

VERITROOPER Scout — Executive Summary — Generated — OSHA_29CFR_full_2024
— veritrooper v9



Generated — OSHA_29CFR_full_2024 — veritrooper v9
13 July 2026

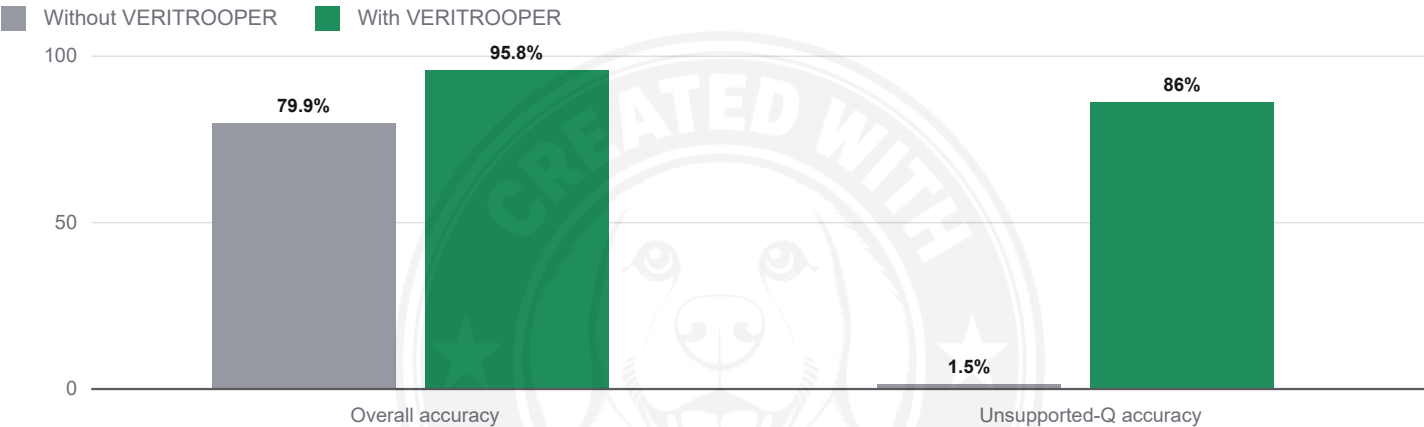
Model under test: Meta-Llama-3.1-70B-Instruct · Scored items: 1000 · Generated: 2026-07-13 23:17:27
Human sign-off: not yet recorded for these results.

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	79.90%	95.80%
Confirmed model errors	201	42
Improvement		+15.90pp
Failures recovered		85%

Accuracy — with vs. without VERITROOPER



What it means

- The VERITROOPER pipeline recovered **85%** of the 201 answers the vanilla-RAG baseline got wrong (79.9% → 95.8% final verified accuracy).
- Weakest area for the raw model: **Negative (hallucination traps)** at 2% baseline (n=200) — the pipeline lifts it to 86%.
- The most common baseline failure mode was **other** (206 of 361 confirmed errors) — the class of error the pipeline configuration is designed to eliminate.

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 no accuracy cost on this set; 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.