International Journal of Computer Trends and Technology
ISSN: 2231–2803 / https://doi.org/10.14445/22312803/IJCTT-V73I1P111

Volume 73 Issue 1, 91-97, January 2025
© 2025 Seventh Sense Research Group®

Original Article

Agentic Retrieval-Augmented Generation: Advancing AI-Driven Information Retrieval and Processing

Abhai Pratap Singh[1], Adit Jamdar[2], Prerna Kaul[3]

[1,2]*Independent Researcher, Sunnyvale, California, USA.*
[3]*Independent Researcher, Seattle, Washington, USA.*

[1]*Corresponding Author : abhaipratapsingh1@gmail.com*

Received: 25 November 2024 Revised: 31 December 2024 Accepted: 16 January 2025 Published: 31 January 2025

Abstract *- This paper explores the emerging field of Agentic Retrieval-Augmented Generation (Agentic RAG), an advanced approach to AI-driven information retrieval and processing. Building upon traditional Retrieval-Augmented Generation, Agentic RAG incorporates goal reasoning and self-direction, enabling AI systems to make informed decisions based on user context and intent. The study examines the fundamental components of Agentic RAG, including its multi-agent hierarchical architecture, key features, and enhancements over conventional systems. Applications across various domains, such as healthcare, financial services, businesses, and education, are discussed. The paper also addresses challenges in implementation, including mitigating AI hallucinations, ethical considerations, and computational scalability. Performance evaluation methods and metrics for Agentic RAG systems are outlined, along with case studies demonstrating their effectiveness. Finally, the paper explores future directions for research and development in this rapidly evolving field, highlighting its potential to revolutionize AI-driven information retrieval and processing.*

Keywords *- Agentic Retrieval-Augmented Generation, Information Retrieval, Artificial Intelligence, Multi-Agent Systems, Natural Language Processing.*

1. Introduction

The rapid evolution of Artificial Intelligence (AI) has led to the development of sophisticated frameworks for information retrieval and processing, among which Retrieval-Augmented Generation (RAG) has emerged as a pivotal approach. RAG integrates the capabilities of Large Language Models (LLMs) with external knowledge sources, enhancing the accuracy and contextual relevance of generated outputs. This architecture effectively overcomes some crucial limitations of classical models by improving their ability to refer to relevant information dynamically. This results in promising tasks such as question-answering and dialogue systems [1, 2]. However, there are issues in prior versions of RAG as it could not effectively handle increasingly complex queries. Integrating diverse data formats, such as tabular data embedded within PDFs, also became problematic when applied in a realistic world [3].

These shortcomings have led to the latest evolution from conventional RAG to Agentic Retrieval-Augmented Generation. Agentic RAG combines aspects of data retrieval with goal reasoning and autonomy, i.e., AI systems cannot only retrieve information but also make informed decisions given the context and intent of the user [4]. This is an essential evolution because it allows the AI agents to be one step further in their advanced interactions, considering and adapting to user and other agent goals in the environment. Agentic RAG has an iterative process of retrieval and generation to enhance the capability even more: models improve their output by successive rounds of information retrieval and generation [5].

Agentic RAG is significant in modern AI applications. This trend shows that understanding and acting on users' intentions become much more critical as AI systems evolve to be used as tools and co-workers [6]. Besides, embedding generative AI into more complex retrieval techniques is essential to help navigate modern information landscapes. Users of modern AI systems demand highly personalized and contextually rich interactions [7]. Agentic RAGs are efficient and can create more intuitive and responsive AI systems that could significantly enhance user experience for customer service, education, and beyond [8].

2. Fundamentals of Agentic RAG

Agentic RAG represents the newer generation of an AI information retrieval paradigm: an information retrieval system able to retrieve information from multiple sources contextually in a self-sustained fashion. Fundamentally, Agentic RAG combines original ideas from retrieval-augmented generation with agentic capabilities, whereby systems can perform actions by themselves rather than just accessing external knowledge through retrieval and make choices regarding user's intentions and contextual signals [7, 9]. This paradigm shift emphasizes that understanding user goals and adapting the responses will improve the interaction quality between AI systems and users [10].

