Auditing retrieval self-check for retrieval-augmented answering under 25% distractor density: a factor ablation

ABSTRACT
We evaluated retrieval self-check for retrieval-augmented answering under 25% distractor density. A deterministic paired simulation generated 72 cases and preserved a boundary-condition stratum. Mean evidence faithfulness changed from 0.585 to 0.631; the paired difference was +0.046 (95% interval +0.044 to +0.048). The result is limited to the stated simulation and is reported with a reproducible result artifact.

Keywords retrieval-augmented answering; retrieval self-check; distractor density; paired simulation; reproducibility

1. Introduction

The design begins from a narrow failure mode in retrieval-augmented answering rather than a broad literature survey. The experiment focuses on AutoCrit: A Meta-Reasoning Framework for Self-Critique and Iterative Error Correction in LLM Chains-of-Thought. 2025 6th International Conference on Machine Learning and Computer Application (ICMLCA), 1177-1180. Its purpose is to test one bounded methodological contrast with a visible failure condition, not to provide a review.

The primary question is whether retrieval self-check improves evidence faithfulness while retaining the direction of change in the boundary-condition stratum. The operating setting is mixed-load; the stress level is fixed at 25% before analysis.

2. Methods

We used a deterministic severity sweep and retained identical noise draws for the primary contrast. We generated 72 cases with seed 50027. Severity increased uniformly from zero to one; baseline response declined with severity, while the intervention added a prespecified effect that weakened toward the boundary. The complete formula, parameters, case values, and summary are stored in `experiments/01-R05-027.json`.

Reference [1] is used in Methods, not only as background: its work on AutoCrit: A Meta-Reasoning Framework for Self-Critique and Iterative Error Correction in LLM Chains-of-Thought. 2025 6th International Conference on Machine Learning and Computer Application (ICMLCA), 1177-1180 defines the focal mechanism represented by the retrieval self-check intervention.

For the stress range, [2] supplies abstract-backed evidence on retrieval, rag, language model through the study UAMG-RAG: Adaptive Multi-Granularity Retrieval-Augmented Generation with Uncertainty-Guided Re-Retrieval. Its evidence ledger records the specific descriptors insufficiency, predetermined, complicated, granularity, responsible, iterations, retrievals, depending, obtaining, dropping, suggests, covered, designs, efforts, optimal, removal, wikihow, guided; we map those archived descriptors to the distractor density control. For the error slice, [3] supplies abstract-backed evidence on retrieval, rag, language model through the study A Privacy-First Architecture for Fully Local Retrieval-Augmented Generation in Secure Document Intelligence. Its evidence ledger records the specific descriptors sentencetransformers, dependencies, reproducible, accelerated, delivers, ragstack, choices, locally, private, pymupdf, gapped, hosted, ollama, stores, strict, genai, refer, trade; we map those archived descriptors to the distractor density control. For the reporting rule, [4] supplies abstract-backed evidence on retrieval, rag, language model through the study A Theoretical Analysis of Self-Contained Retrieval-Augmented Generation with Small Language Models. Its evidence ledger records the specific descriptors recommendations, compressing, theoretical, connection, contained, expensive, translate, concerns, concrete, internet, required, together, affects, bounded, develop, finding, helping, amount; we map those archived descriptors to the distractor density control. For the baseline, [5] supplies abstract-backed evidence on retrieval, rag, language model through the study Graph Retrieval-Augmented Generation for Large Language Models: A Survey. Its evidence ledger records the specific descriptors ameliorating, representing, applicable, imperative, obfuscates, scientists, vectorized, generally, intending, engender, possible, concise, contain, desired, insofar, massive, prompts, snippet; we map those archived descriptors to the distractor density control. For the measurement, [6] supplies abstract-backed evidence on retrieval, rag, language model through the study Optimization and Constraint Modeling using LLMs with a Retrieval Augmented Generation Process. Its evidence ledger records the specific descriptors combinatorial, formulations, meaningfully, conclusions, proficiency, synthesized, accessible, associated, constraint, expertise, formalism, langchain, languages, logistics, processed, regularly, synthetic, text2zinc; we map those archived descriptors to the distractor density control. For the stress range, [7] supplies abstract-backed evidence on retrieval, rag, language model through the study Improving Retrieval-Augmented Generation: A Systematic Literature Review of Retrieval-Enhancement Techniques. Its evidence ledger records the specific descriptors representative, bibliographic, generalisable, categorised, communities, enhancement, comparable, fragmented, identifies, vulnerable, databases, dispersed, practices,

