International Journal of AI, Engineering and Management Studies (IJAIEMS)
Volume 1, Issue 1, Jun 2026
Page no : 219-230

Generative AI-Based Intelligent Information Retrieval and Knowledge Management

D. Nirosha

Independent Researcher, USA

Abstract

The rapid growth of digital information has created significant challenges for organizations in retrieving, organizing, and managing knowledge effectively. Traditional information retrieval systems primarily depend on keyword matching and predefined indexing techniques, which may not adequately understand the context, intent, and relationships within large collections of unstructured information. Generative Artificial Intelligence (GenAI), particularly large language models, provides new opportunities for improving information retrieval and knowledge management by enabling semantic understanding, natural-language interaction, contextual response generation, and automated knowledge organization. This paper proposes a Generative AI-Based Intelligent Information Retrieval and Knowledge Management framework that integrates document processing, semantic representation, intelligent retrieval, generative response generation, and knowledge management. The proposed approach allows users to submit natural-language queries and obtain contextually relevant responses from organizational knowledge repositories. The framework also supports automatic summarization, knowledge classification, question answering, and continuous knowledge updating. The paper discusses the architecture, major components, operational workflow, benefits, challenges, and potential applications of the proposed framework. The study demonstrates that the integration of Generative AI with information retrieval can improve accessibility, reduce information-search effort, and support more effective organizational knowledge utilization. However, issues related to hallucination, data privacy, security, knowledge freshness, and response reliability must be addressed for dependable real-world deployment.

Keywords: *Generative Artificial Intelligence, Information Retrieval, Knowledge Management, Large Language Models, Natural Language Processing, Semantic Search, Intelligent Systems*

Received : 08.06.2026	Acceptance : 13.06.2026	Publication : 15.06.2026

1. INTRODUCTION

The amount of digital information generated by organizations has increased rapidly with the development of cloud computing, enterprise applications, digital communication, and online services. Organizations maintain large collections of documents, reports, emails, technical manuals, research papers, policies, databases, and other knowledge resources. Although these resources contain valuable information, finding the right information at the right time has become increasingly difficult.

Traditional information retrieval systems generally use keywords, metadata, predefined categories, and indexing mechanisms to locate relevant information. These techniques are useful for structured searches, but they may provide limited results when users express complex questions using natural language. A user may know what information is required without knowing the exact keywords used in the underlying documents. Consequently, conventional retrieval approaches can produce irrelevant results or require users to examine multiple documents manually.

Artificial Intelligence has significantly changed the way information can be searched and analyzed. Machine learning and Natural Language Processing (NLP) techniques enable systems to understand text, identify relationships, classify documents, and improve search results. Recent developments in Generative AI and Large Language Models (LLMs) have further expanded these capabilities. Generative AI can interpret natural-language questions and generate responses based on learned language patterns and available contextual information.

The combination of Generative AI and information retrieval creates an opportunity to develop intelligent knowledge management systems. Instead of simply returning a list of documents, an AI-based system can identify relevant knowledge sources, understand their content, summarize important information, and generate a meaningful response for the user.

Knowledge management is also an important organizational activity. It involves collecting, organizing, storing, sharing, and utilizing knowledge to support organizational objectives. An intelligent knowledge management system can reduce the time required to locate information and improve knowledge sharing among employees and users.

This research proposes a Generative AI-Based Intelligent Information Retrieval and Knowledge Management framework. The proposed framework integrates document processing, semantic search, knowledge representation, retrieval mechanisms, and Generative AI to provide intelligent access to organizational knowledge.

The main objectives of the study are:

- To develop an intelligent framework for AI-based information retrieval.
- To improve the semantic understanding of user queries.
- To integrate organizational knowledge repositories with Generative AI.
- To provide contextual and natural-language responses.
- To support automated knowledge organization and summarization.
- To identify major challenges associated with Generative AI-based knowledge management.

2. RELATED WORK

Information retrieval has traditionally focused on identifying relevant documents from large collections. Classical retrieval approaches generally depend on term frequency, document indexing, and keyword similarity. These methods provide efficient searching but may have limitations when the relationship between the query and document is semantic rather than lexical.

The development of machine learning introduced more advanced retrieval approaches. Machine learning models can learn patterns from documents and user interactions, enabling improved ranking and classification. NLP techniques further support tasks such as text classification, entity extraction, sentiment analysis, and semantic understanding.

Deep learning has introduced neural approaches to information retrieval. Neural language models can represent words, sentences, and documents in a way that captures semantic relationships. These representations allow systems to identify documents that are conceptually related to a query even when they do not contain exactly the same keywords.