Agentic RAG systems have three major components: a strong retrieval mechanism(s), generative models, and an agentic decision-making framework. While each retrieval mechanism retrieves relevant information from a specific data source, the generative model synthesizes such information into coherent outputs. The agentic aspect adds a layer of autonomy, in which the system evaluates various retrievals and selects the most relevant information based on the interaction context. This triadic structure will improve the accuracy of responses and support more dynamic interaction models where AI learns and adapts over time. [11]

While traditional RAG approaches are primarily unidirectional, meaning that information is retrieved and generated, Agentic RAG involves a feedback loop wherein the system iterates over and refines outputs based on retrieved data and user interactions. This iterative process enhances the system's ability to handle complex queries and diverse data formats, thus offering a more personalized and contextually aware user experience [2, 9]. Figure 1 provides a simple flowchart that describes the iterative refinement approach used by Agentic RAG systems.

3. The Architecture of Agentic RAG

The architecture of an Agentic RAG system is based on organizing multiple agents in a hierarchy that enhances the efficiency of information retrieval and processing: a master agent oversees the operation as a whole and specialized sub-agents for duties such as data retrieval, context interpretation, and response generation. The controller agent coordinates the activities of such sub-agents, who, in turn, would have to work together toward a common goal [12]. Such hierarchical organization provides a far more streamlined approach to handling complex queries since each sub-agent can be specifically optimized for its function, improving overall system performance [13].

The controller agent controls the interaction between the sub-agents and integrates the output of these sub-agents into a coherent response. This agent guides the flow of information and assesses each sub-agent's contribution to ensure the final response is contextually relevant and in tune with user intent [14]. The individual sub-agents could be domain-oriented, such as medical information retrieval or technical support, in which the response could be fine-tuned and draw upon domain-specific knowledge [15]. Figure 2 provides an architecture overview of a typical Agentic RAG system, where a controller agent integrates with several sub-agents and orchestrates the fetching of data across these sub-agents to fulfill the user request.