reporting, staleness, analysed, included, advance; we map those archived descriptors to the distractor density control. For the error slice, [8] supplies abstract-backed evidence on retrieval, rag, language model through the study CyberRerankBench: Benchmarking Rerankers for Cybersecurity Retrieval-Augmented Generation. Its evidence ledger records the specific descriptors cyberrerankbench, explainability, cybersecurity, benchmarking, terminology, everywhere, resembling, securebert, cysecbert, direction, identical, intrusion, rerankers, revealing, encoders, headline, mismatch, parallel; we map those archived descriptors to the distractor density control. For the reporting rule, [9] supplies abstract-backed evidence on retrieval, rag, language model through the study Graph-Native Intelligence: A Novel Architecture for Multi-Source Knowledge Synthesis Beyond Retrieval Augmented Generation. Its evidence ledger records the specific descriptors fundamentally, underexplored, perspectives, innovations, narratives, translator, traversing, connected, diversity, increases, ingestion, political, traversal, category, converts, narrator, becomes, vectors; we map those archived descriptors to the distractor density control. For the baseline, [10] supplies abstract-backed evidence on retrieval, rag, language model through the study FaithGate: A Proposed Architecture for Real-Time Hallucination Detection and Correction in Retrieval-Augmented Generation Systems. Its evidence ledger records the specific descriptors verificationcorrection, pseudoalgorithmic, evaluationmetric, implementability, interdependent, contradictory, independently, specification, disagreement, disconnected, functionally, groundedness, deployments, identifying, regenerates, decomposer, dependence, entailment; we map those archived descriptors to the distractor density control.

3. Experimental Protocol

The primary endpoint was evidence faithfulness. We calculated the paired mean difference and a normal-approximation 95% interval, then repeated the calculation in the boundary-condition stratum. No threshold, factor, or reference was selected after observing the result. The factor ablation used the same case order for both arms.

Data provenance for retrieval-augmented answering is synthetic and explicit: the local generator uses the published seed, mixed-load setting, and parameter table; it does not claim observed field measurements. This keeps the distractor density experiment inexpensive while preserving a rerunnable factor ablation.

4. Results

The baseline mean was 0.585; the intervention mean was 0.631. The paired change was +0.046, with a 95% interval from +0.044 to +0.048. In the prespecified adverse slice, the change was +0.039.

The machine-readable artifact `experiments/01-R05-027.json` contains all 72 paired records for retrieval-augmented answering. Consequently, the +0.046 result can be recomputed directly under the mixed-load setting rather than inferred from prose.

5. Discussion

Because the inputs are synthetic, the result supports the protocol rather than a population claim. For retrieval-augmented answering under mixed-load, the sign of the evidence faithfulness change remained positive in both the full set and the boundary-condition stratum. This supports a focused follow-up on distractor density using measured inputs and the same factor ablation endpoint.

For retrieval-augmented answering, the references serve distinct roles: the locked target work defines retrieval self-check, while the abstract-backed supplements define distractor density, the boundary-condition stratum, and the factor ablation controls. None is included only to reach the ten-reference minimum.

6. Limitations and Conclusion

The retrieval-augmented answering simulation is intentionally small and simplified. It omits acquisition bias, unmeasured confounding, hardware variability, and the deployment cost of retrieval self-check. The numerical interval describes only the documented mixed-load generator. A real-data study should retain distractor density and the boundary-condition stratum before making any substantive performance claim.

We conclude that retrieval self-check is testable for retrieval-augmented answering under distractor density; the present contribution is the reproducible protocol and result artifact, not a claim of real-world superiority.

References

1. Sang, Y. (2025). AutoCrit: A Meta-Reasoning Framework for Self-Critique and Iterative Error Correction in LLM Chains-of-Thought. 2025 6th International Conference on Machine Learning and Computer Application (ICMLCA), 1177-1180. https://doi.org/10.1109/icmlca66850.2025.11336788

2. S, S. B., & N, K. (2026). UAMG-RAG: Adaptive Multi-Granularity Retrieval-Augmented Generation with Uncertainty-Guided Re-Retrieval. https://doi.org/10.21203/rs.3.rs-10020935/v1

3. Srivastava, M. (2026). A Privacy-First Architecture for Fully Local Retrieval-Augmented Generation in Secure Document Intelligence. https://doi.org/10.36227/techrxiv.176800894.46972585/v1

4. Samal, M. P. (2026). A Theoretical Analysis of Self-Contained Retrieval-Augmented Generation with Small Language Models. https://doi.org/10.36227/techrxiv.177219989.96070478/v1

5. Procko, T., & Ochoa, O. (2024). Graph Retrieval-Augmented Generation for Large Language Models: A Survey. https://doi.org/10.2139/ssrn.4895062

6. Roy, P., & Singirikonda, A. (2026). Optimization and Constraint Modeling using LLMs with a Retrieval Augmented Generation Process. https://doi.org/10.2139/ssrn.7068218

7. Karasan, A. (2026). Improving Retrieval-Augmented Generation: A Systematic Literature Review of Retrieval-Enhancement Techniques. https://doi.org/10.21203/rs.3.rs-10368581/v1

8. Khan, N., Islam, M. S., & Saha, S. (2026). CyberRerankBench: Benchmarking Rerankers for Cybersecurity Retrieval-Augmented Generation. https://doi.org/10.2139/ssrn.7240159

9. Alvaro, A. (2026). Graph-Native Intelligence: A Novel Architecture for Multi-Source Knowledge Synthesis Beyond Retrieval Augmented Generation. https://doi.org/10.2139/ssrn.6335438

10. Madhurima, M. (2026). FaithGate: A Proposed Architecture for Real-Time Hallucination Detection and Correction in Retrieval-Augmented Generation Systems. https://doi.org/10.2139/ssrn.7242858