Multi-Agent Retrieval-Augmented Generation (RAG) Systems for Automated Complex Offer Generation in Industrial Project Business: Architectures, Protocols, Orchestration, On-Premise LLMs, and Security (2023–Present)

Quick Reference

Key Findings Table

Theme/Aspect	Key Insights	Representative Citations
Modular & Decentralized Architectures	Modular, agentic, and decentralized RAG designs (e.g., RAGAF, ERI) enable scalability, flexibility, and data control.	1 2 3 4
Communication Protocols	Embedding similarity, function calling, and vector DBs optimize agent interaction and retrieval precision.	
Orchestration Patterns	Iterative, agentic, and utility-maximizing orchestration (e.g., RAGADA, ReAct) support real-time, multi-agent workflows.	3 5 6
On-Premise LLM Setups	On-premise LLMs with permission-aware RAG, IAM integration, and chunk tuning balance security, cost, and performance.	7 8 9 10
Security Aspects	Multi-layered defenses (input validation, adversarial training, RAGForensics) and permission-aware RAG mitigate threats.	
Efficiency, Quality, ROI	Multi-agent RAG systems improve process speed, factuality, and business value in offer generation.	5 13 14

Direct Answer

The review should focus on the evolution and recent advances in multi-agent RAG systems by critically assessing their architectures, orchestration, communication protocols, and on-premise LLM configurations within industrial project offer generation. The literature post-2023 highlights modular design patterns, decentralized setups that ensure data control, iterative retrieval-generation mechanisms, and robust security measures designed to counteract data poisoning and prompt injection. These systems enhance efficiency and increase ROI by enabling real-time, context-aware offer generation while balancing cost and performance trade-offs.

Study Scope

- **Time Period:** 2023–present
- **Disciplines:** AI/ML, industrial informatics, enterprise automation, cybersecurity
- **Methods:** Systematic literature review, meta-analysis, synthesis of empirical and architectural studies

Assumptions & Limitations

- Most empirical evidence is drawn from adjacent domains (e.g., customer support, insurance, supply chain) due to limited direct studies on industrial offer generation.
- Security and ROI metrics are often inferred from system-level benchmarks and case studies, not always from direct industrial deployments.
- Integration challenges with legacy IAM and industrial protocols (e.g., OPC UA) are underexplored in current literature.

Suggested Further Research

- Empirical benchmarking of multi-agent RAG systems in real-world industrial offer generation, focusing on ROI, efficiency, and security.
- Deeper exploration of integration between decentralized RAG and enterprise IAM/OPC UA.
- Investigation of multimodal RAG (text, tabular, visual) for complex industrial offers.

1. Introduction

Background and Motivation

The industrial project business is characterized by high-value, complex, and bespoke offers that require rapid, accurate, and context-aware generation of proposals. Traditional manual or rule-based approaches are increasingly inadequate due to the scale, complexity, and need for real-time adaptation. Multi-agent, retrieval-augmented generation (RAG) systems have emerged as a promising solution, leveraging large language models (LLMs) and advanced retrieval mechanisms to automate and enhance offer generation. These systems integrate modular, agentic architectures, dynamic retrieval, and robust security, addressing the unique challenges of industrial environments where data privacy, compliance, and efficiency are paramount [1] [5].

Scope and Structure of the Review

This review systematically analyzes the state-of-the-art in multi-agent RAG systems for automated complex offer generation in industrial project business, focusing on literature since 2023. The analysis covers system architectures, communication protocols, orchestration patterns, on-premise LLM setups, and security aspects, with a synthesis of how these systems impact efficiency, quality, and ROI. Where direct evidence is limited, insights from closely related domains are included [1] [5] [7].

2. Theoretical Frameworks

Modular and Decentralized Architectures

Recent research emphasizes modular RAG frameworks that separate retrieval, generation, orchestration, and evaluation components. This modularity, often combined with decentralized architectures (e.g., ERI protocol), enables independent updates, scalability, and easier integration with external data sources. Notable frameworks include RAGAF, GROUSER, and decentralized RAG with ERI, which have been validated in domains such as banking, customer support, and insurance [1] [2] [15].

