

RAGAS Evaluation Report

Evaluation Date: December 13, 2025 at 16:50:38
Total Test Cases: 10
Needle Questions: 5
Summary Questions: 5
Judge Model: models/gemini-flash-latest

Overall System Performance

System Score: 0.740

■ Good - System performing well with room for improvement

System-Wide Metrics

Metric	Score	Status
Context Precision	0.534	■ Needs Work
Context Recall	0.605	■ Needs Work
Faithfulness	0.842	✓ Excellent
Answer Relevancy	0.830	✓ Excellent
Answer Similarity	0.940	✓ Excellent
Answer Correctness	0.687	■ Needs Work

Agent-Specific Performance

Needle Agent Performance

Handles specific detail queries (needle-in-a-haystack)

Needle Agent Score: 0.779 (5 tests)

Metric	Score	vs System Avg
Context Precision	0.607	+0.073
Context Recall	0.800	+0.195
Faithfulness	0.700	-0.142
Answer Relevancy	0.782	-0.049
Answer Similarity	0.949	0.008
Answer Correctness	0.837	+0.150

Summary Agent Performance

Handles high-level overview queries

Summary Agent Score: 0.701 (5 tests)

Metric	Score	vs System Avg
Context Precision	0.461	-0.073
Context Recall	0.410	-0.195
Faithfulness	0.983	+0.142
Answer Relevancy	0.879	+0.049
Answer Similarity	0.932	-0.008
Answer Correctness	0.538	-0.150

Comparative Analysis

Agent	Avg Score	Best Metric	Weakest Metric
Needle Agent	0.779	Answer Similarity: 0.949	Context Precision: 0.607
Summary Agent	0.701	Faithfulness: 0.983	Context Recall: 0.410

Detailed Test Results

Test 1: needle_01

Question: Who conducted the phone interview with Sarah Mitchell?

Type: NEEDLE (routed to: NEEDLE)

Ground Truth: Claims Adjuster Linda Martinez conducted the phone interview with Sarah Mitchell at 16:30:00 hours.

Agent Answer: The information is not available in the provided documents.

Metric	Score
Context Precision	0.000
Context Recall	0.000
Faithfulness	0.000
Answer Relevancy	0.000
Answer Similarity	0.757
Answer Correctness	0.189

Retrieved: page_10_chunk_0, page_10_chunk_1, page_4_chunk_0, page_6_chunk_0, page_9_chunk_5, page_2_chunk_1
Expected: page_4_chunk_2

Test 2: needle_02

Question: What was Sarah Mitchell's blood pressure during the emergency assessment?

Type: NEEDLE (routed to: NEEDLE)

Ground Truth: Sarah Mitchell's blood pressure during the emergency assessment was recorded at 145/92 millimeters of mercury.

Agent Answer: Sarah Mitchell's blood pressure during the emergency assessment was recorded at 145/92 millimeters of mercury.

Metric	Score
Context Precision	0.500
Context Recall	1.000
Faithfulness	1.000
Answer Relevancy	1.000
Answer Similarity	1.000
Answer Correctness	1.000

Retrieved: page_3_chunk_2, page_3_chunk_1, page_4_chunk_0, page_6_chunk_0, page_4_chunk_3, page_3_chunk_3
Expected: page_3_chunk_1

Test 3: needle_03

Question: How many feet were the skid marks at the collision scene?

Type: NEEDLE (routed to: NEEDLE)

Ground Truth: The skid marks measured exactly 47 feet from Chen's vehicle trajectory on the wet pavement.

Agent Answer: The skid marks at the collision scene measured exactly 47 feet from Chen's vehicle trajectory on the wet pavement.

Metric	Score
Context Precision	1.000
Context Recall	1.000
Faithfulness	1.000
Answer Relevancy	0.915
Answer Similarity	0.989
Answer Correctness	0.997

Retrieved: page_3_chunk_4, page_4_chunk_4, page_4_chunk_3, page_2_chunk_4, page_1_chunk_4, page_2_chunk_5

Expected: page_3_chunk_2

Test 4: needle_04

Question: What was the speed of Robert Chen's vehicle when it entered the intersection?

Type: NEEDLE (routed to: NEEDLE)

Ground Truth: Robert Chen's vehicle was traveling at 45 miles per hour when it entered the intersection.

Agent Answer: Robert Chen's vehicle was traveling at an estimated speed of 45 miles per hour when it entered the intersection.

Metric	Score
Context Precision	0.700
Context Recall	1.000
Faithfulness	0.500
Answer Relevancy	0.994
Answer Similarity	0.996
Answer Correctness	0.999

Retrieved: page_2_chunk_4, page_7_chunk_7, page_7_chunk_2, page_3_chunk_5, page_2_chunk_3, page_4_chunk_5

Expected: page_2_chunk_2

Test 5: needle 05

Question: What is the claim number for this insurance claim?

