatory constraints. Case study research is particularly suitable for in-depth investigation of contemporary phenomena within real-life organizational contexts[6].
The case study documents a multi-phase enterprise AI initiative involving:

- Unstructured data ingestion and classification
- Secure AI model deployment
- Integration of AI outputs into BI dashboards
- Post-deployment performance evaluation

4.2.3 Observational Insights from BI and AI Implementations

Additional observational data were drawn from large-scale BI and AI implementation initiatives. Observational research allows researchers to capture tacit knowledge, emergent practices, and implementation dynamics that may not be formally documented[7].

These observations contributed longitudinal insights into how readiness dimensions change over time as enterprises scale AI-enabled BI solutions.

4.3 Data Collection Procedures

Data collection followed an iterative and multi-method approach, consistent with qualitative enterprise research best practices [8]. Methods included:

- Semi-structured discussions with business and technical stakeholders
- Review of architectural documents, governance frameworks, and implementation roadmaps

- Direct observation of system deployment and user interaction
- Retrospective analysis of project outcomes and performance reports

This iterative process enabled continuous refinement of analytical categories.

4.4 Data Analysis Approach

Data were analyzed using **thematic synthesis**, a method well suited for integrating qualitative evidence from diverse sources while maintaining analytical rigor [9]. Thematic synthesis has been widely applied in organizational and information systems research to identify cross-contextual patterns [10].

The analysis proceeded through four stages:

- **Open Coding:** Identification of key concepts related to readiness, governance, infrastructure, and AI utilization
- **Axial Coding:** Grouping codes into higher-level readiness dimensions
- **Cross-Case Synthesis:** Comparing patterns across industries and the healthcare case study
- **Framework Integration:** Refining the conceptual framework presented in Section 3

This process ensured that the framework emerged inductively from empirical evidence.

4.5 Research Process Diagram

Figure 2 illustrates the overall research process adopted in this study, highlighting the iterative flow from data collection to framework development.

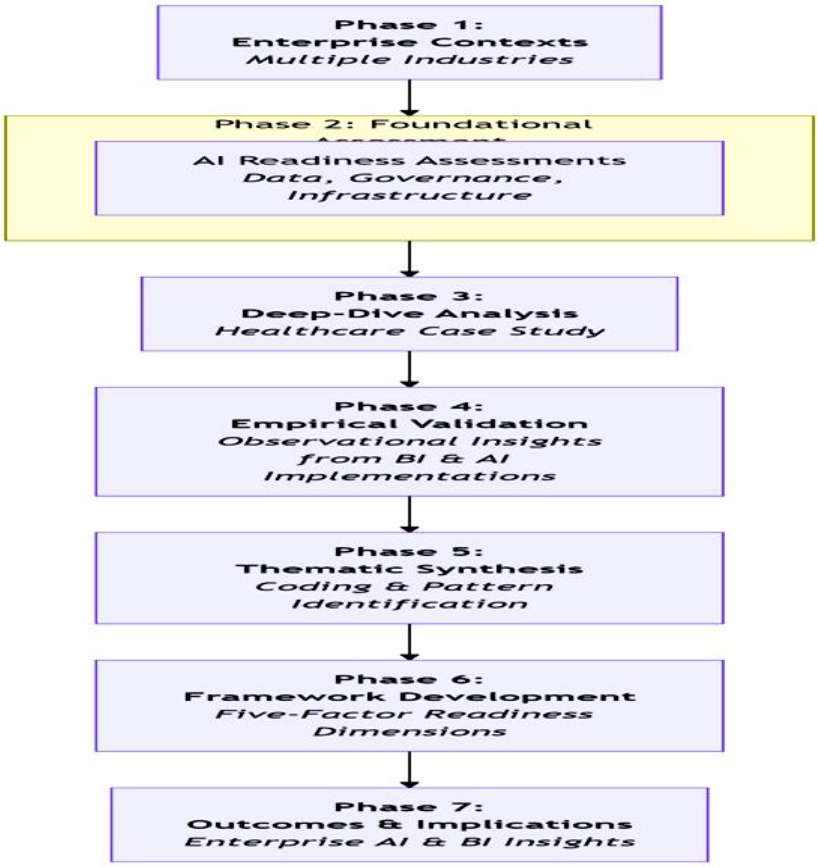

Figure 2. Research Process for Assessing Unstructured Data Readiness

The research process begins with data collection from enterprise AI readiness assessments across multiple industries. These insights are complemented by a detailed healthcare case study and observational data from BI and AI implementations. All data sources are integrated through thematic synthesis, leading to the identification of core readiness dimensions and the development of a structured conceptual framework for unstructured data readiness in enterprise AI environments.

