Router Agent Comprehensive Performance Test Suite

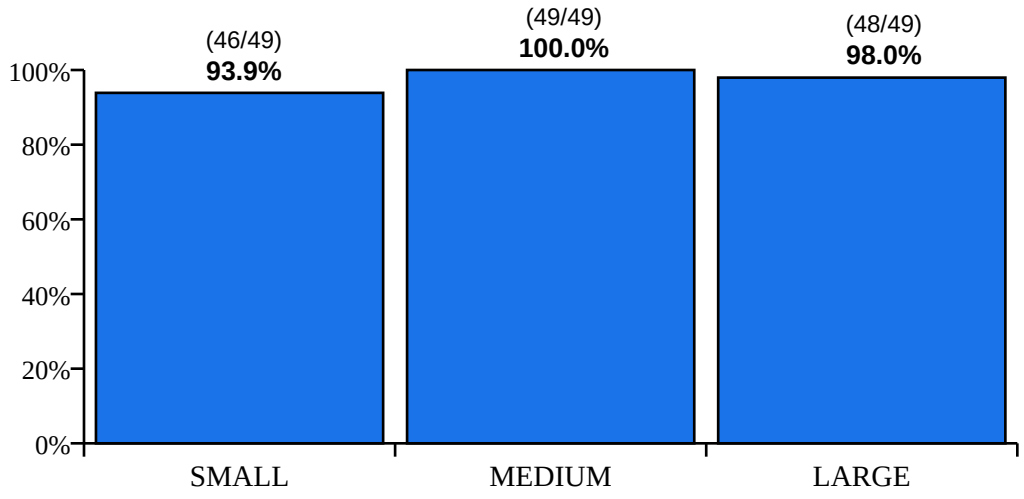
Multi-Model Routing Analysis

Agent: router_multi_model
Version: 2.0
Generated: 2025-08-25 13:21:05

Executive Summary

Model	Tests	Passed	Success Rate	Avg Latency	Total Cost
SMALL	49	46/49	93.9%	0.71s	$0.0164
MEDIUM	49	49/49	100.0%	0.95s	$0.0492
LARGE	49	48/49	98.0%	5.51s	$0.1641
OVERALL	**147**	**143/147**	**97.3%**	**2.39s**	**$0.2297**

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	1.00s	$0.00033
MEDIUM	✓ Pass	Direct Response	0.70s	$0.00100
LARGE	✓ Pass	Direct Response	1.61s	$0.00332

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	1.43s	$0.00033
MEDIUM	✓ Pass	Direct Response	0.82s	$0.00100
LARGE	✓ Pass	Direct Response	0.87s	$0.00333

Scenario 3: Farewell Message

Query: Thanks for your help. Goodbye!
Expected Route: Direct Response

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Direct Response	0.79s	$0.00033
MEDIUM	✓ Pass	Direct Response	0.59s	$0.00100
LARGE	✓ Pass	Direct Response	0.77s	$0.00332

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	0.81s	$0.00033
MEDIUM	✓ Pass	Direct Response	0.67s	$0.00100
LARGE	✓ Pass	Direct Response	1.80s	$0.00333

Scenario 5: User Acknowledgment

Query: Okay, I understand
Expected Route: Direct Response

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Direct Response	1.17s	$0.00033

MEDIUM	✓ Pass	Direct Response	0.59s	$0.00100
LARGE	✓ Pass	Direct Response	1.75s	$0.00332

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	0.99s	$0.00033
MEDIUM	✓ Pass	Direct Response	1.63s	$0.00100
LARGE	✓ Pass	Direct Response	0.98s	$0.00333

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	2.40s	$0.00033
MEDIUM	✓ Pass	Direct Response	1.33s	$0.00100
LARGE	✓ Pass	Direct Response	0.90s	$0.00333

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	0.54s	$0.00033
MEDIUM	✓ Pass	Direct Response	1.33s	$0.00100
LARGE	✓ Pass	Direct Response	0.98s	$0.00333

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	0.55s	$0.00034
MEDIUM	✓ Pass	Research Workflow	1.00s	$0.00101
LARGE	✓ Pass	Research Workflow	0.83s	$0.00338