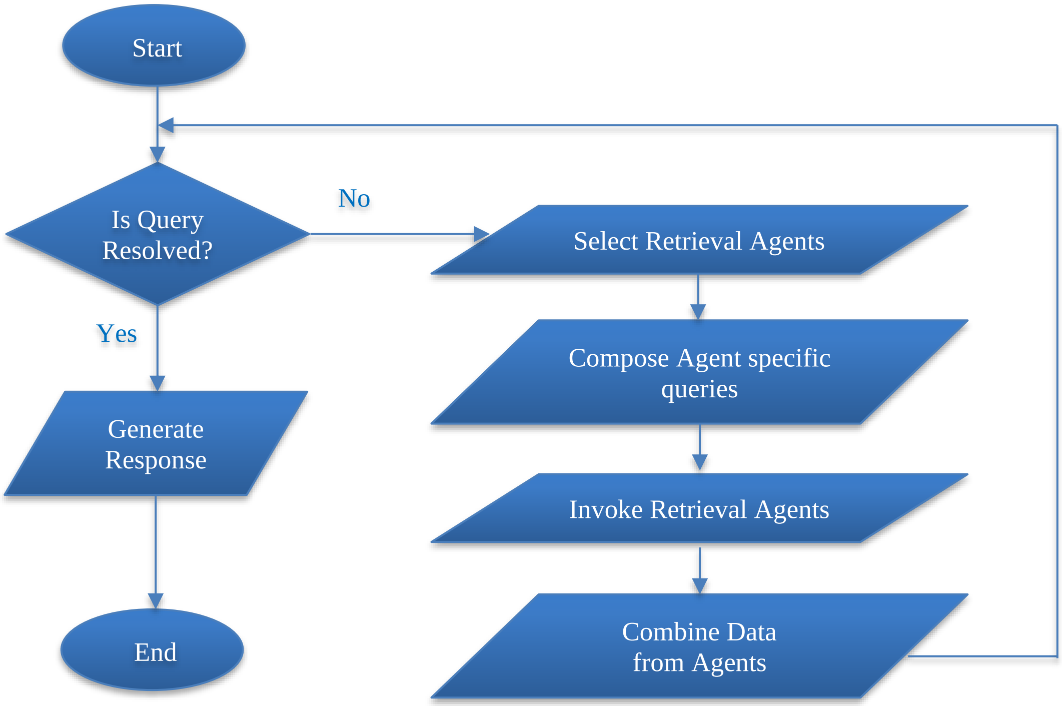

Fig. 1 Agentic RAG iterative framework

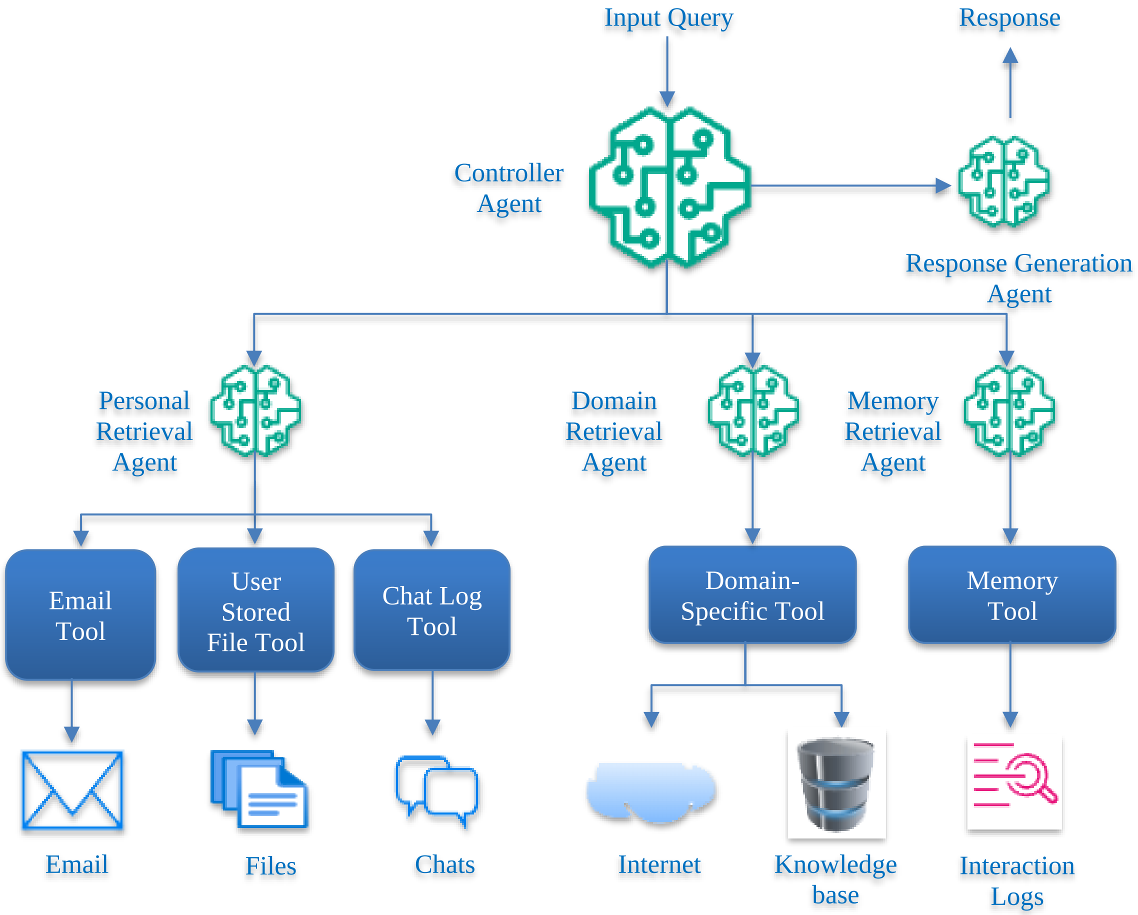

Fig. 2 Typical agentic RAG architecture

It also allows the integration of smaller SLMs within the system for better adaptability and responsiveness. These can be fine-tuned for various tasks, thus allowing flexibility for the diversity required when handling user queries [16]. In this manner, Agentic RAG systems can improve accuracy and contextual understanding, offering a better user experience [17].

4. Key Features and Enhancements of Agentic RAG

Agentic RAG integrates several key features and enhancements that make it far superior in functionality and user interaction. First is the system's adaptive reasoning built into the controller agent, which produces responses regarding the context and evolving user needs. According to Xiong [18], adaptability in dynamic environments means relevance; hence, it can provide tailored information and insights.

Another fundamental capability also built into the controller agent is dynamic planning and execution, whereby the system is enabled to create and update plans in runtime when new information is provided. This feature allows the system to carry out a complicated task efficiently and makes it more effective in responding to users' queries or any changes in the circumstances [19].

The architecture of Agentic RAG further allows collaborative agent networks in which "several specialized agents implement a common purpose." In such collaborations, the work is divided among these agents, so each focuses on the specifics, thus increasing the performance and responsiveness of overall systems [20].