- **Key Features:**
  - **Modularity:** Components such as Generator, Retriever, Orchestration, UI, Source, Evaluation, and Reranker are decoupled for flexibility and maintainability [1].
  - **Decentralization:** ERI protocol allows distributed entities to independently manage retrieval and generation, enhancing data control and resource efficiency [2] [15].
  - **Agentic Design:** Specialized agents (e.g., Function Calling, ReAct, LLMCompiler) enable dynamic data ingestion and real-time reasoning [3] [16] [17].

Agentic and Multi-Modal RAG Designs

Agentic RAG architectures employ multiple specialized agents for tasks such as data authentication, query routing, and custom prompting. Multi-modal RAG systems (e.g., MMA-RAG) integrate structured and unstructured data, supporting complex reasoning and regulatory compliance [3] [5] [16] [17] [18].

- **Hybrid Retrieval:** Combining keyword-based (BM25) and dense embedding models improves factual correctness and semantic similarity [4].
- **Ensemble Approaches:** Weighted average ensembles and clustering enhance retrieval accuracy and semantic answer similarity [19].

Scalability, Flexibility, and Resource Efficiency

Modular and decentralized RAG frameworks improve scalability and flexibility by allowing independent updates and efficient resource usage. Dynamic chunking and ensemble retrieval approaches further optimize performance [13] [19] [20] [21].

- **Dynamic Chunking:** Adjusts retrieval granularity based on task priorities and data types [22].
- **Automated Tools:** RAGuru automates component selection for cost and latency optimization [20].

Synthesis:
Theoretical advances in modular, agentic, and decentralized RAG architectures provide a robust foundation for scalable, flexible, and secure offer generation in industrial contexts. These frameworks enable real-time adaptation, integration with diverse data sources, and efficient resource utilization [1] [2] [3] [4].

3. Methods & Data Transparency

This review is based on a systematic synthesis of peer-reviewed articles, preprints, and technical reports published since 2023. The analysis draws on empirical studies, architectural proposals, and meta-analyses, with a focus on industrial project business and closely related domains (e.g., customer support, insurance, supply chain). All claims are supported by explicit citations to the underlying literature, and where direct evidence is lacking, adjacent domain findings are clearly indicated.

4. Critical Analysis of Findings

System Architectures of Multi-Agent RAG

Modular and Decentralized Architectures

- **RAGAF & GROUSER:** Modularize RAG into distinct components, supporting dynamic retrieval and context-aware generation [1].
- **ERI Protocol:** Enables decentralized operation, improving data control and resource efficiency [2] [15].

- **Agentic RAG:** Specialized agents handle dynamic data ingestion and real-time reasoning, validated in complex simulations [3] [16].

Agentic and Multi-Modal RAG Designs

- **MMA-RAG:** Integrates structured and unstructured data, supporting regulatory compliance in insurance [17].
- **Hybrid Retrieval:** Outperforms paid embeddings in domain-specific tasks [4].
- **Ensemble Models:** Improve semantic answer similarity and retrieval performance [19].

Scalability, Flexibility, and Resource Efficiency

- **Dynamic Chunking:** Outperforms fixed chunk size approaches [22].
- **Automated Optimization:** RAGuru reduces cost and latency while maintaining performance [20].

Synthesis:
Modular, agentic, and decentralized architectures are central to the scalability, flexibility, and security of multi-agent RAG systems in industrial offer generation. These designs facilitate integration with external data sources, dynamic adaptation, and efficient resource utilization [1] [2] [3] [4].

Communication Protocols in Multi-Agent RAG

Agent Communication and Data Exchange

- **Function Calling & ReAct Agents:** Enable dynamic interaction and real-time reasoning [3] [22] [23] [24] [25].
- **Vector Databases & APIs:** Support seamless agent interaction and data exchange [23] [24].
- **Semantic Search:** Embedding-based similarity measures retrieve relevant information dynamically [1] [22].

Embedding Similarity and Function Calling Protocols

- **Similarity Measures:** Centered Kernel Alignment, Jaccard, and rank similarity cluster embedding models and assess retrieval result similarity [26].
- **Function Calling:** Converts text to embeddings for context-aware, up-to-date outputs [27].
- **Hybrid Retrieval:** Combines sparse and dense methods for improved precision [4].

Optimizing Communication: Vector Databases, Caching, and Privacy

- **Vector DB Integration:** Enables semantic indexing and real-time query processing [28].
- **Caching Strategies:** Reduce latency and computational load (e.g., LFU, Proximity) [29] [30].
- **Privacy Risks:** Membership inference and data leakage mitigated by embedding space shifting and prompt-based defenses [31] [32] [33].