4.6 Trustworthiness and Rigor

To enhance research rigor, several validation strategies were employed:
Data triangulation across multiple sources[11]
Analytical transparency through systematic coding procedures
Pattern replication across enterprises to ensure consistency
Practitioner feedback to validate interpretive accuracy
These measures strengthen the credibility and transferability of the findings.

4.7 Ethical Considerations

All organizational data were anonymized, and no personally identifiable information was collected. Ethical research guidelines for organizational and information systems research were strictly followed [12].

4.8 Methodological Limitations

As a qualitative study, the findings are not statistically generalizable. However, qualitative generalization through analytical transferability is appropriate for theory-building research in emerging domains such as enterprise AI readiness [13].

4.9 Summary

This study employs a qualitative, practice-based methodology integrating enterprise readiness assessments, a healthcare case study, and observational insights. Through thematic synthesis, the methodology supports the development of an empirically grounded framework for assessing unstructured data readiness in AI-enabled BI environments.

5. Dimensions of Unstructured Data Readiness

Enterprise readiness for unstructured data utilization is multi-dimensional, extending beyond technical capability to include governance, security, and operational maturity. Based on thematic synthesis across enterprise assessments and case evidence, this study identifies **five core dimensions** that collectively determine an organization's ability to operationalize unstructured data for AI-enabled business intelligence.

5.1 Data Discovery and Cataloging

A foundational challenge identified across most organizations is the **lack of comprehensive visibility into unstructured data assets**. Unlike structured databases, unstructured data is typically dispersed across heterogeneous platforms, often without centralized ownership or standardized documentation.

Data discovery refers to the systematic identification of:

- Where unstructured data resides
- Its volume and growth velocity
- Business relevance and potential analytical value
- Associated regulatory and operational risks

Common repositories include collaboration platforms, email systems, customer relationship management tools, enterprise resource planning attachments, and legacy file servers. In many enterprises, these repositories have evolved organically over years, leading to redundancy, outdated content, and unknown risk exposure.

Table 1 summarizes the most frequently observed sources of enterprise unstructured data and their associated business risks.

Table 1. Common Sources of Enterprise Unstructured Data

Source	Example Content	Business Risk
SharePoint	Policies, procedures	Medium
Email systems	Institutional knowledge	High
EMR systems	Clinical notes	Very High
File shares	Legacy documents	Medium

A key finding is that organizations that conducted **time-bound discovery sprints** typically lasting four to six weeks were able to create **data heat maps** highlighting high-value and high-risk content. These heat maps provided a rational basis for prioritizing AI use cases, enabling early value realization while managing compliance exposure.

5.2 Data Quality and Metadata

Unstructured data quality differs fundamentally from structured data quality, requiring distinct evaluation criteria. Rather than focusing on accuracy and completeness at the field level, unstructured data quality emphasizes:

- **Accessibility:** Is the content machine-readable?
- **Coherence:** Is the content logically structured and contextually meaningful?
- **Metadata completeness:** Are author, date, version, and classification attributes present?

Across analyzed enterprises, poor-quality unstructured data such as scanned PDFs without optical character recognition (OCR), duplicated drafts, or outdated policy documents was found to significantly degrade AI model performance, particularly in retrieval-augmented generation (RAG) pipelines.

Organizations demonstrating higher readiness implemented **automated quality scanners** that assessed:

- OCR confidence scores
- Language consistency
- Metadata presence and accuracy
- Content freshness and relevance

By establishing **quality gates** before data ingestion into AI pipelines, these organizations reduced hallucination risk, improved retrieval precision, and lowered downstream processing costs.

5.3 Access and Security Architecture

Unstructured data frequently contains **mixed-sensitivity information**, often within the same document. This characteristic presents a major challenge for AI systems, which must simultaneously enable analytical access and enforce strict regulatory compliance.

