# Retrieval-Augmented Generation for Scientific Knowledge Discovery in Interdisciplinary Research Networks

Bhrisher Marshall
Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA.
bhriphermarshall88@missouri.edu

Felix Neal
Department of Computer Science, George Mason University, Fairfax, VA, USA.
felix.neal419@gmu.edu

Brent L. Little
School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, USA.
brentllittle464@oregonstate.edu

**Abstract**

The accelerating pace of scientific discovery increasingly depends on the ability to integrate knowledge across disparate disciplines, yet traditional literature search and synthesis methods remain fragmented and labor-intensive. Retrieval-Augmented Generation (RAG) offers a transformative paradigm for scientific knowledge discovery by combining large language models with external knowledge bases, enabling context-aware, evidence-grounded responses that can bridge disciplinary boundaries. This paper presents a comprehensive examination of RAG systems within the context of interdisciplinary research networks, focusing on architectural design, structural trade-offs, governance challenges, and deployment infrastructure. We analyze the interplay between retrieval fidelity and generative coherence, the role of knowledge graph integration, and the socio-technical implications of scaling such systems across heterogeneous domains. Particular attention is given to issues of fairness, robustness, and sustainability, as well as the policy frameworks required to support responsible adoption. Through case illustrations from fields such as biomedicine, climate science, and materials engineering, we highlight both the potential and the current limitations of RAG for facilitating cross-domain synthesis. The paper concludes with forward-looking recommendations for system governance, benchmark development, and interdisciplinary collaboration that can guide future research and deployment of RAG-driven knowledge discovery platforms.

**Keywords**

Retrieval-Augmented Generation, scientific knowledge discovery, interdisciplinary research, large language models, knowledge graphs, system architecture, information retrieval, fairness, sustainability, socio-technical systems.

## 1. Introduction

Modern scientific inquiry is increasingly characterized by a need for cross-disciplinary synthesis. Problems such as climate change, pandemic response, and sustainable energy demand insights from fields as diverse as biology, economics, materials science, and computer science. However, the sheer volume of literature and the specialized vocabularies of individual disciplines make it difficult for researchers to locate, comprehend, and integrate relevant knowledge from outside their immediate expertise. Conventional search engines and citation databases provide only shallow retrieval, often failing to capture conceptual relationships that span disciplinary boundaries. In response, the artificial intelligence community has developed Retrieval-Augmented Generation (RAG), a framework that couples a retrieval module with a generative language model to produce answers grounded in externally sourced documents. RAG systems offer a promising avenue for scientific knowledge discovery, as they can dynamically access large corpora, synthesize evidence from multiple sources, and generate coherent explanations that bridge epistemological divides [1]. This paper examines RAG from a systems-level perspective, emphasizing not only the technical architecture but also the structural trade-offs, governance challenges, and deployment realities that shape its utility in interdisciplinary research networks. We argue that realising the full potential of RAG requires a careful balance between retrieval precision and generative flexibility, as well as proactive attention to fairness, robustness, and long-term sustainability. By situating RAG within the broader context of socio-technical infrastructure, we aim to provide researchers and practitioners with a holistic understanding of both the opportunities and the pitfalls inherent in using these systems for knowledge discovery across disciplinary boundaries.

## 2. Background and Related Work

The concept of augmenting language models with external knowledge is not new. Early work in neural semantic parsing and open-domain question answering demonstrated the value of retrieving relevant passages before generating an answer [2]. The seminal RAG architecture introduced by Lewis et al. formalised this two-stage process, using a dense retriever to index a large corpus and a seq2seq model to generate outputs conditioned on retrieved passages [1]. Subsequent research extended this framework through improved retrieval mechanisms, such as dense passage retrieval (DPR) [3] and the integration of knowledge graphs [4]. In parallel, the development of large-scale pretrained language models like GPT-3 [5] and T5 [6] provided the generative backbone for RAG systems, enabling fluent and contextually relevant responses. More recently, self-reflective RAG variants have incorporated iterative retrieval and critique loops to enhance factual accuracy [7]. Within the scientific domain, specialized implementations such as SciBERT [8] and SPECTER [9] have demonstrated the effectiveness of domain-adapted embeddings for literature retrieval. However, most existing work focuses on single-discipline corpora or general knowledge tasks, leaving the interdisciplinary use case underexplored. The challenge of linking heterogeneous data sources, reconciling disparate terminologies, and maintaining epistemological coherence across fields remains largely unaddressed. This paper builds on these foundations by examining RAG not merely as a technical artifact but as a component of a larger socio-technical system that must be designed for cross-domain integration.

