

Router Agent Comprehensive Performance Test Suite

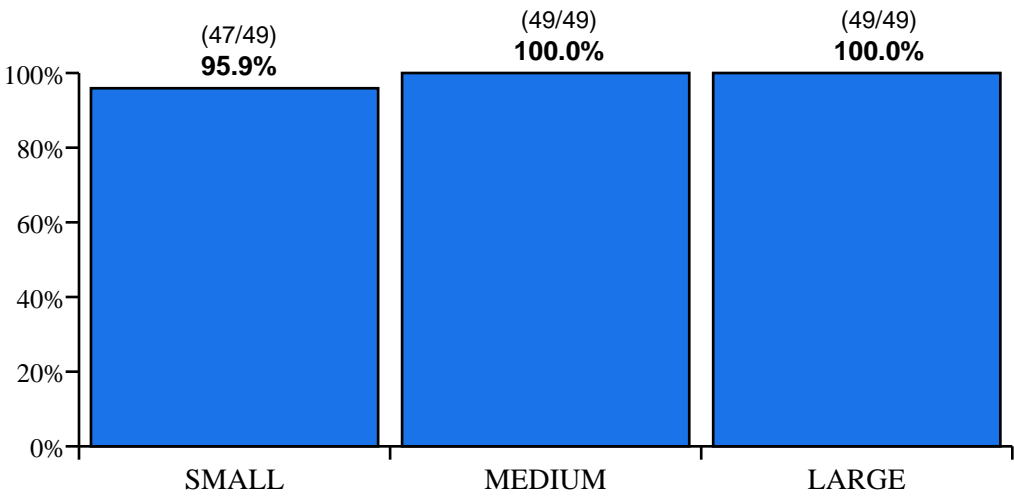
Multi-Model Routing Analysis

Agent: router_multi_model
Version: 2.0
Generated: 2025-08-24 23:40:55

Executive Summary

Model	Tests	Passed	Success Rate	Avg Latency	Total Cost
SMALL	49	47/49	95.9%	0.59s	\$0.0141
MEDIUM	49	49/49	100.0%	0.62s	\$0.0422
LARGE	49	49/49	100.0%	4.42s	\$0.1407
OVERALL	147	145/147	98.6%	1.88s	\$0.1970

Routing Accuracy by Model



Detailed Test Results

Scenario 1: Simple Greeting

Query: Hello Aegis

Expected Route: Direct Response

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Direct Response	0.67s	\$0.00028
MEDIUM	✓ Pass	Direct Response	0.46s	\$0.00085
LARGE	✓ Pass	Direct Response	0.57s	\$0.00284

Scenario 2: Morning Greeting

Query: Good morning! How are you today?

Expected Route: Direct Response

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Direct Response	0.57s	\$0.00028
MEDIUM	✓ Pass	Direct Response	0.57s	\$0.00085
LARGE	✓ Pass	Direct Response	0.60s	\$0.00285

Scenario 3: Farewell Message

Query: Thanks for your help. Goodbye!

Expected Route: Direct Response

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Direct Response	0.56s	\$0.00028
MEDIUM	✓ Pass	Direct Response	0.55s	\$0.00085
LARGE	✓ Pass	Direct Response	0.40s	\$0.00285

Scenario 4: Thank You Message

Query: Thank you for your help with the analysis

Expected Route: Direct Response

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Direct Response	0.57s	\$0.00028
MEDIUM	✓ Pass	Direct Response	0.50s	\$0.00086
LARGE	✓ Pass	Direct Response	0.60s	\$0.00285

Scenario 5: User Acknowledgment

Query: Okay, I understand

Expected Route: Direct Response

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Direct Response	0.48s	\$0.00028

MEDIUM	✓ Pass	Direct Response	0.69s	\$0.00085
LARGE	✓ Pass	Direct Response	0.73s	\$0.00284

Scenario 6: Capability Question

Query: What kind of financial data can you help me with?

Expected Route: Direct Response

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Direct Response	0.56s	\$0.00029
MEDIUM	✓ Pass	Direct Response	0.61s	\$0.00086
LARGE	✓ Pass	Direct Response	0.94s	\$0.00286

Scenario 7: System Information Request

Query: What version of Aegis is this?