Transformer-based language models have significantly improved natural-language processing. These models can process contextual relationships within text and have become important technologies for search, summarization, question answering, and conversational applications.

Generative AI extends these capabilities by generating human-readable responses. Instead of requiring users to interpret search results manually, a generative system can retrieve relevant

information and produce a summarized response. This approach is particularly useful in organizational environments where employees need quick access to policies, technical documents, reports, and internal knowledge.

However, Generative AI also introduces challenges. A language model may generate information that is not supported by the source documents. Such incorrect or unsupported information is commonly described as hallucination. Therefore, integrating retrieval mechanisms with generative models is important for knowledge-intensive applications.

Retrieval-Augmented Generation (RAG) provides an important approach in this area. In a RAG-based architecture, relevant information is retrieved from an external knowledge source before the generative model produces the final response. This allows the system to use organization-specific and updated information rather than depending exclusively on information encoded during model training.

Despite these developments, there remains a need for integrated frameworks that combine information retrieval, Generative AI, and organizational knowledge management while addressing reliability, privacy, security, and knowledge freshness.

3. PROPOSED GENERATIVE AI-BASED FRAMEWORK

The proposed framework provides an intelligent interface between users and organizational knowledge repositories. The system combines semantic retrieval with Generative AI to transform large-scale knowledge collections into accessible and useful information.

The major components of the framework are:

1. Knowledge Sources
2. Data Preprocessing
3. Knowledge Repository
4. Semantic Representation
5. Query Processing
6. Intelligent Retrieval
7. Generative AI Module
8. Response Validation
9. User Interface
10. Knowledge Updating

3.1 Knowledge Sources

The first component consists of the information sources used by the organization. These may include:

- Research papers
- Technical reports
- Business documents
- Policy documents
- Training materials
- Frequently Asked Questions
- Product documentation

- Project reports
- Internal databases
- Digital archives

The system can combine information from multiple sources to create a centralized knowledge environment.

3.2 Data Preprocessing

Raw documents may contain unnecessary formatting, duplicate information, tables, headers, metadata, and other elements. The preprocessing component converts these documents into a consistent format.

The process may include:

- Document extraction
- Text cleaning
- Duplicate removal
- Sentence segmentation
- Metadata extraction
- Document classification
- Text normalization
- Document segmentation

Large documents can be divided into smaller meaningful sections. This improves retrieval efficiency because the system can retrieve specific portions of a document instead of processing the entire document.

3.3 Knowledge Repository

The processed information is stored in a centralized knowledge repository. Depending on the application, the repository can contain structured and unstructured information.

The repository maintains document content along with metadata such as document title, author, date, category, department, and source. Access controls can also be applied to ensure that users can only retrieve information for which they have permission.

3.4 Semantic Representation

A major feature of the proposed system is semantic representation. Instead of representing documents only through keywords, the system converts text into numerical representations that capture semantic relationships.

This enables the system to understand that different expressions may refer to similar concepts. For example, a query such as "How can employees access company leave policies?" may retrieve documents containing terms such as "employee absence regulations" even when the exact words in the query are not present.

Semantic representations therefore improve the ability of the retrieval system to identify conceptually relevant information.

3.5 Query Processing

Users interact with the system through natural-language queries. The query processing component analyzes the user's question and identifies the underlying information requirement.

For example, instead of searching:

"cloud security policy"

a user may ask:

"What security procedures should employees follow when accessing organizational cloud services?"

The system interprets the intent of the question and searches for relevant information across the knowledge repository.

3.6 Intelligent Retrieval

The intelligent retrieval component identifies the most relevant knowledge sources. It combines semantic similarity, metadata, document relevance, and query context to rank potential information sources.

The retrieved content is then provided to the Generative AI module as contextual information.

This retrieval stage is important because the generative model should not rely only on its previously learned knowledge. Instead, it should use relevant organizational information to produce a grounded response.

3.7 Generative AI Module

The Generative AI module processes the user query together with the retrieved information. It generates a natural-language response that directly addresses the user's question.

The module can perform several tasks:

- Question answering
- Document summarization
- Knowledge extraction
- Report generation
- Content classification
- Conversational search
- Comparative analysis
- Knowledge explanation

The generated response can also include references to the retrieved documents so that users can verify the information.

3.8 Response Validation

Response validation is included to improve reliability. The generated response is compared with the retrieved knowledge before being presented to the user.