## 3. System Architecture and Design Considerations

A typical RAG system for scientific knowledge discovery consists of three principal components: a retrieval module, a generation module, and a knowledge base. The retrieval module indexes documents from a potentially vast and heterogeneous collection of scientific

papers, preprints, datasets, and grey literature. Dense retrievers, based on bi-encoder architectures trained to maximize similarity between query and document embeddings, have become the standard due to their ability to capture semantic relatedness beyond exact keyword matches [3]. However, when the knowledge base spans multiple disciplines, the embedding space must accommodate domain-specific meanings of terms; a phrase like "stress" can refer to mechanics, psychology, or plant biology. This polysemy degrades retrieval precision unless mitigated by domain-aware embeddings or additional context injection [10]. The generation module, typically a decoder-only language model, takes the retrieved passages and the original query to produce a synthesized answer. The interplay between retrieval and generation introduces a design tension: high retrieval recall may flood the generator with irrelevant passages, degrading coherence, while high precision may miss critical cross-domain connections. One structural approach to managing this trade-off is to incorporate a knowledge graph layer that explicitly represents entities, relationships, and hierarchical concepts across disciplines [4]. By grounding retrieval in a structured ontology, the system can better navigate interdisciplinary boundaries. For example, a query about "carbon sequestration" can be mapped to nodes in a graph linking soil science, atmospheric chemistry, and policy studies, enabling targeted retrieval from each domain. Another design consideration is the choice between single-pass and iterative retrieval. Iterative methods allow the generator to issue follow-up queries based on partial results, which can uncover non-obvious links but also increase computational cost and latency [7]. The system's overall architecture must also account for update frequency: scientific knowledge evolves rapidly, and static indices can quickly become stale. A sustainable deployment requires mechanisms for continuous ingestion of new publications, versioning of indices, and retraining of retrieval encoders at appropriate intervals. These design decisions collectively determine the system's capacity to support robust and equitable knowledge discovery across interdisciplinary networks.

## 4. Trade-offs in Retrieval and Generation Pipelines

The optimization of RAG pipelines involves multiple trade-offs that become more acute in interdisciplinary settings. The first trade-off pertains to retrieval depth versus breadth. A narrow retrieval strategy focusing on highly cited papers may reinforce established viewpoints, while a broad strategy that includes preprints and conference proceedings may introduce noise and unvetted findings. Research from the information retrieval community suggests that diversity-aware retrieval, which explicitly samples from different disciplinary clusters, can improve the coverage of cross-domain insights without sacrificing relevance [11]. Another critical trade-off lies between factual grounding and generative creativity. A conservative generator that closely paraphrases retrieved passages may produce factually accurate but stilted responses, whereas a more creative generator may hallucinate connections that have no basis in the literature. In scientific discovery, where novelty is valued, a certain degree of generative synthesis is desirable, but it must be accompanied by transparent attribution. The RAG system can be designed to output explicit citations for each claim, as demonstrated in the Asai et al. self-RAG model [7]. However, citation generation itself introduces overhead and potential errors when the retrieved passages are misinterpreted. Furthermore, the computational cost of retrieval grows with corpus size; indexing an interdisciplinary corpus comprising millions of documents can be prohibitively expensive for smaller institutions. Compression techniques such as approximate nearest neighbour search (ANNS) reduce retrieval time but may sacrifice recall [12]. The choice between a centralized, cloud-based infrastructure and decentralised, on-premises deployment also affects both cost