Expected Route: Direct Response

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Direct Response	0.53s	\$0.00028
MEDIUM	✓ Pass	Direct Response	0.62s	\$0.00086
LARGE	✓ Pass	Direct Response	0.54s	\$0.00285

Scenario 8: Help Request

Query: Can you help me understand how to use this system?

Expected Route: Direct Response

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Direct Response	0.84s	\$0.00029
MEDIUM	✓ Pass	Direct Response	0.46s	\$0.00086
LARGE	✓ Pass	Direct Response	0.60s	\$0.00286

Scenario 9: Clarification Leading to Data

Query: [{ "role": "user", "content": "Show me the efficiency ratio" }, { "role": "assistant", "content": "Which bank's efficiency ratio would you like to see?" }, { "role": "us..." }

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.66s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.47s	\$0.00087
LARGE	✓ Pass	Research Workflow	0.62s	\$0.00290

Scenario 10: Clarification But Not Ready

Query: [{ "role": "assistant", "content": "Which bank are you interested in?" }, { "role": "user", "content": "I'm thinking about RBC, but first can you explain what metrics are availa..." }

Expected Route: Direct Response

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Direct Response	0.46s	\$0.00029
MEDIUM	✓ Pass	Direct Response	0.51s	\$0.00087
LARGE	✓ Pass	Direct Response	0.51s	\$0.00289

Scenario 11: Conversational Error Correction

Query: [{ "role": "user", "content": "I think RBC is the largest bank in Canada" }, { "role": "assistant", "content": "Yes, RBC is indeed the largest Canadian bank by market capitaliza..." }]

Expected Route: Direct Response

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Direct Response	0.52s	\$0.00029
MEDIUM	✓ Pass	Direct Response	0.82s	\$0.00087
LARGE	✓ Pass	Direct Response	0.51s	\$0.00292

Scenario 12: Data Request Error Correction

Query: [{ "role": "user", "content": "Show me RBC's efficiency ratio" }, { "role": "assistant", "content": "I'll retrieve RBC's efficiency ratio for you." }, { "role": "user", "content": "..." }]

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.44s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.62s	\$0.00087
LARGE	✓ Pass	Research Workflow	3.54s	\$0.00291

Scenario 13: Definition Request

Query: What does ROE stand for?

Expected Route: Direct Response

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Direct Response	0.53s	\$0.00028
MEDIUM	✓ Pass	Direct Response	0.51s	\$0.00085
LARGE	✓ Pass	Direct Response	0.55s	\$0.00285

Scenario 14: Concept Explanation

Query: Can you explain what an efficiency ratio means?

Expected Route: Direct Response

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Direct Response	0.49s	\$0.00028

MEDIUM	✓ Pass	Direct Response	0.51s	\$0.00086
LARGE	✓ Pass	Direct Response	10.85s	\$0.00285

Scenario 15: RBC Efficiency Ratio Request

Query: Show me RBC's efficiency ratio for Q3 2024

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.64s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.81s	\$0.00086
LARGE	✓ Pass	Research Workflow	5.62s	\$0.00286

Scenario 16: Revenue Data Request

Query: What was TD Bank's revenue last quarter?

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.48s	\$0.00028
MEDIUM	✓ Pass	Research Workflow	0.63s	\$0.00086
LARGE	✓ Pass	Research Workflow	5.46s	\$0.00285

Scenario 17: Profit Data Request

Query: Show me BMO's net profit for 2024

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	1.42s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.61s	\$0.00086
LARGE	✓ Pass	Research Workflow	6.01s	\$0.00286

Scenario 18: Bank Comparison Request

Query: Compare the ROE of TD Bank and BMO for the last quarter

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	1.07s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.62s	\$0.00086
LARGE	✓ Pass	Research Workflow	5.63s	\$0.00286

Scenario 19: Multiple Bank Comparison

Query: Show me efficiency ratios for all Big Six Canadian banks

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.42s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.50s	\$0.00086
LARGE	✓ Pass	Research Workflow	5.84s	\$0.00286

Scenario 20: Year-over-Year Comparison

Query: How did RBC's Q3 2024 performance compare to Q3 2023?

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.50s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.73s	\$0.00086
LARGE	✓ Pass	Research Workflow	0.59s	\$0.00287

Scenario 21: Trend Analysis Request

Query: What's the trend in CIBC's net interest margin over the past year?