Retrieval mechanisms have been improved within Agentic RAG to fetch relevant information from various heterogeneous databases. These methods will help ensure the extracted data is accurate and contextually fitting for forming meaningful responses [21].

This will eventually allow Agentic RAG to tap into a wider array of resources and capabilities to improve user experience. So far, the system has enabled easy access to several data sources and computational tools, increasing its strength and flexibility in performing tasks [23].

5. Applications and Use Cases of Agentic Retrieval-Augmented Generation

Agentic Retrieval-Augmented Generation (Agentic RAG) demonstrates versatile applications across various domains, leveraging its advanced capabilities to enhance information processing and decision-making.

5.1. Healthcare

In the field of health care, Agentic RAG is a revolution in patient care and medical research:

5.1.1. Clinical Decision Support

Agentic RAG systems analyze, in real-time, medical literature, patient records, and clinical guidelines to provide evidence-based insights on complex cases.[24]

5.1.2. Personalized Treatment Plans

The systems can offer customized treatment plans, mainly in diseases such as oncology, by integrating data specific to the patient and the latest research.[24]

5.1.3. Diagnostic Accuracy

Rapid analysis of patient data, combined with the most current medical knowledge, enhances diagnosis precision [25].

5.2. Financial Services

Agentic RAG provides the financial sector with the ability to process large volumes of data:

5.2.1. Market Analysis

These systems may input financial news, economic data, and/or trading information for the models to create investment strategies in line with prevailing tendencies.[26]

5.2.2. Risk Assessment

Agentic RAG amplifies the possibility of identifying and assessing overall financial risks while analyzing various data stretchers.[26]

5.3. Education

Agentic RAG in education is transforming the way of learning:

5.3.1. Personalized Learning

Agentic RAG-powered virtual teaching assistants could provide customized feedback and course adjustments to meet the needs and progress of each student individually. [26]

5.3.2. Content Creation

With the technology, learning materials created stand a better chance of having their content tailored to adapt to different learning styles and competency levels. [27]

5.4. Business Operations

Agentic RAG is already being implemented for operational optimization in many areas:

5.4.1. Supply Chain Management

It deals with the processing and, therefore, analysis of basic data of supply chains, enhancing their current efficiency and decision-making techniques. [28]

5.4.2. Retail Optimization

AI agents will use Agentic RAG for product placement, design specific promotional displays, and optimize space on the shelf. [29]

Agentic RAG opens new ways to effectively handle complex queries and multimodal data processing by enhancing the capabilities of traditional retrieval-augmented systems through dynamic reasoning, adaptive planning, and collaborative agent networks. It improves user experiences and outcomes in many fields, making it a transformative technology in AI-driven information retrieval and processing.

6. Challenges and Considerations in Agentic RAG Systems

Implementing Agentic Retrieval-Augmented Generation (Agentic RAG) systems presents several challenges and considerations that must be addressed to ensure their effectiveness and ethical deployment.

6.1. Hallucinations

The most critical issue to address is hallucinations in AI outputs. Hallucinations occur when AI models produce information that is factually wrong or misleading. This aspect becomes essential in the highest-stakes environments, such as healthcare, where incorrect information can result in potentially fatal outcomes. As demonstrated by Reddy et al. [30], it is essential to develop nuanced evaluation frameworks to assess AI models both pre-deployment and post-deployment to ensure that possible hallucinations are found and mitigated before they affect clinical decision-making.

6.2. Ethics and Bias

Other essential aspects of developing these RAG systems are ethical considerations and mitigating bias. AI models can involuntarily propagate biases within training data, resulting in unfair or discriminatory outcomes. These biases require a holistic approach that should include diverse, representative datasets, increased transparency, and improved mechanisms for accountability [31][34]. In this respect, ethical guidelines must be established for the development of AI to address complexities associated with bias and fairness issues in AI applications [32].

6.3. Scalability

Other challenges include the scalability and computational needs of Agentic RAG systems. In other words, with more complex models and more extensive data integrated into the systems, much higher computational needs will be posed. This must be translated into developing algorithms and infrastructure that can efficiently support real-time processing and analytics [33][34]. The convergence of AI and high-performance computing may provide a solution for scalability, making faster insights possible and ensuring more efficient data handling [34].

