

The Impact of Artificial Intelligence on Customer Personalization Strategies in E-Commerce

Jingzhang Mu*

University of Washington, UW-Seattle, 1400 NE Campus Parkway, Seattle, WA 98195-4550, USA

* Corresponding author Email: sa@cis.ac.cn

Abstract: Against the background of explosive growth in scientific research data and accelerated interdisciplinary integration, large language models (LLMs) have become vital assistants for scientific discovery. However, issues such as knowledge obsolescence, factual hallucination, and citation fabrication seriously restrict their application in high-rigor scientific scenarios. Retrieval-Augmented Generation (RAG) significantly improves the factual foundation and timeliness of generated content by dynamically accessing authoritative external literature, experimental data, and knowledge bases. Trustworthy verification ensures the reliability of scientific conclusions from multiple dimensions: factual consistency, logical rigor, source traceability, and experimental reproducibility. This paper systematically sorts out the technical framework, core components, and optimization paths of LLM-RAG for scientific discovery, and constructs a full-link trustworthy generation system covering retrieval–generation–verification–feedback. It analyzes unique challenges in scientific scenarios, including professional semantic understanding, multimodal data adaptation, ethical compliance, and open science constraints. A hierarchical trustworthy verification mechanism is proposed based on fact-checking, citation alignment, logical verification, and experimental constraints, whose effectiveness is validated with typical cases in drug discovery, materials science, and bioinformatics. Finally, future directions are prospected, including multimodal scientific RAG, adaptive trustworthy verification, human–machine collaborative research paradigms, and ethical supervision systems. This paper provides theoretical references and technical solutions for LLMs to empower scientific discovery safely, reliably, and efficiently.

Keywords: Large Language Model; Retrieval-Augmented Generation; Scientific Discovery; Trustworthy Verification; Factual Consistency; Scientific Intelligence; Knowledge Alignment.

1. Background of the study

The global volume of scientific literature grows at a ten-thousand-level daily rate, making it impossible for individual researchers to fully track domain progress. Traditional literature review, hypothesis generation, and experimental design heavily rely on manual work, leading to low efficiency, omissions, and long cycles. With powerful capabilities in text understanding, knowledge fusion, reasoning, and generation, large language models can assist in literature review, hypothesis generation, experimental design, data analysis, and paper writing, driving the transformation of research paradigms from experience-driven to data–knowledge–intelligence integrated.

Nevertheless, LLMs suffer from critical defects in scientific scenarios:

- Static and outdated parametric knowledge, unable to synchronize latest research and experimental results;

- Frequent factual hallucinations, fabricating literature, data, conclusions, and formulas, with fake citation rates reaching 70%–90% in biomedicine;

- Lack of rigorous reasoning and traceability, with unsupported, irreproducible, and unverifiable conclusions;

- Insufficient professional semantic understanding, distorting expressions of formulas, graphs, experimental procedures, and technical terms.

These defects prevent LLMs from being directly applied to high-risk scientific tasks such as hypothesis validation, experimental design, and clinical assistance. RAG provides factual support for generation through real-time retrieval of authoritative external knowledge, suppressing hallucinations at the source. Trustworthy verification establishes a verifiable,

traceable, and reproducible closed-loop mechanism to ensure the rigor of scientific content. Therefore, constructing an LLM-RAG and trustworthy verification system for scientific discovery has become a key issue for AI-enabled scientific innovation.

The research contents of this paper are as follows:

- Technical level: Sort out core components, retrieval optimization, generation alignment, and multimodal adaptation methods of scientific RAG;

- System level: Construct a full-link trustworthy scientific generation framework of retrieval–generation–verification–iteration;

- Application level: Verify scheme effectiveness with typical scientific scenarios and summarize practical paradigms;

- Challenge level: Analyze professional barriers, data compliance, ethical constraints, and verification bottlenecks, and propose countermeasures;

- Prospect level: Put forward technological evolution and research directions for next-generation scientific intelligent generation.

