

AI Query Accuracy Certification

JetWare Workbench · PE Diligence Scenario · v1.59 Release Series

This certification documents the results of JetWare Workbench's automated accuracy audit suite, executed against verified PE diligence baselines using the Summit Care Holdings scenario. Every query result, generated narrative, and AI-synthesized finding is tested against known ground truth before each release.

SUMMARY RESULTS

<p>2,151 Automated tests passing</p>	<p>100% Structural accuracy tests</p>	<p>0 Hallucinations detected</p>	<p>0% Privacy mode delta</p>
---	--	---	---

ACCURACY TEST COVERAGE

Test Suite	What It Validates	Result
DuckDB execution accuracy	SQL results match verified PE demo baselines within 1% tolerance. Revenue \$52.3M, EBITDA \$6.06M, true Medicaid ~38%.	ALL PASS
Privacy mode prompt stripping	With Privacy Mode ON, AI prompt payloads contain zero raw data values. Schema and aggregated stats only — verified across all 6 endpoints.	ALL PASS
Workflow replay isolation	Results from a Summit Care run do not bleed into an Apex Behavioral Health replay. State isolation confirmed across both datasets.	ALL PASS
Narrative figure validation	AI-generated narratives cite figures matching query results within 1%. Dollar amounts and percentages confirmed against DuckDB output.	ALL PASS
Cross-validation accuracy	CIM-stated figures and data-derived figures captured without conflation. EBITDA gap (\$2.2M) and payer misclassification (~19% vs ~38%) both surface correctly.	ALL PASS
Live AI accuracy — Cloud mode	Natural language queries against the OpenAI endpoint return SQL producing correct results. 8 PE demo queries across 3 consistency runs each.	RUN BEFORE DEMO
Live AI accuracy — Local mode	Same 8 queries against Ollama (llama3.2) at localhost. Zero external API calls. Accuracy comparison vs Cloud baseline.	RUN BEFORE DEMO

AI MODES TESTED

● Cloud Mode

OpenAI GPT-4o-mini. Customer API key Fernet-encrypted at rest. Schema-only prompts by default. No raw data transmitted.

● Local Mode

Ollama at localhost:11434. Zero external calls. All inference runs on-machine. No API key required.

● Privacy Mode

Cloud or Local with raw data stripped from all prompts. AI sees schema and aggregated stats only — never patient or financial records.

KEY FINDINGS

- ✓ **Payer misclassification surfaces reliably.** The Summit Care Holdings payer classification discrepancy — \$10M in Medicaid MCO revenue classified as Commercial — is identified correctly in every test run. The query "Show total paid amount by program_type" surfaces Medicaid MCO as a distinct category with ~19.1% share, against a CIM-stated Medicaid figure of 18.9%.
- ✓ **EBITDA gap confirmed without analyst intervention.** Raw EBITDA (\$6.06M, 11.7% margin) vs CIM-stated adjusted EBITDA (\$8.3M, 15.9% margin) are captured as separate values in the session log. The \$2.2M gap and three add-back line items surface through natural language queries alone — no manual reconciliation required.
- ✓ **Privacy Mode does not degrade accuracy.** Structural tests confirm that Privacy Mode — which strips all raw data from AI prompt payloads — produces no accuracy delta on SQL generation or narrative output. The AI operates on schema and aggregated statistics only, with no reduction in result quality.
- ✓ **Clean targets read clean.** The Apex Behavioral Health replay produces no discrepancy findings. EBITDA reconciles (\$9.51M data vs \$9.5M CIM), payer classification is transparent (42.6% Medicaid including MCO confirmed in both sources). The system correctly distinguishes a clean deal from a discrepant one.

This certification is generated automatically from the JetWare Workbench test suite and updated with each release. The accuracy audit infrastructure is available for review by qualified technical evaluators. All test scenarios use fully synthetic data — no real patient, clinical, or proprietary deal information at any point.

Engineering Audit Team

JetWare AI LLC
jetwareai.com

Current release: v1.59 series - 2,151 automated tests passing