7. Performance Evaluation of Agentic RAG

To establish its lead over traditional RAG, the performance of Agentic RAG needs to be evaluated. Benchmarking against traditional RAG systems involves systematic comparisons using standardized datasets, and tasks are conducted to assess improvements in accuracy, response relevance, and user satisfaction. Recent works by Yang [35][36] have clearly shown the merits of Agentic RAG in handling complex queries and integrating external knowledge, which can result in a higher performance metric.

Standard metrics to evaluate Agentic RAG's effectiveness include precision, recall, F1-score, and user satisfaction ratings. In addition, the correlation measures to be used are Pearson's ρ and Kendall's τ, which quantify the ranking between the generated outputs and the ground truth expectation of the results. [37] This gives an overview of how well the system would work in a natural setting and ensures the test covers both the quantitative and qualitative performance of the system.

Case studies and experimental results further demonstrate the practical applications of Agentic RAG. For example, experiments conducted in healthcare demonstrated impressive results where the Agentic RAG approach was far superior in clinical decision support systems compared to classic models, returning more accurate and contextually relevant information [38]. Similarly, it has been shown that Agentic RAG learns this kind of effective retrieval strategy in open-domain question-answering studies by a wide margin, outperforming classical information retrieval methods [39]. Such results indicate the transformational influence of Agentic RAG on AI-based information retrieval and processing across industries.

8. Future Directions for Agentic RAG

The future of Agentic RAG is immense in terms of potential advancements and innovations. While the technology is still evolving, a few key areas of development and expansion are very likely to shape it:

8.1. Advanced AI Technologies Integration

The integration of state-of-the-art AI technologies will enable agentic RAG systems:

8.1.1. Reinforcement Learning

Integrating reinforcement learning methods can significantly improve the adaptability and efficiency of Agentic RAG in real-time decision-making scenarios by learning from user feedback regarding outcomes to strengthen its strategy for information retrieval and processing. [40]

8.1.2. Advanced Neural Architectures

This could involve developing and incorporating more sophisticated neural network architectures that perform even better with complex queries and understanding context. [40]

8.2. Multimodal Data Processing

Future Agentic RAG systems will likely be adept at processing and analyzing several forms of incoming data, including- text, images, sensor data, video, and audio.

This multimodal capability will also extend to many practical domains, including but not limited to healthcare and intelligent environments in which Agentic RAG originally resides. Integration with and analyses against a wide variety of data sources will supply output with more completeness and addition of context. [41]

8.3. Personalization and Contextual Understanding

The research will aim to improve the personalization capabilities in Agentic RAG in the future, thereby better understanding user context and intent. This will involve developing enhanced user modeling techniques, including personal knowledge graphs for highly tailored responses.

9. Conclusion

The paper has discussed the transformational role of Agentic RAG in AI and information retrieval. The key takeaways are that Agentic RAG enhances the power of traditional retrieval-augmented systems through dynamic reasoning, adaptive planning, and collaborative agent networks. As Hambarde & Proença [42] demonstrated, Agentic RAG enables handling complex queries and multimodal data processing more effectively to enhance user experiences and outcomes. The implications are profound in the field of AI in information retrieval. Agentic RAG solves some of the limitations of traditional systems but, more importantly, provides a path for future work based on emerging AI technologies. Advanced algorithms and potentially external tools will be applied further to enhance the scalability and efficiency of information retrieval processes, making them more responsive to user needs [43].

The future of Agentic RAG is bright, and much remains to be done. Research and development into mitigating bias, computational efficiency, and robust evaluation frameworks are needed. By solving these challenges, Agentic RAG can fully revolutionize information retrieval and processing across various domains [44]. In the final analysis, as Agentic RAG continues to evolve, it is set to play a leading role in shaping the future landscape of AI-driven information retrieval. With its extended capabilities, it will offer considerable benefits to users and organizations alike.

References

[1] S. Nishanth, and S. Swetha, "Enhancing Rag Systems: A Survey of Optimization Strategies for Performance and Scalability," *International Journal of Scientific Research in Engineering and Management*, vol. 8, no. 6, pp. 1-5, 2024. [CrossRef] [Publisher Link]