Synthesis:
Advanced communication protocols—embedding similarity, function calling, and vector database integration—are critical for efficient, secure, and context-aware agent interaction in multi-agent RAG systems ---

Orchestration Patterns in Multi-Agent RAG

Multi-Agent Orchestration Frameworks

- **RAGADA & RAGAF:** Integrate RAG with multi-agent systems and business algorithms for adaptable AI-human interaction [1] [5].
- **Iterative Utility Maximization:** Personalizes retrieval for each agent using feedback loops [6].
- **Agentic Workflows:** Specialized agents (Function Calling, ReAct) enable dynamic data ingestion and real-time reasoning [3].

Real-Time Reasoning and Dynamic Data Ingestion

- **StreamRAG:** Reduces latency and enhances scalability via lock-aware and traffic-aware query coordination [34].
- **Multi-Semantic RAG:** Integrates knowledge graph structuring and hierarchical semantic reasoning for multi-hop, cross-domain integration [35].
- **CRP-RAG:** Uses reasoning graphs for dynamic knowledge retrieval and aggregation [36].

Task Decomposition, Coordination, and Feedback

- **ReAct & AutoGen:** Alternate between thought, action, and observation, decomposing tasks and integrating feedback [37] [38].
- **Semantic Orchestration:** Context-aware triggers and recommender agents enable flexible, transparent coordination [39].
- **Dynamic Task Allocation:** Optimizes resource utilization and collaborative efficiency [40].

Synthesis:
Iterative, agentic, and utility-maximizing orchestration patterns enable real-time, multi-agent workflows that support complex offer generation, dynamic data ingestion, and robust reasoning [3] [5] [6].

On-Premise LLM Setups for Industrial RAG

On-Premise LLM Architectures and Security

- **Generator & Retriever Models:** Access external knowledge bases while addressing security via data filtering, adversarial training, and access control [7] [9] [10] [41].
- **Permission-Aware RAG:** Integrates with IAM systems for fine-grained, resource-level access control [9].
- **Pre-Processing Frameworks:** Screen inputs to prevent leakage and off-domain queries [41].

Performance Optimization and Cost-Effectiveness

- **RAGuru:** Automates configuration for cost, latency, and accuracy optimization [20].
- **Chunk Size Tuning:** Balances retrieval efficiency, response quality, and computational cost [42] [43].
- **Hybrid Retrieval:** Combines keyword-based and dense embeddings for improved performance [4].

Permission-Aware RAG and Fine-Grained Access Control

- **GraphRAG:** Models industrial assets for semantic access control [44].
- **OPC UA Integration:** Enables secure data exchange and interoperability [45].
- **MARL Frameworks:** Support adaptive access control aligned with industrial traffic patterns [46].

Synthesis:
On-premise LLM setups with permission-aware RAG, IAM integration, and chunk tuning are essential for secure, efficient, and cost-effective offer generation in industrial environments [7] [8] [9] [10].

Security Aspects in Multi-Agent RAG Systems

Threat Landscape and Security Challenges

- **Key Threats:** Data poisoning, prompt injection, privacy leaks, adversarial query manipulation [11] [12] [47] [48].
- **Closed-Source LLM Risks:** Proprietary data exposure mitigated by local open-source generators and security filtering [49].
- **Human-in-the-Loop:** Enhances factual accuracy but introduces design challenges for security and trust [50].

Defense Mechanisms and Robustness Strategies

- **Input Validation & Adversarial Training:** Prioritized for risk mitigation [12].
- **RAGForensics:** Traceback system for identifying poisoned texts [51].
- **Robust Fine-Tuning:** Enhances resilience against noisy or misleading retrieval [52].

Human-in-the-Loop and Privacy-Preserving Approaches

- **Supervised Prompting & Modular Architectures:** Allow human oversight without compromising privacy [53].
- **Traceability:** Links outputs to verifiable data sources for validation and auditing [21] [54].
- **Vector DBs:** Enable privacy-conscious retrieval by avoiding direct exposure of raw data [54] [55].

Permission-Aware RAG and Compliance with Industrial Standards

- **IAM & OPC UA Integration:** Enforces fine-grained access control and regulatory compliance [56] [57] [58] [59].
- **Hierarchical Retrieval:** Improves precision and contextual relevance, supporting compliance [60] [61].

