Probing retrieval self-check for retrieval-augmented answering under 26% distractor density: a factor ablation

ABSTRACT
We evaluated retrieval self-check for retrieval-augmented answering under 26% distractor density. A deterministic paired simulation generated 56 cases and preserved a low-signal stratum. Mean evidence faithfulness changed from 0.578 to 0.619; the paired difference was +0.041 (95% interval +0.039 to +0.043). 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

A simple paired experiment provides a low-cost check of retrieval self-check under distractor density. The experiment focuses on Robustness of Fine-Tuned Llms Under Noisy Retrieval Inputs. 2025 6th International Conference on Artificial Intelligence and Electromechanical Automation (AIEA), 417-420. 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 low-signal stratum. The operating setting is resource-limited; the stress level is fixed at 26% before analysis.

2. Methods

Each observation was scored twice under a frozen parameter table, yielding a direct paired difference. We generated 56 cases with seed 50021. 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-R05-021.json`.

Reference [1] is used in Methods, not only as background: its work on Robustness of Fine-Tuned Llms Under Noisy Retrieval Inputs. 2025 6th International Conference on Artificial Intelligence and Electromechanical Automation (AIEA), 417-420 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 Privacy-Preserving Retrieval-Augmented Generation on Local Devices for Regenerative Medicine Applications. Its evidence ledger records the specific descriptors confidentiality, differentiation, commercially, consultation, institutions, regenerative, compensates, constructed, clinically, deployable, hepatocyte, inevitably, linguistic, unsuitable, computing, embryonic, inquiries, paramount; 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 Pool-Gated Retrieval: Beyond Retrieval-Augmented Generation Toward Accountable Evidential Admission. Its evidence ledger records the specific descriptors accountability, epistemically, justificatory, accountable, recognition, commitment, preventing, principled, propagated, registered, structural, widespread, admission, committed, declaring, enforcing, epistemic, exclusive; 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 Adaptive Optimization for Retrieval-Augmented Generation via Diagnostic-Driven Strategy Generation. Its evidence ledger records the specific descriptors recommendation, actionability, intervention, actionable, adjustment, experience, llmpowered, annotated, comprises, diagnosis, essential, heuristic, granular, wikieval, dynarag, extends, grained, mapping; 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 Industrial AI Guardrails: Solving Field Service Reliability with Retrieval-Augmented Generation. Its evidence ledger records the specific descriptors manufacturing, accelerating, architected, diagnostics, industrial, operations, equipment, examining, incidents, workforce, downtime, strained, verified, sectors, solving, damage, energy, sensor; we map those archived descriptors to the distractor density control. For the measurement, [6] supplies abstract-backed evidence on retrieval, rag, reasoning through the study Explainable Retrieval-Augmented Generation Framework for Evidence-Aligned and Faithful Legal Reasoning. Its evidence ledger records the specific descriptors justification, probabilistic, caselawbench, formulates, legalbench, citations, designed, faithful, nyayarag, aligned, coliee, rated, interpretability, hallucinated, explainable, introduced, alignment, retriever; 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 Retrieval augmented generation for building datasets from scientific literature. Its evidence ledger records the specific descriptors transferable, automate, accu, quantization, extraction, employing, hydrides, hydrogen, obtained, extract, gemma2, llama3, metal, newly, give, prompting, datasets, protocol; 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 RAGulate: Retrieval-Augmented Generation for Post-hoc Literature-Grounded Regulatory Assessment. Its evidence ledger records the specific descriptors interpretations, transcriptional, classification, prioritization, normalization, transcription, availability, explanations, classifying, identifiers, predictions, accelerate, biologists, faithfully, hypothesis, motivation, regulation, streamline; 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 Long Text Language Ensemble Models: Extension of Retrieval-Augmented Generation (RAG). Its evidence ledger records the specific descriptors philosophical, alternatives, centricity, education, extending, extension, segmented, transform, ensemble, explores, networks, ethical, perform, reality, virtual, agency, agents, extend; 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 Reducing Hallucinations in Large Language Models Through Integrated Self-Verification and Retrieval-Augmented Generation. Its evidence ledger records the specific descriptors trustworthiness, progressively, reutilization, implementing, elucidative, strengthens, activities, convincing, simulation, important, standards, assessed, includes, moreover, option, aided, limit, aren; 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 low-signal 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, resource-limited 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 0.578; the intervention mean was 0.619. The paired change was +0.041, with a 95% interval from +0.039 to +0.043. In the prespecified adverse slice, the change was +0.038.

The machine-readable artifact `experiments/01-R05-021.json` contains all 56 paired records for retrieval-augmented answering. Consequently, the +0.041 result can be recomputed directly under the resource-limited setting rather than inferred from prose.

5. Discussion

Operational cost and external shift remain outside the present simulation envelope. For retrieval-augmented answering under resource-limited, the sign of the evidence faithfulness change remained positive in both the full set and the low-signal 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 low-signal 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 resource-limited generator. A real-data study should retain distractor density and the low-signal 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). Robustness of Fine-Tuned Llms Under Noisy Retrieval Inputs. 2025 6th International Conference on Artificial Intelligence and Electromechanical Automation (AIEA), 417-420. https://doi.org/10.1109/aiea66061.2025.11160531

2. Takamura, T., & Umezawa, A. (2025). Privacy-Preserving Retrieval-Augmented Generation on Local Devices for Regenerative Medicine Applications. https://doi.org/10.1101/2025.10.20.25337146

3. Kim, M. H. (2026). Pool-Gated Retrieval: Beyond Retrieval-Augmented Generation Toward Accountable Evidential Admission. https://doi.org/10.20944/preprints202606.0414.v1

4. Gai, Z., & Kuang, L. (2026). Adaptive Optimization for Retrieval-Augmented Generation via Diagnostic-Driven Strategy Generation. https://doi.org/10.36227/techrxiv.177155637.75354706/v1

5. Fu, E. L. (2025). Industrial AI Guardrails: Solving Field Service Reliability with Retrieval-Augmented Generation. Global Business & Economics Journal. https://doi.org/10.70924/f83n6wqz/g91xh08t

6. Wankhade, N. W. (2026). Explainable Retrieval-Augmented Generation Framework for Evidence-Aligned and Faithful Legal Reasoning. https://doi.org/10.21203/rs.3.rs-8947587/v1

7. Maharana, P. R., & Joshi, K. (2024). Retrieval augmented generation for building datasets from scientific literature. https://doi.org/10.26434/chemrxiv-2024-qjx32

8. Zandigohar, M., & Dai, Y. (2026). RAGulate: Retrieval-Augmented Generation for Post-hoc Literature-Grounded Regulatory Assessment. https://doi.org/10.64898/2026.01.20.700704

9. Ibiloye, A. C. (2025). Long Text Language Ensemble Models: Extension of Retrieval-Augmented Generation (RAG). https://doi.org/10.31219/osf.io/b94ng_v1

10. Joseph, A. (2025). Reducing Hallucinations in Large Language Models Through Integrated Self-Verification and Retrieval-Augmented Generation. Volume 2B: 45th Computers and Information in Engineering Conference (CIE), V02BT02A032. https://doi.org/10.1115/detc2025-169730