[2] Jiawei Chen et al., "Benchmarking Large Language Models in a Retrieval-Augmented Generation," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 16, p. 17754-17762, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[3] Uday Allu, Biddwan Ahmed, and Vishesh Tripathi, "Beyond Extraction: Contextualizing Tabular Data for Efficient Summarization by Language Models," *TechRxiv*, pp. 1-5, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] David W. Aha, "Goal Reasoning: Foundations, Emerging Applications, and Prospects," *AI Magazine*, vol. 39, no. 2, pp. 3-24, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[5] Zhihong Shao et al., "Enhancing Retrieval-Augmented Large Language Models with Iterative Retrieval-Generation Synergy," *Findings of the Association for Computational Linguistics: EMNLP*, pp. 9248-9274, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] Sune Müller, Alexander Kempton, and Troels Mønsted, "*Is the Human IS Researcher Dead? Long Live the AI Researcher*," Proceedings of the 57th Hawaii International Conference on System Sciences, pp. 207-216, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Qiushi Huang et al., "Learning Retrieval Augmentation for Personalized Dialogue Generation," *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 2523-2540, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Helin Yang et al., "Artificial-Intelligence-Enabled Intelligent 6G Networks," *IEEE Network*, vol. 34, no. 6, pp. 272-280, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[9] Huayang Li et al., "A Survey on Retrieval-Augmented Text Generation," *arXiv*, pp. 1-13, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Uday Allu, Biddwan Ahmed, and Vishesh Tripathi, "Beyond Extraction: Contextualising Tabular Data for Efficient Summarisation by Language Models," *TechRxiv*, pp. 1-5, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[11] Zijie J. Wang, and Duen Horng Chau, "MeMemo: On-Device Retrieval Augmentation for Private and Personalized Text Generation," *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2765-2770, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[12] Igor N. Glukhikh, Tatiana Y. Chernysheva, and Yaroslav A. Shentsov, "Decision Support in a Smart Greenhouse Using Large Language Model with Retrieval Augmented Generation," *Proceedings of Third International Conference on Digital Technologies, Optics, and Materials Science*, vol. 13217, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[13] Wenyu Huang et al., "Retrieval Augmented Generation With Rich Answer Encoding," *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics*, vol. 1, pp. 1012-1025, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Mangala Madankar, Manoj Chandak, and U.N. Shrawankar, "Agent-Based Multilingual Information Retrieval System: A Design Approach," *Helix*, vol. 8, no. 5, p. 3863-3867, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[15] Jie Zhang et al., "Study on Master-Slave Interaction Model Based on Stackelberg Game in Distributed Environment," *Symmetry*, vol. 12, no. 2, pp. 1-22, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[16] Kenta Nagao, Natsuki Kawaguchi, and Takao Sato, "Design of Data-Driven Multi-Agent Systems," *Proceedings of International Conference on Artificial Life and Robotics*, vol. 28, pp. 205-208, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[17] Jenhui Chen, Obinna Agbodike, and Lei Wang, "Memory-Based Deep Neural Attention (MDNA) for Cognitive Multi-Turn Response Retrieval in Task-Oriented Chatbots," *Applied Sciences*, vol. 10, no. 17, pp. 1-11, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[18] Xingyu Xiong, and Mingliang Zheng, "Integrating Deep Learning with Symbolic Reasoning in Tinyllama for Accurate Information Retrieval," *Research Article*, pp. 1-10, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Shuroug A. Alowais et al., "Revolutionizing Healthcare: The Role of Artificial Intelligence in Clinical Practice," *BMC Medical Education*, vol. 23, no. 1, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[20] Alessandra Sorrentino et al., "Personalizing Care through Robotic Assistance and Clinical Supervision," *Frontiers in Robotics and AI*, vol. 9, pp. 1-12, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[21] Z. Raza, "Integrating Artificial Intelligence Techniques for Advancements in Colorectal Cancer Management: Navigating Past and Predicting Future Direction," *Journal of the Pakistan Medical Association*, vol. 74, no. 4, pp. 165-170, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[22] Rimma Zaripova, Adam Mentsiev, and Olga Kovrizhnykh, "Advancing Parallel Programming Integrating Artificial Intelligence for Enhanced Efficiency and Automation," *E3S Web of Conferences*, vol. 460, pp. 1-9, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[23] Thi Xuyen Nguyen, "English Majors' Perceptions of AI Tool Application in English Language Learning at Tertiary Level in Vietnam," *Journal of Knowledge Learning and Science Technology*, vol. 3, no. 1, pp. 179-193, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[24] E.N. Bureau, Agentic RAG in Healthcare: Bridging Knowledge Gaps for Improved Patient Care, Express Healthcare, 2024. [Online]. Available: https://www.expresshealthcare.in/news/agentic-rag-in-healthcare-bridging-knowledge-gaps-for-improved-patient-care/447366/