Synthesis:
Multi-layered security strategies—input validation, adversarial training, RAGForensics, and permission-aware RAG—are critical for mitigating threats and ensuring compliance in industrial multi-agent RAG systems ---

Impact on Efficiency, Quality, and ROI in Offer Generation

Efficiency Gains and Process Automation

- **Process Speed:** Modular, agentic RAG architectures automate and accelerate offer generation [5] [13] [14].
- **Resource Utilization:** Automated tools (e.g., RAGuru) optimize cost and latency [14].

Quality Enhancement and Factual Consistency

- **Factual Grounding:** Modular and agentic RAG systems reduce hallucinations and improve output quality [62] [63] [64] [65].
- **Human-in-the-Loop:** Further enhances accuracy and reliability [53] [64].

ROI and Business Value in Industrial Applications

- **Cost-Effectiveness:** RAGuru and hybrid retrieval strategies deliver comparable performance at reduced cost [20].

- **Business Value:** Improved efficiency and quality translate to higher ROI in offer generation [49].

Synthesis:
Multi-agent RAG systems demonstrably improve efficiency, quality, and ROI in offer generation and related industrial processes, though further empirical validation in direct industrial contexts is needed [5] [13] [14].

5. Real-World Implications

- **Industrial Adoption:** Modular, agentic, and decentralized RAG systems are increasingly adopted in industrial settings for offer generation, customer support, and document processing, driven by the need for real-time, context-aware, and secure automation [1] [5] [17].
- **Security & Compliance:** On-premise LLMs with permission-aware RAG and IAM/OPC UA integration address stringent data privacy and regulatory requirements in industrial environments [9] [56].
- **Cost & Performance:** Automated optimization tools and hybrid retrieval strategies enable practical, cost-effective deployment of RAG systems in resource-constrained industrial settings [14] [20].
- **Quality Assurance:** Human-in-the-loop and traceability features support high-stakes applications where factual accuracy and auditability are critical [53] [64].

6. Future Research Directions

Open Challenges in Industrial Multi-Agent RAG

- **Empirical Benchmarking:** Need for robust, domain-specific benchmarks covering ROI, efficiency, and security in industrial offer generation [7] [11] [40] [66].
- **Integration with Legacy Systems:** Underexplored challenges in integrating decentralized RAG with enterprise IAM and industrial protocols (e.g., OPC UA) [7] [56].
- **Scalability & Orchestration:** Further research needed on dynamic, scalable orchestration patterns for large-scale, multi-agent industrial deployments [40].

Emerging Trends and Opportunities

- **Multimodal RAG:** Integration of text, tabular, and visual data for richer, more specialized offer generation [17] [67].
- **Advanced Orchestration:** Real-time, utility-maximizing, and agentic orchestration frameworks for complex, cross-domain workflows [67] [68].
- **Secure Data Sharing:** Attribute-based searchable encryption and verifiable retrieval for secure, fine-grained access control [68] [69].

Synthesis:
While multi-agent RAG systems have made significant strides in modularity, security, and efficiency, further research is needed to empirically validate their impact in industrial offer generation, address integration challenges, and explore new modalities and orchestration patterns.

Conclusion

Recent advances in multi-agent, retrieval-augmented generation systems have established a robust foundation for automating complex offer generation in industrial project business. Modular, agentic, and decentralized architectures, advanced communication and orchestration protocols, secure on-premise LLM setups, and multi-layered security frameworks collectively enable scalable, efficient, and compliant solutions. These systems promise substantial improvements in efficiency, quality, and ROI, but further empirical research and integration with industrial standards are essential to fully realize their potential in real-world industrial contexts.

