

VERITROOPER Scout — Executive Summary - SEC 10K Evaluation

Model under test: gpt-5.5 · **Scored items:** 1000 · **Generated:** 2026-08-04 15:54:06
Human sign-off: not yet recorded for these results.



Bottom line. With VERITROOPER, the system answered more accurately and correctly refused unsupported questions.

Scorecard

Metric	Without VERITROOPER (vanilla-RAG baseline)	With VERITROOPER
Final verified accuracy	82.60%	96.10%
Answers scored incorrect	174	39
Improvement		+13.50pp
Failures recovered		86%

What it means

- The VERITROOPER pipeline recovered **86%** of the 174 answers the vanilla-RAG baseline got wrong (82.6% → 96.1% final verified accuracy).
- Weakest area for the raw model: **Negative (hallucination traps)** at 60% baseline (n=200) — the pipeline lifts it to 96%.

Recommended action

- The tested pipeline configuration outperformed the baseline on this material and is the preferred candidate for broader validation and — subject to provider risk review and recorded human sign-off — a limited controlled pilot. It removed the confirmed error class above at a cost of 14 answers the baseline got right and the pipeline did not (see the paired-outcome table in the full report); final release remains subject to the recorded release disposition and responsible-person sign-off.
- Consider the deterministic guardrails in the full report to catch the same failure class in your own stack — no model retraining.
- Treat these figures as indicative on this material (scored n=1000); broaden the set for statistically tight claims before an unconditional production decision.

This one-pager summarizes the run; the full report, the per-question listings, and the compliance-evidence artifacts accompany it in the same package.

Generated by VERITROOPER as an executive summary of an accuracy & robustness audit. This artifact is testing evidence only. It does not by itself constitute, certify, or confer conformity with any framework, and does not replace the conformity assessment or the provider's other obligations. Have counsel review before external use.