The validation stage can identify:

- Unsupported statements
- Missing information
- Conflicting information
- Outdated content
- Potentially inaccurate responses

If insufficient evidence is available, the system can inform the user rather than generating an unsupported answer.

3.9 User Interface

The user interface provides a simple conversational environment for accessing organizational knowledge. Users can submit questions using natural language instead of complex search commands.

The interface can provide:

- Generated answers
- Relevant source documents
- Document summaries
- Related questions
- Search history
- Knowledge categories
- Source references

This approach reduces the technical knowledge required to interact with organizational information systems.

3.10 Knowledge Updating

Knowledge management systems must continuously update their information. New documents, revised policies, research findings, and organizational changes can be incorporated into the repository.

An automated updating process can periodically identify new and modified documents, preprocess them, generate semantic representations, and add them to the knowledge repository.

This helps maintain the relevance and freshness of the system.

4. SYSTEM WORKFLOW

The proposed framework operates through a sequence of interconnected stages.

First, organizational documents are collected from approved knowledge sources. The documents are then processed to remove unnecessary content and divide large documents into meaningful sections.

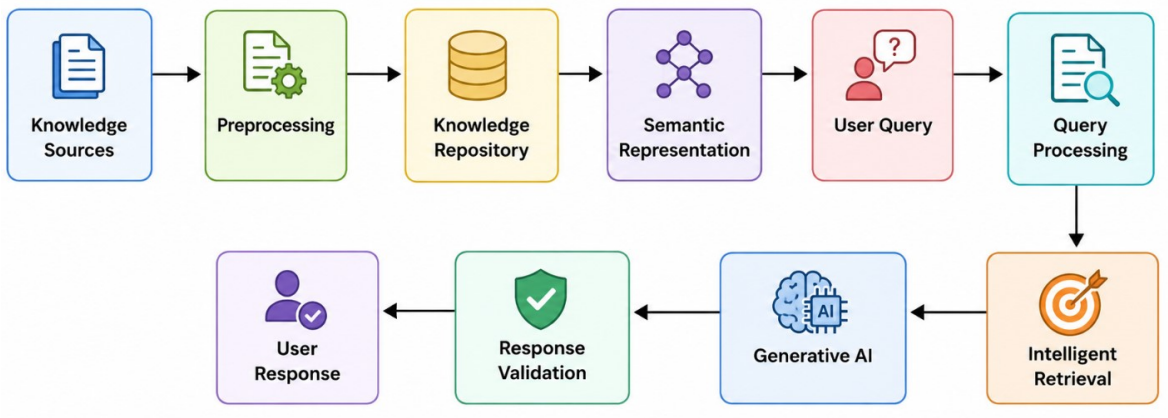
Next, the processed content is converted into semantic representations and stored in the knowledge repository. When a user submits a query, the system analyzes the query and identifies its semantic meaning.

The retrieval module searches the repository and selects the most relevant knowledge segments. These segments are then provided to the Generative AI model as contextual information.

The Generative AI model analyzes the retrieved information and generates a response. Before displaying the response, the validation component checks whether the generated information is sufficiently supported by the retrieved sources.

Finally, the response and relevant source information are presented to the user. User interactions may also be analyzed to improve future retrieval and knowledge organization.

The overall workflow can be represented as:



5. KNOWLEDGE MANAGEMENT CAPABILITIES

The proposed system extends beyond traditional search by providing multiple knowledge management capabilities.

5.1 Knowledge Discovery

The system enables users to discover information that may not be easily found through traditional keyword searches. Semantic retrieval can identify relationships between concepts across different documents.

5.2 Knowledge Organization

Documents can be automatically classified according to topic, department, document type, or other organizational categories. This reduces the manual effort required to organize large information collections.

5.3 Knowledge Sharing

Employees can access organizational knowledge through a common conversational interface. This supports knowledge sharing across departments and reduces dependence on individual employees who may possess specialized knowledge.

5.4 Knowledge Summarization

Large reports and documents can be summarized automatically. Users can obtain the major findings and important information without reading every page.

5.5 Knowledge Question Answering

Users can ask direct questions about organizational information. The system retrieves relevant content and generates an understandable response.

5.6 Knowledge Preservation

Organizational knowledge stored in documents can be made more accessible and reusable. This is particularly important when experienced employees leave an organization and their knowledge needs to remain available.

6. APPLICATIONS

The proposed framework can be applied in several domains.

6.1 Education

Educational institutions can use Generative AI-based knowledge management systems to organize academic materials, research documents, course information, and institutional policies.