References

1. Retrieval-Augmented Generation Architecture Framework: Harnessing the Power of RAG
Shan, R., Shan, T.. Lecture Notes in Computer Science, 2025.
https://www.scopus.com/pages/publications/85211368920
2. An Architecture and Protocol for Decentralized Retrieval Augmented Generation
Hecking, T., Sommer, T., Felderer, M.. .
https://www.scopus.com/pages/publications/105007945955
3. Revolutionizing Multi-agent Systems: The Role of Agentic RAG in Dynamic Data Ingestion and Real-Time Reasoning
Okorafor, E., Djitog, I., Udechukwu, P., (...), Akanwa, A.. Communications in Computer and Information Science, 2025.
https://www.scopus.com/pages/publications/105014409557
4. Performance Evaluation for Cost-Effective Retrieval Process for Multi-Document Retrieval-Augmented Generation on a Domain-Specific Dataset
Kartiyanta, M.A., Ancilla, E., Jingga, K.. .
https://www.scopus.com/pages/publications/105014324718
5. Bridging Human and AI Decision-Making with LLMs: The RAGADA Approach
Pitkäranta, T., Pitkäranta, L.. International Conference on Enterprise Information Systems, ICEIS - Proceedings, 2024.
https://www.scopus.com/pages/publications/85193974262
6. Learning to Rank for Multiple Retrieval-Augmented Models through Iterative Utility Maximization
Salemi, A., Zamani, H.. .
https://www.scopus.com/pages/publications/105013785419
7. Enhancing Communication and Data Transmission Security in RAG Using Large Language Models
Gummadi, V., Udayaraju, P., Sarabu, V.R., (...), Sarella, S.. .
https://www.scopus.com/pages/publications/85214782970
8. CRUD-RAG: A Comprehensive Chinese Benchmark for Retrieval-Augmented Generation of Large Language Models
Lyu, Y., Li, Z., Niu, S., (...), Chen, E.. ACM Transactions on Information Systems, 2025.
https://www.scopus.com/pages/publications/85219511377
9. Permission-Aware RAG: Identity and Access Management (IAM)-Based Access Filtering in Multi-Resource Environments
Jeong, J., Lee, S.-G.. IEEE Access, 2025.
https://www.scopus.com/pages/publications/105020958905
10. M-RAG: Reinforcing Large Language Model Performance through Retrieval-Augmented Generation with Multiple Partitions
Wang, Z., Teo, S.X., Ouyang, J., (...), Shi, W.. Proceedings of the Annual Meeting of the Association for Computational Linguistics, 2024.

https://www.scopus.com/pages/publications/85204471985
11. Retrieval-Augmented Generation: A Survey of Security Challenges and Countermeasures
Wang, C., Li, H., Song, W., Lin, Y.. .
https://www.scopus.com/pages/publications/105018920964
12. Adversarial threat vectors and risk mitigation for retrieval-Augmented generation systems
Ward, C.M., Harguess, J.. Proceedings of SPIE - The International Society for Optical Engineering, 2025.
https://www.scopus.com/pages/publications/105015295430
13. Enhancing Retrieval and Managing Retrieval: A Four-Module Synergy for Improved Quality and Efficiency in RAG Systems
Shi, Y., Zi, X., Shi, Z., (...), Xu, M.. Frontiers in Artificial Intelligence and Applications, 2024.
https://www.scopus.com/pages/publications/85213366018
14. AdaRAG: Adaptive Optimization for Retrieval Augmented Generation with Multilevel Retrievers at the Edge
Ouyang, T., Hong, G., Zhao, K., (...), Chen, X.. Proceedings - IEEE INFOCOM, 2025.
https://www.scopus.com/pages/publications/105011073862
15. A novel system for strengthening security in large language models against hallucination and injection attacks with effective strategies
Gokcimen, T., Daş, B.. Alexandria Engineering Journal, 2025.
https://www.sciencedirect.com/science/article/pii/S111001682500328X
16. SAMAC-R3-MED: Semantic alignment and multi-agent collaboration of retriever-reranker-responder models for multimodal engineering documents
Li, F., Li, X., Wen, S., (...), Bao, J.. Computers in Industry, 2025.
https://www.scopus.com/pages/publications/105010061299
17. MMA-RAG: Multi-Modal Agents for Insurance Document Processing with Retrieval-Augmented Generation
Krayem, I., Ghourabi, M., Al Assaad, M.. Lecture Notes in Networks and Systems, 2025.
https://www.scopus.com/pages/publications/105017230135
18. M3-RAG: Unified Multimodal and Multilingual Retrieval-Augmented Generation
Xu, J., Zhu, J., Ba, Y.. .
https://www.scopus.com/pages/publications/105014909290
19. Impact of Ensemble of Vector Embeddings on Speculative Retrieval Augmented Generation
Kukreja, S., Kumar, T., Bharate, V., (...), Guha, D.. .
https://www.scopus.com/pages/publications/85216614021
20. RAGuru: A Tool to Create and Automatically Deploy Workload Optimized RAG
Bhowmick, A., Rishikesh, S., Taksande, A., (...), Singhal, R.. .
https://www.scopus.com/pages/publications/105007286644
21. Nursing Retrieval-Augmented Generation: Retrieval augmented generation for nursing question answering with large language models
Xiong, Liping, Zeng, Qiqiao, Luo, Weixiang, Liu, Ronghui. International Journal of Nursing Sciences, 2025.
https://www.sciencedirect.com/science/article/pii/S2352013225001280
22. Research on Retrieval-Augmented Generation Methods Based on Agent Workflows and Applications
Ma, Z., Li, Q.. .
https://www.scopus.com/pages/publications/105017972042
23. Scalable Multimodal RAG Systems: Integrating AI for Adaptive Information Retrieval and Generation
Yuvzhenko, D., Putrenko, V., Serhii, S., Pashynska, N.. CEUR Workshop Proceedings, 2025.