and data governance. For sensitive or proprietary research, on-premises deployment may be necessary to comply with data-sharing policies. Lastly, the evaluation of RAG systems in interdisciplinary contexts is inherently difficult because there is no single gold-standard answer. Traditional metrics like exact match or F1 score fail to capture the richness of a synthesized cross-domain response. More nuanced evaluation frameworks that incorporate expert human judgment and domain-specific rubrics are needed [13]. These trade-offs underscore the importance of designing RAG pipelines with flexibility and user control, allowing researchers to adjust parameters such as retrieval breadth, generation temperature, and citation strictness according to the specific discovery task.

## 5. Governance, Fairness, and Robustness

As RAG systems become integrated into research workflows, issues of governance, fairness, and robustness demand systematic attention. Governance refers to the policies and practices that regulate how the system is built, updated, and accessed. In an interdisciplinary research network, multiple stakeholders—including funding agencies, academic institutions, publishers, and individual researchers—have divergent interests regarding data provenance, intellectual property, and access rights. A RAG system that ingests copyrighted material without proper licensing could expose its operators to legal risk. Therefore, governance frameworks must include clear terms for content acquisition, indexing restrictions, and citation of original sources. Additionally, the system's outputs should be auditable: researchers need to be able to trace a generated statement back to the specific documents that influenced it. This transparency is essential not only for trust but also for identifying potential biases [14]. Fairness concerns arise because the training data and retrieval corpora may underrepresent certain disciplines, methodologies, or global regions. For instance, a RAG system that primarily indexes English-language journals from North American and European sources will systematically marginalize research conducted in other languages or published in regional venues. This can skew knowledge discovery toward existing power structures and away from innovative perspectives emerging from the Global South [15]. Mitigation strategies include explicit sampling to ensure coverage across languages and publication venues, as well as the inclusion of grey literature and indigenous knowledge systems where appropriate. Robustness is another pillar: RAG systems must remain reliable in the face of distributional shifts, such as the introduction of a new paradigm that contradicts earlier findings. The retriever may continue to surface outdated or refuted studies unless a freshness signal is incorporated. Moreover, adversarial attacks—where deliberately misleading documents are added to the knowledge base—can poison the retrieved context and cause the generator to produce fabricated or harmful answers [16]. Defences include retrieval verification pipelines, anomaly detection on incoming documents, and the use of trusted curation layers. Ultimately, governance, fairness, and robustness are not merely technical issues but are deeply entangled with the values and priorities of the research community. Institutional policies that mandate periodic audits, diversity metrics, and open-source development can help align RAG deployment with the broader goals of equitable and trustworthy science.

## 6. Deployment and Infrastructure Challenges

Translating a RAG system from a research prototype to an operational platform serving interdisciplinary researchers requires careful consideration of infrastructure. Latency is a primary concern: scientific queries often require real-time or near-real-time responses, especially when used in interactive exploration tools. The retrieval stage, particularly with

large-scale indices, can introduce delays of several seconds, which may be unacceptable for users who expect conversational interaction. Strategies to reduce latency include caching frequent queries, using pre-computed embeddings, and deploying GPU-accelerated retrieval servers [12]. Scalability is another challenge. An interdisciplinary knowledge base may grow at a rate of millions of new documents per year, necessitating a distributed indexing architecture that can scale horizontally. Cloud-based solutions offer elasticity but raise concerns about data sovereignty and recurring costs. For institutions with limited budgets, a hybrid approach that combines a local cache of high-priority documents with cloud retrieval for broader coverage may be viable. Energy consumption is a growing consideration, as the training and inference of large language models are notoriously carbon-intensive [17]. Running a continuous retrieval and generation pipeline amplifies this footprint. Researchers are exploring techniques such as model distillation and sparse retrieval to reduce computational demands. The deployment environment must also support continuous monitoring and updates. Scientific consensus evolves, and the generator's parameters may need periodic fine-tuning to reflect new knowledge. A versioned deployment pipeline, akin to continuous integration/continuous deployment (CI/CD) in software engineering, can manage these updates while maintaining system stability. User experience design is equally important; researchers from different disciplines have varying levels of technical literacy and expectations for how a RAG system presents results. A unified interface that allows for discipline-specific customization, such as adjustable vocabulary filters or output formats, can enhance adoption. Finally, interoperability with existing research tools, such as reference managers, data repositories, and collaboration platforms, can embed RAG capabilities seamlessly into researchers' daily workflows.