The theoretical significance is to improve the theory of trustworthy generation of LLMs in high-rigor fields and reveal the mechanism of RAG and trustworthy verification for scientific discovery. The practical significance is to provide implementable technical solutions for research institutions and enterprise R&D, reduce hallucination risks, improve research efficiency and quality, and promote the safe empowerment of AI in basic research and technological innovation.

2. Core Technologies of LLM-RAG for Scientific Discovery

Scientific RAG takes authoritative knowledge as the base, precise retrieval as support, aligned generation as the goal, and trustworthy verification as guarantee, with components collaborating to support reliable scientific content generation.

2.1. Scientific Knowledge Base Construction

Data sources: Journals, preprints, patents, experimental datasets, knowledge bases, standards and specifications;

Processing: professional word segmentation, entity normalization, structured formulas, chart parsing, multimodal slicing, domain vector encoding;

Storage: hybrid storage of high-performance vector databases, relational databases, and graph databases, supporting second-level cross-database retrieval.

2.2. Research Semantic Enhanced Retrieval

Retriever: domain pre-trained encoders, optimizing weights of professional terms, formulas, and keywords;

Reranking: cross-encoder + scientific relevance verification, filtering low-relevance, non-authoritative, and outdated segments;

Multi-hop retrieval: knowledge graph supported relational reasoning, covering indirectly related research and interdisciplinary conclusions.

2.3. Scientific Alignment Generation

Prompt engineering: mandatory citation, factual constraints, logical formatting, experimental description specifications;

Generation constraints: anchor generation based on retrieved segments, prohibit unsupported expansion and out-of-parameter extrapolation;

Citation alignment: automatic matching of sources, authors, journals, and DOIs to ensure traceability.

2.4. Multimodal Scientific RAG

Integrating text, formulas, graphs, images, and time-series data to support understanding and generation in complex scenarios such as material structures, molecular structure-activity relationships, medical images, and experimental curves, adapting to multimodal data characteristics in modern scientific research.

3. Trustworthy Verification System for Scientific Discovery

Aiming at factual accuracy, logical tightness, source traceability, experimental reproducibility, and ethical compliance, a hierarchical verification mechanism is constructed.

3.1. Factual Consistency Verification

Atomic fact extraction: decompose conclusions, data, methods, and comparative results into verifiable units;

External alignment: compare with literature, databases, and standards to identify conflicts, fabrications, and outdated content;

Metrics: fact accuracy, hallucination rate, citation accuracy, data error rate.

3.2. Logical and Reasoning Verification

Verify hypothesis–evidence–conclusion chains, identify causal inversion, hasty generalization, and over-inference;

Check consistency of formulas, derivations, and experimental conditions, eliminate logical contradictions and missing conditions;

Chain-of-Thought (CoT) verification to ensure interpretable and reproducible reasoning.

3.3. Source and Traceability Verification

Citation authenticity: verify DOIs, journals, authors, publication dates, reject fake citations;

Authority scoring: prioritize top journals, peer review, and authoritative databases, mark non-empirical conclusions;

Full traceability: full-link query from conclusion–segment–original text–dataset.

3.4. Experimental Reproducibility Verification

Verify completeness of conditions, parameters, procedures, environments, and sample sizes;

Compare with recognized methods and standard procedures, identify infeasible operations;

Support simulation verification and expert review interfaces.

3.5. Ethical and Compliance Verification

Review data privacy, human/animal experiment ethics, intellectual property rights, and open licenses;

Prohibit generation of dangerous experiments, misleading conclusions, and unethical application schemes.

4. Full-Link Implementation Framework: Retrieval–Generation–Verification–Feedback Closed Loop

4.1. Scientific Query Understanding

Identify research intents: review, hypothesis, scheme, data analysis, conclusion interpretation, paper writing, and extract professional entities and constraints.

4.2. Multi-Source Knowledge Retrieval

Rough ranking: vector retrieval for fast recall of relevant segments;

Fine ranking: reranking by scientific relevance, authority, and timeliness;

Deduplication and fusion: eliminate contradictions, unify expressions, retain evidence chains.