https://www.scopus.com/pages/publications/105017727078
24. Enhancing Interactive Querying with a Multimodal RAG System: Integrating Text, Video, and Document Analysis via LLaMA3
Patel, A., Shivani, R., Usha, N.V., Shruthiba, A.. .
https://www.scopus.com/pages/publications/105015654958
25. Retrieval Augmented Generation on Hybrid Cloud: A New Architecture for Knowledge Base Systems
Chuang, C.-C., Chen, K.-C.. .
https://www.scopus.com/pages/publications/85208097254
26. Beyond Benchmarks: Evaluating Embedding Model Similarity for Retrieval Augmented Generation Systems
Caspari, L., Ghosh Dastidar, K.G., Zerhoudi, S., (...), Granitzer, M.. CEUR Workshop Proceedings, 2024.
https://www.scopus.com/pages/publications/85207503949
27. Retrieval-Augmented Generation: Advancing personalized care and research in oncology
Zarfati, M., Soffer, S., Nadkarni, G.N., Klang, E.. European Journal of Cancer, 2025.
https://www.scopus.com/pages/publications/86000304757
28. Prospects of Retrieval-Augmented Generation (RAG) for Academic Library Search and Retrieval
Bevara, R.V.K., Lund, B.D., Mannuru, N.R., (...), Mannuru, A.. Information Technology and Libraries, 2025.
https://www.scopus.com/pages/publications/105008523692
29. Caching at Scale: Efficiency and Fairness Analysis in Multi-tenant RAG Systems
Ruparel, H., Patel, T.. SN Computer Science, 2025.
https://www.scopus.com/pages/publications/105019085661
30. Leveraging Approximate Caching for Faster Retrieval-Augmented Generation
Bergman, S.A., Ji, Z., Kermarrec, A.-M., (...), de Vos, M.. .
https://www.scopus.com/pages/publications/105003622049
31. The Good and The Bad: Exploring Privacy Issues in Retrieval-Augmented Generation (RAG)
Zeng, S., Zhang, J., He, P., (...), Tang, J.. Proceedings of the Annual Meeting of the Association for Computational Linguistics, 2024.
https://www.scopus.com/pages/publications/85198150173
32. PRESS: Defending Privacy in Retrieval-Augmented Generation via Embedding Space Shifting
He, J., Liu, C., Hou, G., (...), Li, J.. Proceedings - ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing, 2025.
https://www.scopus.com/pages/publications/105003870875
33. Is My Data in Your Retrieval Database? Membership Inference Attacks Against Retrieval Augmented Generation
Anderson, M., Amit, G., Goldsteen, A.. International Conference on Information Systems Security and Privacy, 2025.
https://www.scopus.com/pages/publications/105001708399
34. Streamrag: A Lock-Aware and Traffic-Aware Query Coordinator in Stream-Based Rag Systems
Jeong, Y., Park, K., Park, S.. .
https://www.scopus.com/pages/publications/105010822627
35. Multi_semantic RAG: What's not Important is Important
Jiang, M., Zou, L., Lu, Y., (...), Qin, P.. Communications in Computer and Information Science, 2025.
https://www.scopus.com/pages/publications/105012433456
36. CRP-RAG: A Retrieval-Augmented Generation Framework for Supporting Complex Logical Reasoning and Knowledge Planning