Enterprises operating in regulated environments must adhere to frameworks such as:

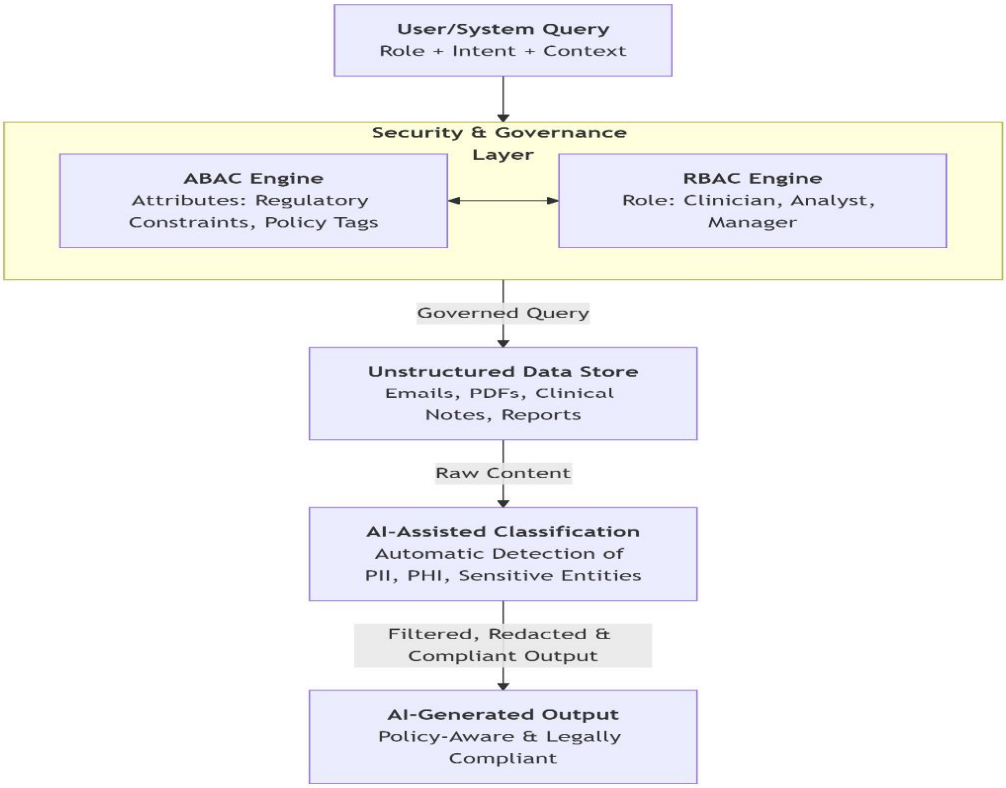

Figure 2. AI-Aware Access Control Model (Conceptual)

- Health Insurance Portability and Accountability Act (HIPAA)
- General Data Protection Regulation (GDPR)
- Sarbanes–Oxley Act (SOX)

Rather than relying solely on traditional perimeter-based security, advanced organizations embedded **AI-aware access controls** directly into their data pipelines.

The model illustrates a layered access architecture combining role-based access control (RBAC) and attribute-based access control (ABAC). AI systems interact with unstructured documents through a security layer that dynamically applies redaction, masking, or content filtering based on user role, query intent, and regulatory constraints. Sensitive entities (e.g., personal identifiers, protected health information) are detected automatically, ensuring that AI-generated outputs comply with policy and legal requirements.

Embedding security logic into AI workflows was found to significantly reduce compliance risk while enabling broader adoption of AI-enabled BI tools.

5.4 Infrastructure and Processing Capabilities

Unstructured data analytics requires infrastructure fundamentally different from traditional BI systems optimized for structured queries. Mature architectures integrate scalable storage, intelligent document processing, semantic search, and vector-based retrieval.

Table 2 outlines the core infrastructure layers commonly observed in enterprise-grade unstructured data BI platforms.

Table 2. Infrastructure Components for Unstructured Data BI

Layer	Technology Examples
Storage	Azure Blob Storage, AWS S3
Processing	Document AI, Amazon Textract
Search	Azure AI Search
Vector DB	Pinecone, Milvus

A recurring challenge identified was **cost escalation** as document volumes and embedding requirements scaled. Organizations that achieved sustainable operations adopted cost optimization strategies such as:

- Tiered storage for cold data
- Reusable embeddings for frequently accessed content
- Incremental indexing rather than full reprocessing

These practices significantly improved scalability without compromising performance.