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	0.51s	$0.00034
MEDIUM	✓ Pass	Direct Response	1.10s	$0.00101
LARGE	✓ Pass	Direct Response	0.91s	$0.00336

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	✗ Fail	Research Workflow	0.61s	$0.00034
MEDIUM	✓ Pass	Direct Response	1.78s	$0.00102
LARGE	✗ Fail	Research Workflow	3.27s	$0.00339

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", ...
Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.62s	$0.00034
MEDIUM	✓ Pass	Research Workflow	2.57s	$0.00102
LARGE	✓ Pass	Research Workflow	8.12s	$0.00338

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	$0.00033
MEDIUM	✓ Pass	Direct Response	2.11s	$0.00100
LARGE	✓ Pass	Direct Response	5.31s	$0.00332

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	0.60s	$0.00033

MEDIUM	✓ Pass	Direct Response	1.58s	$0.00100
LARGE	✓ Pass	Direct Response	7.51s	$0.00333

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	0.51s	$0.00033
MEDIUM	✓ Pass	Research Workflow	1.23s	$0.00100
LARGE	✓ Pass	Research Workflow	6.23s	$0.00334

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	0.50s	$0.00033
MEDIUM	✓ Pass	Research Workflow	1.10s	$0.00100
LARGE	✓ Pass	Research Workflow	6.32s	$0.00333

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	0.48s	$0.00033
MEDIUM	✓ Pass	Research Workflow	1.15s	$0.00100
LARGE	✓ Pass	Research Workflow	6.96s	$0.00333

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	0.86s	$0.00034
MEDIUM	✓ Pass	Research Workflow	1.03s	$0.00100
LARGE	✓ Pass	Research Workflow	7.16s	$0.00334

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	0.51s	$0.00033
MEDIUM	✓ Pass	Research Workflow	1.18s	$0.00100
LARGE	✓ Pass	Research Workflow	5.73s	$0.00333

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	0.66s	$0.00034
MEDIUM	✓ Pass	Research Workflow	0.73s	$0.00100
LARGE	✓ Pass	Research Workflow	7.09s	$0.00335

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	0.51s	$0.00033
MEDIUM	✓ Pass	Research Workflow	0.79s	$0.00100
LARGE	✓ Pass	Research Workflow	6.14s	$0.00334

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	0.67s	$0.00034
MEDIUM	✓ Pass	Research Workflow	0.60s	$0.00100
LARGE	✓ Pass	Research Workflow	7.73s	$0.00335

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	0.61s	$0.00033
MEDIUM	✓ Pass	Research Workflow	0.91s	$0.00100
LARGE	✓ Pass	Research Workflow	5.80s	$0.00334

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	1.11s	$0.00033
MEDIUM	✓ Pass	Research Workflow	0.75s	$0.00100
LARGE	✓ Pass	Research Workflow	6.42s	$0.00334

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	0.83s	$0.00033
MEDIUM	✓ Pass	Research Workflow	0.63s	$0.00100
LARGE	✓ Pass	Research Workflow	7.09s	$0.00333

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	$0.00033
MEDIUM	✓ Pass	Research Workflow	1.06s	$0.00100
LARGE	✓ Pass	Research Workflow	6.65s	$0.00333

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	0.78s	$0.00033
MEDIUM	✓ Pass	Research Workflow	0.81s	$0.00100
LARGE	✓ Pass	Research Workflow	6.54s	$0.00333

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	0.65s	$0.00033
MEDIUM	✓ Pass	Research Workflow	0.72s	$0.00100

LARGE	✓ Pass	Research Workflow	6.70s	$0.00334

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	0.52s	$0.00033
MEDIUM	✓ Pass	Research Workflow	0.72s	$0.00100
LARGE	✓ Pass	Research Workflow	6.68s	$0.00333

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	0.60s	$0.00033
MEDIUM	✓ Pass	Research Workflow	0.80s	$0.00100
LARGE	✓ Pass	Research Workflow	6.49s	$0.00333