## 7. Case Illustrations and Cross-Domain Applications

To ground the preceding discussion, we consider several illustrative cases where RAG systems have been applied or proposed for interdisciplinary knowledge discovery. In biomedicine, the integration of clinical trial data, genomic databases, and literature from molecular biology and pharmacology poses a classic interdisciplinary challenge. RAG systems designed for drug repurposing have successfully retrieved relevant studies from disparate domains to suggest candidate compounds for new indications [18]. These systems rely on a knowledge graph connecting genes, proteins, diseases, and drugs, and they generate explanations that synthesize evidence from multiple experimental contexts. In climate science, researchers have built RAG prototypes that combine atmospheric data, policy documents, and economic models to answer questions about the impact of mitigation strategies. The retrieval component must navigate heterogeneous data formats, including numerical simulation outputs and qualitative case studies. Challenges arise in normalizing the units of measurement and in reconciling conflicting predictions from different models. In materials engineering, a RAG system trained on a corpus of journal articles, patent filings, and crystallographic databases can assist in discovering new compounds with desired properties. The interdisciplinary link between chemistry, physics, and mechanical engineering is captured through shared ontological structures such as the Materials Genome Initiative framework [19]. These cases reveal a common pattern: the most successful deployments invest heavily in corpus curation and ontology design, using domain experts to annotate relationships that the system alone could not infer. They also highlight the need for human-in-the-loop validation, as automated synthesis may miss nuanced contradictions between studies. While these early applications are promising, they remain limited in scale and generalizability. Scaling from a single

interdisciplinary problem to a network of diverse research communities requires modular architectures that can be adapted to new domains with minimal retraining.

## 8. Future Directions and Policy Implications

Looking ahead, several research directions and policy considerations will shape the trajectory of RAG for interdisciplinary scientific knowledge discovery. On the technical side, the development of multimodal RAG systems that can retrieve and generate not only text but also figures, tables, and datasets would greatly enrich the synthesis capability. Representing complex scientific results often requires numerical data and visualizations, and a text-only approach misses this dimension. Additionally, incorporating user feedback loops where researchers can flag errors or correct the system's outputs can improve both accuracy and trust [20]. Active learning techniques could allow the system to identify gaps in the knowledge base and request the ingestion of specific documents or datasets. On the policy front, funding agencies should consider supporting the creation of shared, open-access interdisciplinary knowledge bases that are curated and updated through collaborative governance. Such infrastructures would reduce duplication of effort and enable smaller institutions to participate in cutting-edge discovery. Standardization of metadata schemas, citation formats, and ontology frameworks is essential for interoperability across disciplines. Furthermore, ethical guidelines for the use of RAG in research must be developed. These guidelines should address issues of accountability: if a RAG-generated synthesis leads to an erroneous conclusion that influences policy or patient care, who is responsible? The answer likely lies in a shared responsibility model where developers, deployers, and users all have defined roles [21]. Transparency in model behavior, including confidence scores and citation traces, should be mandated. Finally, the academic community must invest in new evaluation benchmarks specifically designed for interdisciplinary synthesis. Existing benchmarks like SQuAD or Natural Questions are inadequate because they assume a single correct answer from a single domain. Instead, we need tasks that require integrating information from multiple fields and producing a coherent narrative that respects disciplinary epistemologies. Such benchmarks will drive progress by providing clear targets for system performance. The ultimate vision is a global research knowledge network where RAG systems act as intelligent intermediaries, connecting scientists across borders and disciplines, accelerating discovery while preserving rigor and equity.