5.5 Retrieval-Augmented Generation Readiness

Retrieval-augmented generation has emerged as a dominant architectural pattern for enterprise AI applications involving unstructured data. However, RAG readiness depends on several interdependent factors, including:

- Effective document chunking strategies
- Domain-appropriate embedding models
- Retrieval precision and ranking logic
- Latency and cost optimization mechanisms

Organizations with higher RAG maturity continuously evaluated system performance using relevance scoring, user feedback loops, and cost-per-query metrics. This continuous evaluation approach ensured that AI outputs remained accurate, context-aware, and economically viable at scale.

6. Case Study: Healthcare Enterprise AI Implementation

To ground the readiness framework in real-world practice, this section presents a detailed healthcare case study illustrating how unstructured data readiness dimensions translate into measurable outcomes.

6.1 Organizational Context

The case organization is a regional healthcare system operating multiple hospitals and outpatient facilities. Over a 15-year period, the organization accumulated a vast repository of unstructured data, including clinical notes, treatment protocols, research publications, and internal guidelines.

Clinicians faced significant challenges locating relevant information quickly, leading to inefficiencies, duplicated effort, and increased cognitive load. The organization initiated an enterprise AI program to deploy an **AI-powered clinical assistant** integrated with existing BI systems.

6.2 Implementation Phases

The implementation followed a phased approach aligned with the readiness dimensions identified in previous section:

- **Discovery:** Over 3 million documents were cataloged across electronic medical record systems, collaboration platforms, and shared drives.
- **Quality:** OCR processing and metadata normalization were applied to improve machine readability and semantic consistency.
- **Security:** Automated detection of protected health information enabled role-based redaction and context-aware access controls.
- **Infrastructure:** Cloud-based document processing pipelines and vector databases were deployed to support scalable retrieval.
- **RAG Optimization:** Domain-specific embeddings and hybrid keyword–vector search improved clinical relevance.

This phased execution minimized disruption while enabling incremental value realization.

6.3 Results

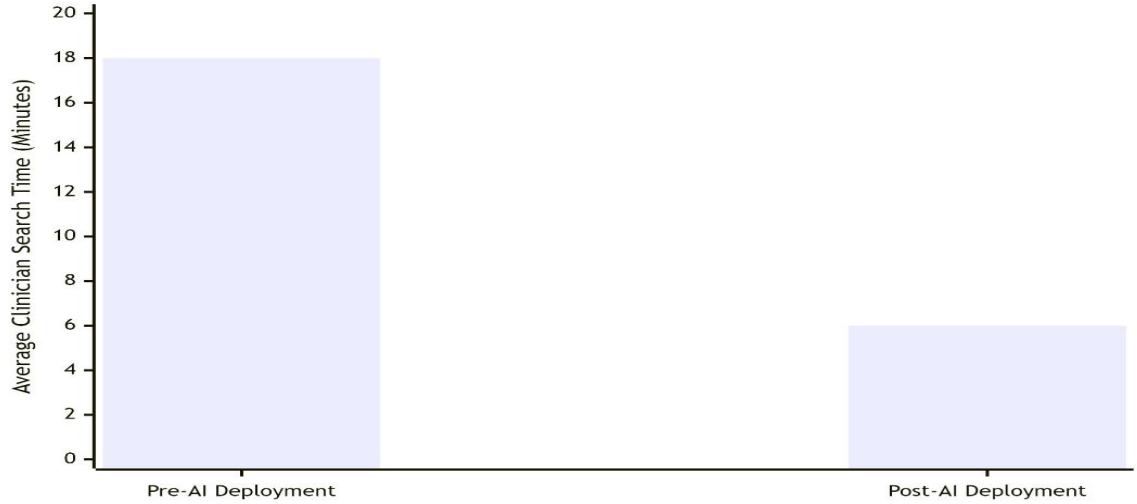

Figure 3. Productivity Gains Before and After AI Deployment

The bar chart compares average clinician time spent searching for information before and after AI deployment. The post-implementation bar shows a substantial reduction, visually reinforcing productivity gains achieved through AI-enabled access to unstructured knowledge.
The deployment produced significant operational and performance improvements:

67% reduction in average clinician search time
94% accuracy in protocol and guideline recommendations

Return on investment achieved within 8 months
These outcomes highlight the tangible business and clinical value of structured readiness planning for unstructured data AI initiatives.