Kehan, X., Kun, Z., Jingyuan, X., Wei, H.. Electronics (Switzerland), 2025.
https://www.scopus.com/pages/publications/85214446183
37. MASC: Large language model-based multi-agent scheduling chain for flexible job shop scheduling problem
Wang, Z., Wan, C., Liu, J., (...), Hu, Z.. Advanced Engineering Informatics, 2025.
https://www.sciencedirect.com/science/article/pii/S1474034625004203
38. Agentic AI: The age of reasoning—A review
Nisa, Ume, Shirazi, Muhammad, Saip, Mohamed Ali, Pozi, Muhammad Syafiq Mohd. Journal of Automation and Intelligence, 2025.
https://www.sciencedirect.com/science/article/pii/S2949855425000516
39. Semantic and modular orchestration of AI-driven digital twins for industrial interoperability and optimization
Juarez Juarez, M.G.J., Giret, A., Botti, V.. Journal of Industrial Information Integration, 2025.
https://www.sciencedirect.com/science/article/pii/S2452414X25001827
40. Streamline automated biomedical discoveries with agentic bioinformatics
Zhou, Juexiao, Jiang, Jindong, Han, Zhongyi, (...), Gao, Xin. Briefings in Bioinformatics, 2025.
https://www.sciencedirect.com/science/article/pii/S1477405425001681
41. A Pre-Processing Framework for Securing LLM-RAG Interfaces Against Information Leakage
Davies, R.H., Sanghvi, K., Nalam, R., Ramnath, R.. .
https://www.scopus.com/pages/publications/105013077199
42. The Effect of Chunk Size on the RAG Performance
Hladěna, J., Šteflovič, K., Čech, P., (...), Zvackova, A.. Lecture Notes in Networks and Systems, 2025.
https://www.scopus.com/pages/publications/105014412546
43. DropMicroFluidAgents (DMFAs): autonomous droplet microfluidic research framework through large language model agents
Nguyen, Dinh-Nguyen, Tong, Raymond Kai-Yu, Dinh, Ngoc-Duy. Digital Discovery, 2025.
https://www.sciencedirect.com/science/article/pii/S2635098X25001925
44. Approaches to automatic discovery and modeling of Industrial Assets for IT/OT Integration
Todkar, A., Sarkar, M., Solanki, J., Tylka, J.. IEEE International Conference on Automation Science and Engineering, 2025.
https://www.scopus.com/pages/publications/105018305266
45. Seamless Integration of Legacy Industrial Systems with OPC UA for Enhanced Digital Transformation
Sujith, R., Ganapathy, V.S., Ilayaraja, R., (...), Vignesh, A.. .
https://www.scopus.com/pages/publications/85215507691
46. Learning Energy-Efficient MAC Protocols Using MARL in Industrial Networks
Schilirò, A., Miuccio, L., Riolo, S., Panno, D.. .
https://www.scopus.com/pages/publications/105012250352
47. Retrieval Poisoning Attacks Based on Prompt Injections into Retrieval-Augmented Generation Systems that Store Generated Responses
Anichkov, Y., Popov, V., Bolovtsov, S.. Lecture Notes in Computer Science, 2025.
https://www.scopus.com/pages/publications/86000240690
48. SafeRAG: Benchmarking Security in Retrieval-Augmented Generation of Large Language Model
Xun, X., Niu, S., Li, Z., (...), Wang, M.. Proceedings of the Annual Meeting of the Association for Computational Linguistics, 2025.
https://www.scopus.com/pages/publications/105021050516