Students can ask questions and receive contextual answers based on approved educational resources.

6.2 Healthcare

Healthcare organizations can use intelligent retrieval systems to access clinical guidelines, medical documentation, administrative policies, and research resources. Appropriate privacy and security controls are essential in this environment.

6.3 Business Organizations

Businesses can integrate internal reports, employee manuals, business policies, project documentation, and customer-support information into a unified knowledge system.

Employees can quickly retrieve information without manually searching multiple organizational repositories.

6.4 Research Organizations

Research institutions can use the framework to retrieve information from large collections of research papers, technical reports, datasets, and project documents.

Generative AI can also assist researchers in summarizing and comparing related information.

6.5 Customer Support

Customer-service organizations can use the framework to provide intelligent responses based on product manuals, FAQs, service policies, and historical support information.

6.6 Government Organizations

Government departments can organize regulations, procedures, reports, and public information. An intelligent retrieval system can help employees and citizens access relevant information more efficiently.

7. BENEFITS OF THE PROPOSED FRAMEWORK

The integration of Generative AI and information retrieval provides several advantages.

Improved Search Accuracy: Semantic retrieval enables the system to identify information based on meaning rather than exact keyword matching.

Reduced Search Time: Users can obtain summarized answers instead of manually reviewing numerous documents.

Natural-Language Interaction: Users can interact with the knowledge system using normal conversational questions.

Better Knowledge Sharing: Organizational knowledge becomes accessible across different departments and user groups.

Automated Summarization: Large documents can be transformed into concise summaries.

Scalable Knowledge Access: The framework can be expanded as the organization's knowledge repository grows.

Improved Decision Support: Access to relevant and contextual information can help employees make better-informed decisions.

8. CHALLENGES AND LIMITATIONS

Although Generative AI provides significant benefits, several challenges must be considered.

8.1 Hallucination

Generative AI models may produce information that is not supported by the available sources. This can reduce trust in the system. Retrieval-based grounding and response validation are therefore important.

8.2 Data Privacy

Organizational repositories may contain confidential information. Appropriate access controls, encryption, and privacy mechanisms must be implemented.

8.3 Data Security

Unauthorized users may attempt to access restricted information through the AI interface. User authentication and authorization are necessary to protect sensitive knowledge.

8.4 Outdated Information

If the knowledge repository is not regularly updated, the system may generate responses based on outdated documents. Continuous synchronization is therefore necessary.

8.5 Computational Requirements

Large-scale Generative AI systems may require significant computational resources. Organizations must consider infrastructure, processing time, and operational costs.

8.6 Bias

AI systems may reproduce biases present in their training data or organizational knowledge sources. Monitoring and evaluation are necessary to reduce biased outputs.

8.7 Source Reliability

The quality of generated responses depends partly on the quality of the retrieved documents. Incorrect, incomplete, or unreliable source information can affect the final response.

9. EVALUATION FRAMEWORK

The proposed system can be evaluated using several performance dimensions.

9.1 Retrieval Accuracy

Retrieval accuracy measures whether the system identifies relevant documents or knowledge segments for a given query.

9.2 Response Relevance

Response relevance evaluates whether the generated answer directly addresses the user's question.

9.3 Response Grounding

Grounding measures how closely the generated response is supported by the retrieved information.

9.4 Response Time

Response time measures the time required to process a query, retrieve relevant knowledge, generate a response, and present the result.

9.5 User Satisfaction

User satisfaction can be evaluated through surveys and interaction feedback. Users can rate the usefulness, clarity, accuracy, and completeness of generated responses.

9.6 Knowledge Freshness

The system should also be evaluated according to its ability to incorporate newly added or modified organizational information.

A comprehensive evaluation should combine technical performance with human evaluation because numerical retrieval performance alone may not fully represent the quality of a Generative AI-based knowledge management system.

10. DISCUSSION

The proposed framework demonstrates how Generative AI can transform traditional information retrieval into a more interactive knowledge management environment. Traditional systems generally require users to search for documents and interpret the results themselves. In contrast, the proposed approach enables users to communicate with organizational knowledge using natural language.

The integration of semantic retrieval is particularly important because organizational knowledge is often distributed across documents that use different terminology. A semantic approach can identify relationships between user questions and relevant knowledge even when the vocabulary differs.

Generative AI further improves accessibility by transforming retrieved information into readable responses. However, generation should remain closely connected to reliable knowledge sources. A system that generates fluent but unsupported information can create serious risks, particularly in business, healthcare, finance, and other knowledge-sensitive environments.