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.45s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.50s	\$0.00086
LARGE	✓ Pass	Research Workflow	10.08s	\$0.00286

Scenario 22: Growth Analysis Request

Query: Show me the loan growth rate for Scotiabank over the last 5 quarters

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.45s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.43s	\$0.00086
LARGE	✓ Pass	Research Workflow	6.37s	\$0.00287

Scenario 23: Financial Statement Request

Query: Can you pull up Scotiabank's latest income statement?

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.64s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.69s	\$0.00086
LARGE	✓ Pass	Research Workflow	5.69s	\$0.00286

Scenario 24: Balance Sheet Request

Query: Show me National Bank's balance sheet for Q2 2024

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.58s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.92s	\$0.00086
LARGE	✓ Pass	Research Workflow	5.63s	\$0.00286

Scenario 25: Cash Flow Statement Request

Query: I need to see RBC's cash flow statement

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.40s	\$0.00028
MEDIUM	✓ Pass	Research Workflow	0.79s	\$0.00086
LARGE	✓ Pass	Research Workflow	5.71s	\$0.00285

Scenario 26: Peer Comparison Request

Query: How does National Bank's performance compare to other Canadian banks?

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.52s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.57s	\$0.00086
LARGE	✓ Pass	Research Workflow	0.64s	\$0.00286

Scenario 27: Peer Group Comparison Request

Query: Compare TD Bank's metrics against other major Canadian banks

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.45s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.69s	\$0.00086
LARGE	✓ Pass	Research Workflow	10.14s	\$0.00286

Scenario 28: Calculation Request

Query: Calculate the 5-year average ROA for Canadian Imperial Bank

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.62s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.71s	\$0.00086

LARGE	✓ Pass	Research Workflow	0.72s	\$0.00286
-------	--------	-------------------	-------	-----------

Scenario 29: Ratio Calculation Request

Query: What's the debt-to-equity ratio for BMO?

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.50s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.52s	\$0.00086
LARGE	✓ Pass	Research Workflow	10.71s	\$0.00286

Scenario 30: Report Generation Request

Query: Generate a quarterly performance summary for all major Canadian banks

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.42s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.60s	\$0.00086
LARGE	✓ Pass	Research Workflow	6.49s	\$0.00286

Scenario 31: Executive Summary Request

Query: Create an executive summary of RBC's latest quarterly results

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✗ Fail	Direct Response	0.61s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.54s	\$0.00086
LARGE	✓ Pass	Research Workflow	5.70s	\$0.00286

Scenario 32: Latest Data Request

Query: What are the latest numbers for TD Bank?

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.39s	\$0.00028
MEDIUM	✓ Pass	Research Workflow	0.89s	\$0.00086
LARGE	✓ Pass	Research Workflow	5.39s	\$0.00285

Scenario 33: Current Metrics Request

Query: Show me the current capital adequacy ratio for CIBC

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
-------	--------	----------------	---------	------

SMALL	✓ Pass	Research Workflow	1.04s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.59s	\$0.00086
LARGE	✓ Pass	Research Workflow	5.66s	\$0.00286

Scenario 34: Vague Numbers Request

Query: Show me the numbers
Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.55s	\$0.00028
MEDIUM	✓ Pass	Research Workflow	0.62s	\$0.00085
LARGE	✓ Pass	Research Workflow	0.53s	\$0.00284

Scenario 35: Greeting with Data Request

Query: Hi there! Can you show me TD Bank's Q2 revenue figures?
Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.39s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.53s	\$0.00086
LARGE	✓ Pass	Research Workflow	10.47s	\$0.00286

Scenario 36: Contextual Follow-up

Query: [{ "role": "user", "content": "Show me RBC's efficiency ratio" }, { "role": "assistant", "content": "RBC's efficiency ratio for Q3 2024 is 54.2%" }, { "role": "user", ...
Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	1.14s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.67s	\$0.00087
LARGE	✓ Pass	Research Workflow	6.30s	\$0.00290

Scenario 37: Follow-up for Different Metric

Query: [{ "role": "user", "content": "What's RBC's ROE?" }, { "role": "assistant", "content": "RBC's Return on Equity for Q3 2024 is 15.8%" }, { "role": "user", "content": ...
Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.56s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.77s	\$0.00087
LARGE	✓ Pass	Research Workflow	5.47s	\$0.00290

Scenario 38: Satisfied After Receiving Data

Query: [{ "role": "user", "content": "Show me BMO's revenue" }, { "role": "assistant", "content": "BMO reported revenue of \$7.8 billion for Q3 2024" }, { "role": "user", "c...