[25] BOSS Editorial, The Impact of Retrieval-Augmented Generation (RAG) in the Healthcare Sector, BOSS Magazine, 2024. [Online]. Available: https://thebossmagazine.com/retrieval-augmented-generation-rag-in-healthcare-sector/

[26] Dinesh Rivankar, Agentic RAG: Turbocharging Data-Driven User Experiences, Persistent Systems, 2024. [Online]. Available: https://www.persistent.com/blogs/agentic-rag-turbocharging-data-driven-user-experiences/

[27] Jana, Data Innovation Summit 2025 X Edition, Hyperight, 2024. [Online]. Available: https://hyperight.com/event/data-innovation-summit-2025-x-edition/

[28] Sascha Brodsky, Agentic RAG Turns AI into a Smarter Digital Sleuth, IBM.com, 2024. [Online]. Available: https://www.ibm.com/think/news/ai-detectives-agentic-rag

[29] Agentic AI: 8 Use Cases with Real-life Examples, AIMultiple: High Tech Use Cases & Tools to Grow Your Business, 2024. [Online]. Available: https://research.aimultiple.com/agentic-ai/

[30] Sandeep Reddy et al., “Evaluation Framework to Guide Implementation of AI Systems into Healthcare Settings,” *BMJ Health & Care Informatics*, vol. 28, no. 1, pp. 1-7, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[31] Emilio Ferrara, “Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies,” *Sci*, vol. 6, no. 1, pp. 1-15, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[32] Anna Jobin, Marcello Ienca, and Effy Vayena, “The Global Landscape of AI Ethics Guidelines,” *Nature Machine Intelligence*, vol. 1, no. 9, p. 389-399, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[33] Lavanya Shanmugam et al., “Optimizing Cloud Infrastructure for Real-Time AI Processing: Challenges and Solutions,” *International Journal for Multidisciplinary Research*, vol. 6, no. 2, pp. 1-11, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[34] E.A. Huerta et al., “Convergence of Artificial Intelligence and High Performance Computing on NSF-Supported Cyberinfrastructure,” *Journal of Big Data*, vol. 7, no. 1, pp. 1-12, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[35] Yihe Yang et al., “Advancing Structured Query Processing in Retrieval-Augmented Generation with Generative Semantic Integration,” *Frontiers in Computing and Intelligent Systems*, vol. 9, no. 3, pp. 64-71, 2024. [CrossRef] [Publisher Link]

[36] Yuanjie Lyu et al., “Crud-Rag: A Comprehensive Chinese Benchmark for Retrieval-Augmented Generation of Large Language Models,” *ACM Transactions on Information Systems*, vol. 43, no. 2, pp. 1-32, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[37] Hamed Zamani et al., “Retrieval-Enhanced Machine Learning,” *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2875-2886, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[38] Xiarepati Tieliwaerdi et al., “Exploring the Potential of ChatGPT-4 for Clinical Decision Support in Cardiac Electrophysiology and Its Semi-Automatic Evaluation Metrics,” *medRxiv Preprint*, pp. 1-12, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[39] Kenton Lee, Ming-Wei Chang, and Kristina Toutanova, “Latent Retrieval for Weakly Supervised Open Domain Question Answering,” *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 6086-6096, 2019. [CrossRef] [Publisher Link]

[40] Enver Malsia, and Afrim Loku, “Generative Artificial Intelligence in Health System Management: Transformative Insights,” *Journal of Service Science and Management*, vol. 17, no. 2, pp. 107-117, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[41] Ion-Alexandru Secara, and Dariia Hordiiuk, “Personalized Health Monitoring Systems: Integrating Wearable and AI,” *Journal of Intelligent Learning Systems and Applications*, vol. 16, no. 2, pp. 44-52, 2024. [Google Scholar] [Publisher Link]

[42] Kailash A. Hambarde, and Hugo Proença, “Information Retrieval: Recent Advances and Beyond,” *IEEE Access*, vol. 11, pp. 76581-76604, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[43] Kalyani Pakhale, “Large Language Models and Information Retrieval,” *International Journal for Multidisciplinary Research*, vol. 5, no. 6, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[44] Rahman Taufik et al., “Information Retrieval System for Scientific Publications of Lampung University by Using VSM, K-Means, and LSA,” *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 11, pp. 199-209, 2023. [CrossRef] [Google Scholar] [Publisher Link]