Therefore, the proposed framework emphasizes retrieval, source grounding, validation, access control, and continuous knowledge updating. These components can help organizations obtain the benefits of Generative AI while reducing some of its major risks.

The framework also has potential for continuous improvement. User feedback can be used to identify unsuccessful searches, missing information, and frequently requested topics. Organizations can use these insights to improve their knowledge repositories and retrieval mechanisms.

11. FUTURE SCOPE

Future research can extend the proposed framework in several directions.

First, multimodal Generative AI can be integrated to process not only text but also images, tables, diagrams, audio, and video. This would allow organizations to manage diverse forms of knowledge through a single platform.

Second, autonomous AI agents can be incorporated to perform more complex knowledge management tasks. Agents could identify outdated documents, recommend knowledge updates, and generate organizational summaries.

Third, privacy-preserving AI techniques can be investigated to support secure knowledge management in sensitive environments.

Fourth, domain-specific language models can be developed to improve performance for specialized areas such as healthcare, law, engineering, finance, and education.

Finally, future studies can investigate long-term user interaction, knowledge quality, response reliability, and organizational productivity to determine the practical impact of Generative AI-based knowledge management.

12. CONCLUSION

This paper presented a Generative AI-Based Intelligent Information Retrieval and Knowledge Management framework for improving access to large-scale organizational knowledge. The proposed framework integrates knowledge sources, preprocessing, semantic representation, intelligent retrieval, Generative AI, response validation, and continuous knowledge updating.

Unlike conventional keyword-based search systems, the proposed approach allows users to interact with organizational knowledge using natural-language questions. Relevant information can be retrieved from knowledge repositories and transformed into contextual responses through Generative AI. The framework also supports document summarization, knowledge discovery, knowledge organization, and knowledge sharing.

The integration of Generative AI provides substantial opportunities for improving organizational information accessibility and reducing the effort required to locate useful knowledge. Nevertheless, hallucination, privacy, security, outdated information, computational requirements, and bias remain important challenges.

Overall, Generative AI combined with intelligent information retrieval represents a promising direction for next-generation knowledge management systems. With appropriate grounding, validation, security, and governance mechanisms, such systems can provide organizations with more efficient, accessible, and intelligent methods for managing and utilizing knowledge.

REFERENCES

1. *Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N. S., Chen, A., Creel, K., Davis, J. Q., Demszky, D., … Liang, P. (2021).* On the opportunities and risks of foundation models. *arXiv.* https://arxiv.org/abs/2108.07258
2. *Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, H., & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey.* arXiv. https://arxiv.org/abs/2312.10997
3. *Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks.* Advances in Neural Information Processing Systems, 33, *9459–9474.*
4. *Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need.* Advances in Neural Information Processing Systems, 30.
5. *Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., … Wen, J.-R. (2023). A survey of large language models.* arXiv. https://arxiv.org/abs/2303.18223
6. *Bhagwat, V. B. (2024). A simplified transition from EBS Payroll to Cloud Payroll: Benefits and Drawbacks. Journal of Computational Analysis and Applications, 33(6), 463-474.*
7. *Bhagwat, V. B. (2025). Simplifying Payroll Balance Conversions in Payroll Systems Implementation through the Use of Generative AI.*
8. *Prodduturi, S. M. (2025). Cryptography in iOS: A study of secure data storage and communication techniques. International Journal on Science and Technology, 16(1).*
9. *Prodduturi, S. M. K. (2025). Opportunities and Challenges for iOS Developers in Exploring the Integration of Augmented Reality Technologies. International Journal of Engineering Science and Advanced Technology (IJESAT), 25(4), 200-207.*

10. *Narapareddy, V. S. R., & Yerramilli, S. K. (2023). ARTIFICIALINTELLIGENCE INCIDENTFORECASTING. International Journal of Engineering TechnologyResearch & Management (IJETRM), 7(12), 551-559.*
11. *Narapareddy, V. S. R. (2022). Strategies for Integrating Services with External Systems Via Rest & Soap. Universal Library of Engineering Technology,(Issue).*
12. *Todupunuri, A. (2025). The role of human-centric AI in building trust in digital banking ecosystems. International Journal of Innovative Science and Research Technology, 10(1), 1281-1286.*
13. *Todupunuri, A. (2025). Improving Customer Experience With Modern Banking Solutions. SSRN Electronic Journal. https://doi. org/10.2139/ssrn, 5120615.*