Type: NEEDLE (routed to: NEEDLE)

Ground Truth: The claim number for this insurance claim is CLM-2024-00789-AUTO.

Agent Answer: The claim number for this insurance claim is CLM-2024-00789-AUTO.

Metric	Score
Context Precision	0.833
Context Recall	1.000
Faithfulness	1.000
Answer Relevancy	1.000
Answer Similarity	1.000
Answer Correctness	1.000

Retrieved: page_1_chunk_0, page_1_chunk_1, page_10_chunk_0, page_10_chunk_2, page_9_chunk_3, page_9_chunk_8

Expected: page_1_chunk_0



Test 6: summary_01

Question: What was the total estimated claim value?

Type: SUMMARY (routed to: SUMMARY)

Ground Truth: The total estimated claim value was \$47,850, consisting of \$44,010 for vehicle repairs and \$3,840 for medical expenses.

Agent Answer: The total estimated claim value for Sarah Mitchell's insurance claim, CLM-2024-00789-AUTO, is \$47,850. This amount encompasses \$44,010 for vehicle damage repairs and \$3,840 for medical expenses incurred due to the accident.

Metric	Score
Context Precision	1.000
Context Recall	1.000
Faithfulness	1.000
Answer Relevancy	0.876
Answer Similarity	0.901
Answer Correctness	0.654

Retrieved: [page 1 summary](#), [page 10 summary](#), [page 8 summary](#), [page 4 summary](#), [page 5 summary](#), [page 9 summary](#)

Expected: page 1 summary, page 10 summary, page 8 summary



Test 7: summary_02

Question: Summarize the events that led to the claim.

Type: SUMMARY (routed to: SUMMARY)

Ground Truth: On January 15, 2024, at 09:23:45 AM, Sarah Mitchell was involved in a multi-vehicle collision at the intersection of Pine Street and 4th Avenue in Seattle. Robert Chen's Toyota Camry ran a red light at 45 mph and struck Mitchell's Honda Accord on the driver's side door. Emergency responders arrived within 8 minutes, and Mitchell was transported to Seattle Medical Center with whiplash and cervical spine strain. The incident occurred during morning rush hour on wet roads from overnight rainfall.

Agent Answer: On January 15, 2024, Sarah Mitchell filed claim number CLM-2024-00789-AUTO following a multi-vehicle collision at the intersection of Maple Avenue and 5th Street in Downtown Seattle. The accident occurred at approximately 09:23 AM when Robert Chen failed to obey a traffic control device, entering the intersection against a red light. Witness Marcus Thompson confirmed Chen's violation, noting that his brake lights did not activate until after he entered the intersection. The Seattle Police Department subsequently issued a citation to Chen, who was determined to be 100% at fault for the incident, leading to the claim for vehicle repairs and medical expenses totaling \$47,850.

Metric	Score
Context Precision	0.000
Context Recall	0.000
Faithfulness	1.000
Answer Relevancy	0.803
Answer Similarity	0.926
Answer Correctness	0.360

Retrieved: page_1_summary, page_10_summary, page_8_summary, page_9_summary, page_4_summary, page_7_summary
Expected: page_1_summary, page_10_summary, page_2_summary, page_3_summary

Test 8: summary_03

Question: What was the outcome of Sarah Mitchell's medical treatment?

Type: SUMMARY (routed to: SUMMARY)

Ground Truth: Sarah Mitchell was diagnosed with Grade 2 whiplash injury, cervical spine strain, and mild concussion. She underwent emergency care, CT scan, and was prescribed physical therapy consisting of 12 sessions over 6 weeks. By her follow-up appointment, her pain had improved from 6 to 4 out of 10. She received full medical clearance from Dr. Patterson on February 3, 2024, with a good prognosis for full recovery. Total medical costs were \$3,840.

Agent Answer: Sarah Mitchell's medical treatment following the accident was comprehensive and included a follow-up appointment with Dr. Michael Patterson, who noted a decrease in her pain level from six to four. She was prescribed a 12-session physical therapy program at Seattle Rehabilitation Center, which was managed by Physical Therapist Amanda Chen. The total estimated medical treatment costs amounted to \$3,840, all of which were fully covered by her insurance policy. Overall, her treatment aimed to address cervical spine tenderness and improve her recovery from the injuries sustained in the collision.

Metric	Score
Context Precision	0.333
Context Recall	0.400
Faithfulness	0.917
Answer Relevancy	0.936

deductible, which was accounted for in the final payment.

Metric	Score
Context Precision	0.589
Context Recall	0.400
Faithfulness	1.000
Answer Relevancy	0.889
Answer Similarity	0.942
Answer Correctness	0.569

Retrieved: page_1_summary, page_10_summary, page_8_summary, page_9_summary, page_4_summary, page_5_summary

Expected: page_1_summary, page_10_summary, page_8_summary