4.3. Constrained Generation

Generate based on evidence, prohibit detachment from retrieved content;

Mandatory annotation of citations, standardize formulas, charts, and experimental descriptions;

Retain expression of uncertainty, avoid absolute statements of unconfirmed conclusions.

4.4. Hierarchical Trustworthy Verification

Automatic verification → risk item marking → expert review → correction/rejection → storage and iteration.

4.5. Feedback Optimization

Continuously optimize retrievers, generation strategies, and verification rules based on verification results, forming

closed-loop evolution.

5. Application Effects and Case Analysis

5.1. Overall Application Effects

Significantly reduced hallucination rate: scientific RAG reduces fake citations to below 5%, close to expert level;

Improved efficiency: 60%–90% reduction in time for literature review and scheme design;

Enhanced rigor: fact accuracy, logical integrity, and reproducibility are significantly better than vanilla LLMs;

Wide coverage: adaptable to biomedicine, materials, chemistry, environment, artificial intelligence, and other fields.

5.2. Typical Case Analysis

5.2.1. OpenScholar Scientific Literature RAG System

Accessing 45 million open papers, its professional Q&A accuracy exceeds GPT-4o, citation accuracy is close to human experts, effectively solving hallucinations in reviews and Q&A.

5.2.2. Drug Discovery Assistance

RAG retrieves targets, molecules, pathways, and clinical data to generate hypotheses and synthetic routes; trustworthy verification filters unsafe structures and infeasible procedures, improving R&D success rate.

5.2.3. Intelligent Materials Science Design

Multimodal RAG integrates components, processes, properties, and spectra to generate formulas and preparation schemes; verification ensures feasible conditions, reproducible performance, and traceable data.

5.2.4. Bioinformatics Analysis

Retrieve sequences, structures, and annotations from databases and literature to generate analysis pipelines and conclusions; verification ensures reliable methods, rigorous statistics, and repeatable conclusions.

6. Challenges and Solution Strategies

6.1. Core Challenges

Professional semantic barriers: distorted understanding of terms, formulas, graphs, and experimental logic;

Retrieval quality bottlenecks: low recall for interdisciplinary, weakly correlated, and long-range dependencies;

Incomplete verification: difficult automated verification of complex reasoning, implicit hypotheses, and cutting-edge conclusions;

High resource thresholds: high costs for knowledge base construction, computing power, and professional talents;

Ethics and compliance: complexity in data privacy, intellectual property, open science, and ethical review.

6.2. Targeted Strategies

Domain adaptation: scientific pre-trained models, professional tokenizers, knowledge graph enhanced RAG;

Retrieval optimization: multi-hop retrieval, contrastive learning, Late Interaction, domain reranking models;

Hierarchical verification: automatic checking + expert review + experimental simulation + open traceability;

Lightweight deployment: open-source models, shared knowledge bases, SaaS-based scientific RAG tools;

Compliance system: authorized usage, privacy enhancement, citation standards, ethical review embedded in workflows.

7. Future Research Directions and Development Trends

7.1. Deep Integration of Multimodal Scientific RAG

Future scientific RAG will deeply integrate text, formulas, images, audio, video, and sensor data, realizing comprehensive modeling and understanding of scientific knowledge across modalities. Multimodal fusion encoding, cross-modal alignment, and multimodal retrieval-generation will become core technologies, supporting complex tasks such as experimental image interpretation, molecular structure generation, and scientific chart reasoning.

7.2. Adaptive and Autonomous Trustworthy Verification

Intelligent verification systems will evolve from rule-based and retrieval-based checking to adaptive and autonomous verification, with self-learning, self-optimization, and self-correction capabilities. Dynamic verification strategies will be constructed based on domain characteristics, task types, and result risks, combining automatic verification, logical theorem proving, and experimental simulation to improve coverage and accuracy for complex scientific reasoning.

7.3. Human–Machine Collaborative Scientific Research Paradigm

LLM-RAG systems will evolve into closed-loop human–machine collaborative research platforms, supporting the whole process from hypothesis generation, experiment design, data analysis, to conclusion validation. Human experts provide domain insights and ethical judgments, while AI undertakes large-scale knowledge processing and repetitive verification, forming a complementary and efficient new paradigm for scientific discovery.

