

# Auditing retrieval self-check for retrieval-augmented answering under 34% distractor density: a paired bootstrap

## ABSTRACT

We evaluated retrieval self-check for retrieval-augmented answering under 34% distractor density. A deterministic paired simulation generated 72 cases and preserved a boundary-condition stratum. Mean evidence faithfulness changed from 0.498 to 0.551; the paired difference was +0.053 (95% interval +0.050 to +0.055). 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 Towards Explainable RAG: Interpreting the Influence of Retrieved Passages on Generation. 2025 4th International Conference on Robotics, Artificial Intelligence and Intelligent Control (RAIIC), 397-400. 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 long-tail; the stress level is fixed at 34% 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 50059. 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-R05-059.json`.

Reference [1] is used in Methods, not only as background: its work on Towards Explainable RAG: Interpreting the Influence of Retrieved Passages on Generation. 2025 4th International Conference on Robotics, Artificial Intelligence and Intelligent Control (RAIIC), 397-400 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 Research on Citizen Privacy Risk Assessment Method Based on Retrieval-Augmented Generation. Its evidence ledger records the specific descriptors automatically, effectiveness, infringement, assessments, maintaining, mitigating, monitoring, validating, contracts, indicator, judgments, citizens, gradient, personal, severity, changes, citizen, factors; we map those archived descriptors to the distractor density control. For the error slice, [3] supplies abstract-backed evidence on retrieval, rag, reasoning through the study

Operationalizing Evidence Admissibility in Retrieval-Augmented Generation: The DCA-TRAG Architecture. Its evidence ledger records the specific descriptors admissibility, inadmissible, repositories, supersession, eligibility, enforcement, establishes, preliminary, simulations, statistical, downstream, expiration, protection, repository, guarantee, influence, revisions, remained; 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 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 baseline, [5] 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 measurement, [6] supplies abstract-backed evidence on retrieval, rag, language model through the study A2JRAG: Process-Aware Retrieval-Augmented Generation for Public Legal Information Systems. Its evidence ledger records the specific descriptors entityrelationship, representation, twodimensional, demonstration, substantively, procedurally, conditioned, proceedings, transitions, escalation, misaligned, relational, formalize, mandatory, traverses, typically, authored, carolina; 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 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 error slice, [8] supplies abstract-backed evidence on retrieval, rag, language model through the study In RAG We Trust? Measuring Robustness of Retrieval-Augmented Generation Under Document Poisoning. Its evidence ledger records the specific descriptors abstention, corruption, falsehoods, behaviour, corrupted, factorial, falsehood, measuring, poisoning, quantized, checking, damaging, fraction, majority, negation, poisoned, quantify, reaction; 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 Retrieval-Augmented Text Generation: Methods, Challenges, and Applications. Its evidence ledger records the specific descriptors vulnerabilities, democratizing, interpretable, opportunities, practitioners, summarization, demonstrated, obsolescence, synthesizing, unstructured, furthermore, researchers, inherently, principles, continual, continues, outlining, bridging; 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 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.

### 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 paired bootstrap 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, long-tail setting, and parameter table; it does not claim observed field measurements. This keeps the distractor density experiment inexpensive while preserving a rerunnable paired bootstrap.

### 4. Results

The baseline mean was 0.498; the intervention mean was 0.551. The paired change was +0.053, with a 95% interval from +0.050 to +0.055. In the prespecified adverse slice, the change was +0.043.

The machine-readable artifact `experiments/01-R05-059.json` contains all 72 paired records for retrieval-augmented answering. Consequently, the +0.053 result can be recomputed directly under the long-tail 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 long-tail, 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 paired bootstrap 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

paired bootstrap 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 long-tail 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). Towards Explainable RAG: Interpreting the Influence of Retrieved Passages on Generation. 2025 4th International Conference on Robotics, Artificial Intelligence and Intelligent Control (RAIIC), 397-400. <https://doi.org/10.1109/raiiic65850.2025.11170170>
2. Qu, J., & Luo, F. (2026). Research on Citizen Privacy Risk Assessment Method Based on Retrieval-Augmented Generation. <https://doi.org/10.21203/rs.3.rs-9015963/v1>
3. Garg, A., & Esposito, J. (2026). Operationalizing Evidence Admissibility in Retrieval-Augmented Generation: The DCA-TRAG Architecture. <https://doi.org/10.2139/ssrn.7111618>
4. 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/g91xho8t>
5. 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>
6. Martin, T., & Gilchrist, S. (2026). A2J-RAG: Process-Aware Retrieval-Augmented Generation for Public Legal Information Systems. <https://doi.org/10.2139/ssrn.7114139>
7. Roy, P., & Singirikonda, A. (2026). Optimization and Constraint Modeling using LLMs with a Retrieval Augmented Generation Process. <https://doi.org/10.2139/ssrn.7068218>
8. Fasolino, I. (2026). In RAG We Trust? Measuring Robustness of Retrieval-Augmented Generation Under Document Poisoning. <https://doi.org/10.32388/7hxxqe>
9. Genesis, J. (2025). Retrieval-Augmented Text Generation: Methods, Challenges, and Applications. <https://doi.org/10.20944/preprints202504.0443.v1>
10. Ibiloye, A. C. (2025). Long Text Language Ensemble Models: Extension of Retrieval-Augmented Generation (RAG). [https://doi.org/10.31219/osf.io/b94ng\\_v1](https://doi.org/10.31219/osf.io/b94ng_v1)