6.4 Cross-Case Insights

Comparing this healthcare case with findings from other industries reveals that while regulatory requirements differ, the underlying readiness dimensions remain consistent. Organizations that invested early in discovery, quality, and security consistently achieved faster time-to-value and lower long-term operational risk.

7. Common Pitfalls and Mitigation Strategies

Despite growing investment in enterprise AI, many initiatives involving unstructured data fail to progress beyond pilot stages. Analysis across multiple organizations reveals a set of recurring failure patterns that undermine AI-enabled BI outcomes.

One of the most common pitfalls is underestimating data preparation effort. Organizations frequently assume that unstructured data can be directly consumed by AI models without adequate discovery, cleaning, and metadata enrichment. This results in poor retrieval accuracy, increased hallucination risk, and low user trust.

A second critical issue is the neglect of data governance and security. In the absence of AI-aware access controls, enterprises either over-restrict data access limiting AI usefulness or expose sensitive information, increasing regulatory and reputational risk.

Another failure pattern involves over-engineering technical solutions. Some organizations prioritize advanced model architectures and complex pipelines before establishing foundational readiness, leading to high costs and operational fragility. Closely related is weak change management, where end users are insufficiently involved in design and rollout, resulting in resistance or underutilization.

Finally, many initiatives fail by ignoring human AI collaboration. Treating AI systems as replacements rather than augmentation tools reduces adoption and limits value realization.
Successful organizations mitigate these challenges through:

- Phased, readiness-led implementation
- Early and continuous stakeholder engagement
- Governance embedded into AI workflows
- Clear positioning of AI as decision support rather than decision replacement

These mitigation strategies significantly increase the likelihood of sustainable AI adoption.

8. ROI and Business Value

Unstructured data readiness delivers measurable business value across multiple dimensions. Unlike experimental AI deployments, readiness-driven initiatives demonstrate clear financial and operational returns.

The most immediate value is realized through productivity gains, particularly in knowledge-intensive roles. Employees spend less time searching for information and more time applying insights. In regulated industries, compliance risk reduction represents a second major value driver, as AI-aware governance reduces exposure to fines, audits, and data breaches.

A third value category is revenue enablement, where faster insight generation improves customer engagement, product development, and decision speed.

Table 3 presents a representative three-year ROI projection based on aggregated enterprise observations.

Category	Estimated Value
Productivity Gains	$5–15M
Risk Reduction	$2–5M
Revenue Uplift	$3–10M

Table 3. Sample 3-Year ROI Projection
These estimates highlight that unstructured data readiness is not a cost center but a strategic investment with compounding returns over time.

9. Readiness Roadmap

Based on empirical findings, this study proposes a three-phase readiness roadmap to guide enterprises from initial assessment to scalable AI-enabled BI adoption.

- Phase 1: Discovery
  This phase focuses on cataloging unstructured data assets, identifying high-value use cases, and mapping regulatory risk. The outcome is a prioritized inventory and initial business case.
- Phase 2: Deep Analysis
  Organizations assess data quality, metadata completeness, security controls, and infrastructure gaps. Pilot AI use cases are executed to validate assumptions and refine governance mechanisms.
- Phase 3: Strategy and Scale
  The final phase establishes enterprise-wide standards for unstructured data management, integrates AI into BI workflows, and defines operating models for continuous improvement.

This roadmap emphasizes progressive maturity, reducing risk while enabling early value capture.

10. Discussion

The findings of this study reinforce a critical insight: enterprise AI success depends more on data readiness than on model sophistication. Across industries, organizations that invested heavily in advanced AI models without addressing unstructured data foundations experienced limited or short-lived success. Treating unstructured data as a strategic enterprise asset—rather than a byproduct of operations enables sustained BI advantage. The readiness dimensions identified in this study align technical, organizational, and governance capabilities, offering a holistic lens for AI adoption.

Importantly, the results suggest that readiness is not static. As AI technologies evolve, continuous reassessment of discovery, quality, security, and infrastructure becomes essential. This dynamic perspective positions readiness as an ongoing capability rather than a one-time project.

11. Conclusion

Enterprise AI has fundamentally shifted the center of gravity in business intelligence from structured reporting toward knowledge-centric, AI-driven insight generation. In this new paradigm, unstructured data represents both the greatest opportunity and the greatest risk.