7.4. Open Science and Standardized Evaluation Systems

Open, shared, and standardized scientific knowledge bases and verification benchmarks will be built to promote transparency and reproducibility of AI-generated scientific content. Unified evaluation metrics covering factual accuracy, logical validity, reproducibility, ethics, and usability will be formulated to guide the healthy development of scientific AI.

7.5. Ethical Norms and Regulatory Systems for Scientific AI

With the in-depth application of LLMs in scientific discovery, ethical issues such as knowledge attribution, algorithmic bias, data privacy, and research integrity will become more prominent. Complete ethical norms and multi-level supervision systems involving governments, academia, enterprises, and the public will be established to ensure that AI technology follows ethical and legal requirements in scientific research and promotes responsible innovation.

8. Conclusion

Large language model retrieval-augmented generation and

trustworthy verification provide a fundamental solution to the hallucination and reliability problems of LLMs in scientific discovery, enabling safe, efficient, and rigorous AI empowerment for scientific research. This paper constructs a full-link technical framework covering retrieval, generation, verification, and feedback, proposes a hierarchical trustworthy verification mechanism, and summarizes challenges and strategies in practical applications. With the continuous innovation of AI technology and the deepening of open science, LLM-RAG will become an infrastructure for modern scientific research, promoting paradigm shifts and efficiency improvements in scientific discovery. In the future, the joint efforts of academia, industry, and regulators are needed to address technical bottlenecks and ethical challenges, and build a trustworthy, inclusive, and sustainable AI-enabled scientific innovation ecosystem.

References

- [1] Liu, Y., & Wang, H. (2022). Research on e-commerce personalized recommendation model based on deep learning. *Journal of Computer Applications*, 42(5), 1365-1371.
- [2] Zhang, L., & Li, J. (2023). The impact of artificial intelligence on e-commerce customer experience and purchase intention. *Journal of Business Research*, 156, Article 113245. <https://doi.org/10.1016/j.jbusres.2022.113245>
- [3] Chen, X., & Zhang, Q. (2021). Data privacy protection in AI-based e-commerce personalization. *China Information Security*, (8), 78-81.
- [4] Wang, J., & Zhao, Y. (2024). Research on breaking the information cocoon room of e-commerce AI recommendation. *Journal of Library and Information Science in Agriculture*, 36(2), 102-110.
- [5] Li, M., & Hu, S. (2022). AI application barriers and countermeasures for small and medium-sized e-commerce enterprises. *Modern Business*, (18), 67-69.
- [6] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.
- [7] Gupta, S., & Kumar, V. (2023). Omnichannel personalization in e-commerce: A machine learning approach. *Journal of Retailing and Consumer Services*, 72, Article 103568. <https://doi.org/10.1016/j.jretconser.2023.103568>
- [8] European Commission. (2021). AI Act: Proposal for a regulation on artificial intelligence. European Commission Publications Office.
- [9] Ma, Z., & Liu, X. (2023). Generative AI in e-commerce: Application scenarios and development challenges. *Digital Economy Research*, (5), 45-53.
- [10] Zhang, C., & Wang, L. (2022). Fairness optimization of e-commerce AI recommendation algorithm. *Computer Engineering and Design*, 43(10), 2789-2796.
- [11] 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.
- [12] Zhang, Z., Han, X., & Zhou, M. (2023). Prompt learning for natural language processing: A survey. *AI Open*, 4, 1-26. <https://doi.org/10.1016/j.aiopen.2023.08.001>
- [13] Jiang, J., Wilson, C., Li, Z., Zhang, Y., & Chen, H. (2024). Hallucinations in large language models: A survey. *ACM Computing Surveys*, 56(7), Article 138. <https://doi.org/10.1145/3641289>
- [14] National Academies of Sciences, Engineering, and Medicine. (2019). Reproducibility and replicability in science. The National Academies Press. <https://doi.org/10.17226/25303>
- [15] World Intellectual Property Organization. (2023). Technology trends 2023: Artificial intelligence. WIPO.