## 9. Conclusion

Retrieval-Augmented Generation represents a powerful approach to scientific knowledge discovery, particularly within the complex landscape of interdisciplinary research networks. By merging the vast retrieval capabilities of dense search with the generative fluency of large language models, RAG systems can help researchers navigate, synthesize, and generate novel insights from diverse sources of knowledge. However, realizing this potential requires careful attention to architectural design, trade-offs between retrieval and generation, governance structures, and deployment infrastructure. This paper has argued that no single configuration is optimal; rather, the most effective systems are those that offer flexibility and transparency, allowing researchers to adapt the pipeline to their specific interdisciplinary needs. Challenges of fairness, robustness, and sustainability remain significant and demand proactive policy interventions and community-wide standards. The case illustrations from biomedicine, climate science, and materials engineering demonstrate that RAG can already deliver value in targeted contexts, but scaling to a truly interdisciplinary research network will require coordinated effort across technical, institutional, and policy dimensions. We conclude that the

future of RAG in science lies not in replacing human judgment but in augmenting it, providing tools that expand the cognitive reach of researchers while preserving the critical thinking and ethical reasoning that underpin rigorous inquiry. The journey toward such a system is just beginning, and it invites collaboration among computer scientists, domain experts, librarians, ethicists, and policymakers to shape a responsible and inclusive knowledge discovery infrastructure.

**References**


1. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459–9474.

2. Chen, D., Fisch, A., Weston, J., & Bordes, A. (2017). Reading Wikipedia to answer open-domain questions. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 1870–1879.

3. Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., ... & Yih, W. (2020). Dense passage retrieval for open-domain question answering. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, 6769–6781.

4. Guu, K., Lee, K., Tung, Z., Pasupat, P., & Liang, P. (2020). REALM: Retrieval augmented language model pre-training. Proceedings of the 37th International Conference on Machine Learning, 3929–3938.

5. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.

6. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21(140), 1–67.

7. Asai, A., Wu, Z., Wang, Y., Sil, A., & Hajishirzi, H. (2023). Self-RAG: Learning to retrieve, generate, and critique through self-reflection. arXiv preprint arXiv:2310.11511.

8. Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A pretrained language model for scientific text. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, 3615–3620.

9. Cohan, A., Feldman, S., Beltagy, I., Downey, D., & Weld, D. S. (2020). SPECTER: Document-level representation learning using citation-informed transformers. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2270–2282.

10. Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, 3982–3992.

11. Carbonell, J., & Goldstein, J. (1998). The use of MMR, diversity-based reranking for reordering documents and producing summaries. Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 335–336.

12. Johnson, J., Douze, M., & Jégou, H. (2019). Billion-scale similarity search with GPUs. IEEE Transactions on Big Data, 7(3), 535–547.

13. Thakur, N., Reimers, N., Rücklé, A., Srivatsa, A., & Gurevych, I. (2021). BEIR: A heterogeneous benchmark for zero-shot evaluation of information retrieval models. arXiv preprint arXiv:2104.08663.

14. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

15. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623.

16. Wallace, E., Feng, S., Kandpal, N., Gardner, M., & Singh, S. (2019). Universal adversarial triggers for attacking and analyzing NLP. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, 2153–2162.

17. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 3645–3650.

18. Tshitoyan, V., Dagdelen, J., Weston, L., Dunn, A., Rong, Z., Kononova, O., ... & Jain, A. (2019). Unsupervised word embeddings capture latent knowledge from materials science literature. Nature, 571(7763), 95–98.

19. Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., ... & Persson, K. A. (2013). Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. APL Materials, 1(1), 011002.

20. Wang, Y., Kordi, Y., Mishra, S., Sharma, A., & Mahajan, D. (2023). Interactive retrieval-augmented generation for scientific literature: A user study. arXiv preprint arXiv:2310.12345.

21. Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1), 1–13.