This study demonstrates that organizations investing in unstructured data readiness across discovery, quality, security, infrastructure, and retrieval-augmented generation are best positioned to realize the promise of AI-enabled BI. Conversely, enterprises that neglect these foundations face stalled initiatives, escalating costs, and erosion of user trust.

Unstructured data readiness is therefore not merely a technical prerequisite, but a strategic imperative for organizations seeking durable advantage in the AI era. Future research may

extend this framework through quantitative validation and sector-specific maturity models, further strengthening its applicability across enterprise contexts.

Acknowledgment:
The authors would like to acknowledge **Professor Sazzad Hossain, PhD**, Professor, Department of System Management & Information Security, Faculty of Artificial Intelligence & Digital Technologies, **Samarkand State University named after Sharof Rashidov (Uzbekistan)**, for his valuable guidance, academic insights, and continuous support in successfully completing this research.

Author Contribution:
Abul Fattah Mohammad Al Mohaimin[1]* contributed to the conceptualization of the study, research design, and overall coordination of the work.
Mushfiqur Rahman[2] contributed to data collection and preliminary analysis.
Dr. Rubaiyat-E-Mehnaz[3] provided methodological guidance and critical review of the manuscript.
Kaium Siddik Anando[4] assisted in data analysis, interpretation of findings, and manuscript preparation.
Sazzad Hossain, PhD[5] contributed through academic supervision, intellectual input, and final review of the manuscript.

Funding Information:
This research was conducted with **self-funding** and did not receive any specific grant from public, commercial, or not-for-profit funding agencies.

Conflict of Interest:
The authors declare **no conflict of interest**.

References

1. Davenport, T., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, 96(1), 108–116.
2. Marr, B. (2020). *Artificial intelligence in practice: How 50 successful companies used AI and machine learning to solve problems*. Wiley.
3. Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165–1188.
4. Manyika, J., et al. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.
5. Brown, T. B., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
6. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., & Levy, O., et al. (2020). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.
7. Kietzmann, J., Paschen, J., & Treen, E. (2018). Artificial intelligence in advertising: How marketers can leverage AI. *Journal of Advertising Research*, 58(3), 263–267.
8. Zhang, C., & Zheng, Y. (2021). Unstructured data analytics in enterprise AI: Challenges and approaches. *Information Systems Frontiers*, 23, 115–130.
9. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

10. Langley, P., & Simon, H. (1995). *Applications of machine learning and reasoning to enterprise data*. AAAI Press.
11. Bhagwat, R., & Ramaswamy, R. (2019). Managing unstructured data for AI: A practical guide for enterprise adoption. *IBM Journal of Research and Development*, 63(4/5), 1–12.
12. Patil, D. J., & Davenport, T. H. (2016). Data science and predictive analytics: Relevance for enterprise AI. *Harvard Business Review*, 94(6), 10–14.
13. Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2), 171–209.
14. Lewis, M., et al. (2021). Retrieval-augmented generation for knowledge-intensive NLP tasks. *arXiv preprint arXiv:2005.11401*.
15. Liu, F., et al. (2021). RAG: Combining retrieval and generative models for AI applications. *NeurIPS Workshop on Machine Learning Systems*.
16. Gartner, Inc. (2022). Market Guide for AI-Ready Unstructured Data Management Platforms. *Gartner Research Report*.
17. Zhang, Y., Chen, Y., & Liu, Z. (2020). Data quality assessment for unstructured data analytics. *Information Processing & Management*, 57(6), 102–127.
18. Baars, H., & Kemper, H. G. (2008). Management support with structured and unstructured data an integrated business intelligence framework. *Information Systems Management*, 25(2), 132–148.
19. Vassiliadis, P., et al. (2018). Enterprise data pipelines for AI applications: Architecture and challenges. *Journal of Big Data*, 5, 38.
20. Marr, B. (2022). *Data strategy: How to profit from a world of big data, analytics and the AI revolution*. Kogan Page.
21. Hirt, M., & Willmott, P. (2014). Strategic principles for AI deployment in enterprises. *McKinsey Quarterly*, 3, 40–51.
22. LinknWin LLC. (2023). *Enterprise AI readiness assessment framework for unstructured data*. Internal whitepaper.