49. Secure Multifaceted-RAG: Hybrid Knowledge Retrieval with Security Filtering
Byun, G., Lee, S., Choi, N., Choi, J.D.. Information (Switzerland), 2025.
https://www.scopus.com/pages/publications/105017004866
50. Agentic RAG with Human-in-the-Retrieval
Xu, X., Zhang, D., Liu, Q., (...), Zhu, L.. .
https://www.scopus.com/pages/publications/105007869177
51. Traceback of Poisoning Attacks to Retrieval-Augmented Generation
Zhang, B., Xin, H., Fang, M., (...), Liu, Z.. .
https://www.scopus.com/pages/publications/105005159200
52. Robust Fine-tuning for Retrieval Augmented Generation against Retrieval Defects
Tu, Y., Su, W., Zhou, Y., (...), Ai, Q.. .
https://www.scopus.com/pages/publications/105011826359
53. Leveraging MDS2 and SBOM data for LLM-assisted vulnerability analysis of medical devices
Stein, S., Pilgermann, M., Weber, S., Sedlmayr, M.. Computational and Structural Biotechnology Journal, 2025.
https://www.sciencedirect.com/science/article/pii/S2001037025002788
54. Large-language models: The game-changers for materials science research
Yu, Songlin, Ran, Nian, Liu, Jianjun. Artificial Intelligence Chemistry, 2024.
https://www.sciencedirect.com/science/article/pii/S2949747724000344
55. Effectiveness of retrieval augmented generation-based large language models for generating construction safety information
Uhm, M., Kim, J., Ahn, S., (...), Kim, H.. Automation in Construction, 2025.
https://www.sciencedirect.com/science/article/pii/S0926580524006629
56. UAVs meet LLMs: Overviews and perspectives towards agentic low-altitude mobility
Tian, Y., Lin, F., Li, Y., (...), Wang, F.-Y.. Information Fusion, 2025.
https://www.sciencedirect.com/science/article/pii/S1566253525002313
57. A comprehensive survey on integrating large language models with knowledge-based methods
Yang, W., Some, L., Bain, M., Kang, B.. Knowledge-Based Systems, 2025.
https://www.sciencedirect.com/science/article/pii/S0950705125005490
58. Empowering knowledge graphs with hybrid retrieval-augmented generation for the intelligent mix scheme of mass concrete
Shang, Y., Ke, Z., Lin, P., (...), Tan, S.. Case Studies in Construction Materials, 2025.
https://www.sciencedirect.com/science/article/pii/S2214509525007776
59. Intelligent Chinese patent medicine (CPM) recommendation framework: Integrating large language models, retrieval-augmented generation, and the largest CPM dataset
Qin, S., Wang, Y., Cui, T., (...), Li, H.. Pharmacological Research, 2025.
https://www.sciencedirect.com/science/article/pii/S1043661825003081
60. Prompt Compression based on Key-Information Density
Lin, Y., Guo, W., Zhang, Y., (...), Li, Z.. Expert Systems with Applications, 2025.
https://www.sciencedirect.com/science/article/pii/S0957417425013600
61. Generative knowledge-guided review system for construction disclosure documents
Xiao, H., Zhuang, J., Yang, B., (...), Lai, S.. Advanced Engineering Informatics, 2025.
https://www.sciencedirect.com/science/article/pii/S1474034625005117
62. GINGER: Grounded Information Nugget-Based Generation of Responses
Lajewska, W., Balog, K.. .

https://www.scopus.com/pages/publications/105011825178
63. Multi-agent large language model framework for code-compliant automated design of reinforced concrete structures
Chen, J., Bao, Y.. Automation in Construction, 2025.
https://www.sciencedirect.com/science/article/pii/S0926580525003711
64. Reasoning beyond limits: Advances and open problems for LLMs
Ferrag, M.A., Tihanyi, N., Debbah, M.. ICT Express, 2025.
https://www.sciencedirect.com/science/article/pii/S240595952500133X
65. Investigating the Performance of Retrieval-Augmented Generation and Domain-Specific Fine-Tuning for the Development of AI-Driven Knowledge-Based Systems
Lakatos, R., Pollner, P., Hajdu, A., Joó, T.. Machine Learning and Knowledge Extraction, 2025.
https://www.scopus.com/pages/publications/105001169264
66. Graph retrieval augmented large language models for facial phenotype associated rare genetic disease
Song, Jie, Xu, Zhichuan, He, Mengqiao, (...), Shen, Bairong. npj Digital Medicine, 2025.
https://www.sciencedirect.com/science/article/pii/S2398635225000469
67. OMGM: Orchestrate Multiple Granularities and Modalities for Efficient Multimodal Retrieval
Yang, W., Fu, J., Wang, R., (...), Bian, J.. Proceedings of the Annual Meeting of the Association for Computational Linguistics, 2025.
https://www.scopus.com/pages/publications/105021058441
68. SE-COLLAB: Achieving Fine-Grained and Efficiently Verifiable Searchable Encryption With Boolean Multi-Keyword Search for Collaborative IIoT Data Sharing
Fugkeaw, S., Deevijit, J.. IEEE Access, 2025.
https://www.scopus.com/pages/publications/105011165037
69. Securing the IoT wireless communication system: a multi-access physical layer approach
Katariya, J.K., Manjunath, H.R., Malviya, A., Kumar, R.. Multidisciplinary Science Journal, 2024.
https://www.scopus.com/pages/publications/85203374462