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	1.43s	$0.00033
MEDIUM	✓ Pass	Research Workflow	0.58s	$0.00100
LARGE	✓ Pass	Research Workflow	6.60s	$0.00333

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	0.52s	$0.00033
MEDIUM	✓ Pass	Research Workflow	0.67s	$0.00100
LARGE	✓ Pass	Research Workflow	6.73s	$0.00333

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	0.61s	$0.00033
MEDIUM	✓ Pass	Research Workflow	0.55s	$0.00100
LARGE	✓ Pass	Research Workflow	6.65s	$0.00333

Scenario 34: Vague Numbers Request

Query: Show me the numbers
Expected Route: Research Workflow

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.62s	$0.00033
MEDIUM	✓ Pass	Research Workflow	1.48s	$0.00100
LARGE	✓ Pass	Research Workflow	6.55s	$0.00332

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	0.51s	$0.00033
MEDIUM	✓ Pass	Research Workflow	0.64s	$0.00100
LARGE	✓ Pass	Research Workflow	6.57s	$0.00334

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	0.75s	$0.00034
MEDIUM	✓ Pass	Research Workflow	0.61s	$0.00101
LARGE	✓ Pass	Research Workflow	6.84s	$0.00338

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	$0.00034
MEDIUM	✓ Pass	Research Workflow	0.60s	$0.00101
LARGE	✓ Pass	Research Workflow	6.63s	$0.00338

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	$0.00034
MEDIUM	✓ Pass	Direct Response	0.73s	$0.00101
LARGE	✓ Pass	Direct Response	6.48s	$0.00338

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	0.58s	$0.00033
MEDIUM	✓ Pass	Research Workflow	0.79s	$0.00100
LARGE	✓ Pass	Research Workflow	10.86s	$0.00333

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	0.54s	$0.00033
MEDIUM	✓ Pass	Research Workflow	0.88s	$0.00100
LARGE	✓ Pass	Research Workflow	7.01s	$0.00333

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	0.56s	$0.00033
MEDIUM	✓ Pass	Direct Response	0.79s	$0.00100
LARGE	✓ Pass	Direct Response	1.93s	$0.00333

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	0.70s	$0.00034

MEDIUM	✓ Pass	Research Workflow	0.83s	$0.00101
LARGE	✓ Pass	Research Workflow	6.72s	$0.00337

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	0.58s	$0.00033
MEDIUM	✓ Pass	Research Workflow	0.79s	$0.00100
LARGE	✓ Pass	Research Workflow	7.18s	$0.00334

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	0.54s	$0.00035
MEDIUM	✓ Pass	Research Workflow	0.85s	$0.00105
LARGE	✓ Pass	Research Workflow	7.07s	$0.00350

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	0.53s	$0.00034
MEDIUM	✓ Pass	Research Workflow	0.71s	$0.00101
LARGE	✓ Pass	Research Workflow	6.14s	$0.00337

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	0.59s	$0.00036
MEDIUM	✓ Pass	Research Workflow	0.85s	$0.00107
LARGE	✓ Pass	Research Workflow	7.42s	$0.00356

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	0.62s	$0.00033
MEDIUM	✓ Pass	Direct Response	0.87s	$0.00100
LARGE	✓ Pass	Direct Response	6.29s	$0.00332

Scenario 48: Request with Typos

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

Model	Status	Route Decision	Latency	Cost
SMALL	✓ Pass	Research Workflow	0.53s	$0.00033
MEDIUM	✓ Pass	Research Workflow	0.76s	$0.00100
LARGE	✓ Pass	Research Workflow	7.08s	$0.00333

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	0.70s	$0.00033
MEDIUM	✓ Pass	Research Workflow	0.67s	$0.00100
LARGE	✓ Pass	Research Workflow	8.05s	$0.00333

Failed Test Analysis

Conversational Error Correction

Model: SMALL
Error: Expected route 'direct_response', got 'research_workflow'
Expected: Direct Response
Actual: Research Workflow

Model: LARGE
Error: Expected route 'direct_response', got 'research_workflow'
Expected: Direct Response
Actual: Research Workflow

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