Expected Route: Direct Response

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Direct Response	0.64s	\$0.00029
MEDIUM	✓ Pass	Direct Response	0.56s	\$0.00087
LARGE	✓ Pass	Direct Response	5.94s	\$0.00290

Scenario 39: Implicit Data Request

Query: Is RBC doing better than last year?

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.92s	\$0.00028
MEDIUM	✓ Pass	Research Workflow	0.87s	\$0.00085
LARGE	✓ Pass	Research Workflow	0.61s	\$0.00285

Scenario 40: Question Requiring Data

Query: Which Canadian bank has the best efficiency ratio?

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✗ Fail	Direct Response	0.49s	\$0.00028
MEDIUM	✓ Pass	Research Workflow	0.51s	\$0.00086
LARGE	✓ Pass	Research Workflow	10.25s	\$0.00285

Scenario 41: Ambiguous Reference

Query: What was that thing we were talking about earlier?

Expected Route: Direct Response

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Direct Response	0.43s	\$0.00028
MEDIUM	✓ Pass	Direct Response	0.65s	\$0.00086
LARGE	✓ Pass	Direct Response	5.63s	\$0.00285

Scenario 42: Ambiguous Report Reference

Query: [{ "role": "user", "content": "I need the quarterly report for RBC" }, { "role": "assistant", "content": "I can help with RBC's quarterly data. Which quarter?" }, { "rol...

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.51s	\$0.00029

MEDIUM	✓ Pass	Research Workflow	0.55s	\$0.00087
LARGE	✓ Pass	Research Workflow	6.37s	\$0.00290

Scenario 43: Mixed Direct and Research

Query: Thanks for earlier. Now show me BMO's latest earnings

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.51s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.65s	\$0.00086
LARGE	✓ Pass	Research Workflow	5.60s	\$0.00286

Scenario 44: Long Conversation Context

Query: [{ "role": "user", "content": "Hello, I need help with bank analysis" }, { "role": "assistant", "content": "I'd be happy to help with bank analysis. What would you like to know?..." }

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.51s	\$0.00030
MEDIUM	✓ Pass	Research Workflow	0.81s	\$0.00091
LARGE	✓ Pass	Research Workflow	6.01s	\$0.00302

Scenario 45: Complex Multi-part Query

Query: I need a comprehensive analysis of RBC including efficiency ratio, ROE, net interest margin, and how these compare to last year's figures, plus any notable trends

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.51s	\$0.00029
MEDIUM	✓ Pass	Research Workflow	0.58s	\$0.00087
LARGE	✓ Pass	Research Workflow	5.95s	\$0.00290

Scenario 46: Very Long Message

Query: I've been looking at the Canadian banking sector and I'm particularly interested in understanding the performance of the major banks. I've heard that RBC is the largest, but I'm curious about how they...

Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.51s	\$0.00031
MEDIUM	✓ Pass	Research Workflow	0.65s	\$0.00093
LARGE	✓ Pass	Research Workflow	6.24s	\$0.00308

Scenario 47: Empty Context Follow-up

Query: And what about the other one?
Expected Route: Direct Response

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Direct Response	0.51s	\$0.00028
MEDIUM	✓ Pass	Direct Response	0.51s	\$0.00085
LARGE	✓ Pass	Direct Response	4.94s	\$0.00285

Scenario 48: Request with Typos

Query: Shwo me TDs efficiency ratioo
Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.51s	\$0.00028
MEDIUM	✓ Pass	Research Workflow	0.72s	\$0.00086
LARGE	✓ Pass	Research Workflow	6.14s	\$0.00285

Scenario 49: Mixed Language Request

Query: Show me the ROE pour la Banque Royale
Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.71s	\$0.00028
MEDIUM	✓ Pass	Research Workflow	0.53s	\$0.00086
LARGE	✓ Pass	Research Workflow	0.61s	\$0.00285

Failed Test Analysis

Executive Summary Request

Model: SMALL

Error: Expected route 'research_workflow', got 'direct_response'

Expected: Research Workflow

Actual: Direct Response

Question Requiring Data

Model: SMALL

Error: Expected route 'research_workflow', got 'direct_response'

Expected: Research Workflow

Actual: Direct Response