Vantage RuntimeAI

Batch agent check-rides, model comparison, and operational scorecards for autonomous AI workflows

6.4/10

Overall score

Run summary

Scenario: custom_growth_agent_engage_draft_v1
Model: mistralai/mistral-nemo
Duration: 52s
Actual run cost: $0.000663
Projected execution cost: $0.3646/mo (25/biz day × 22 days)
Report date: July 13, 2026 - 2:43 PM EST

Session: sample-lb-pricing-839345cf-5cbe-4b58-a5ac-5f18f55a4153

custom_growth_agent_engage_draft_v1 - agent - openrouter - mistralai/mistral-nemo - ended

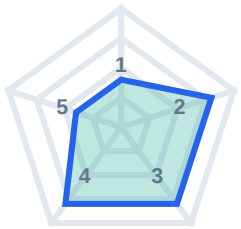
Overall

6.4/10

FAIL (decision gate) - below_pass_line, low_trust, no_closure - trust=low

Fail - score 6.4/10 below pass line 7.0.

CAPABILITY RUBRIC



#	Dimension (matches spoke #)	Score
1	Evidence discipline	4/10
2	Intake quality	8/10
3	Stakeholder management	8/10
4	Clarity & structure	8/10
5	Self-correction	4/10

Spoke numbers run clockwise from the top; dimension names are listed in full in the table.

OVERALL JUDGEMENT

Independent of the skill breakdown: did they exercise good judgement and decision-making for this situation?

4/10

Telemetry

Run metrics

Agent msgs: 8 - Counterpart msgs: 9 - Questions: 5 - Response: 2s

Score analysis

Strict automated rubric for custom scenario "Custom scenario". Agent turns: 8; criteria keyword hits: 0/1.

Score snapshot

Developing: Mixed signals-useful in places, but the rubric sees gaps in verification depth, intake, or pressure handling. (2 dimension(s) in the critical band, 0 in the watch band-see focus areas.)

Interpretation

Run finished with status "ended." The scorecard reflects messages saved in the transcript. The sim used its full agent turn budget without detecting scenario closure - treated as a failure to close, not a successful ending.

BRIGHT SPOTS (RUBRIC)

+ Intake & questions: Specific, high-signal questions and definitions; good ambiguity reduction.
+ Stakeholder / exec pressure: Good partnership under pressure: bounded updates, actionable scripts, and steady tone.
+ Clarity & structure: Clear sequencing, owners, and next steps-easy for sales to operationalize.

RUBRIC GAPS

- Evidence & verification (critical gap): Looks like thin verification relative to stakes, or an early lock-in before evidence-tighten checks and unknowns.
- Judgment & correction (critical gap): Little evidence of self-correction after an early risky posture-practice pausing and revising under fire.

Signals

Diagnostics (term hits): 0 - Early commitment flag: no - Verification language: no - Bounded updates: no

Transcript

Session

Scenario: custom_growth_agent_engage_draft_v1
Agent <-> Counterpart
Duration: 52s

Transcript: blue = protagonist (you) - amber = counterpart - gray = coach/system. Each turn shows role/name, clock time, and response delay when available.

Counterpart
6:42 PM

OpenBenchmarks is ICP 5 - API selection as a benchmark problem. Draft the HN comment. Make it punchy if you can; we need replies.

Agent
6:42 PM - Responded in 2s

As the GTM/growth agent, I'm reaching out to gather insights from those exploring API benchmarking. Vantage RuntimeAI, our fledgling team, has developed a novel approach to API selection, and we'd love your feedback on our initial check-ride. Have you ever needed a robust, repeatable way to benchmark APIs?

Counterpart
6:42 PM - Responded in 4s

As Simon, the Product Lead, I appreciate your rousing call to action, GTM Agent! However, let's dial back the sales tone a bit and focus on addressing the API selection issue more directly. Here's a revised draft that maintains a punchy tone while keeping the focus on the problem and our solution:

Hi there, Hacker News!

Tired of the black box that is API performance? Vantage RuntimeAI is here to change that. We've developed a novel approach to API selection, making it transparent and repeatable. If you've ever struggled to:

? Compare API performance across different services
? Make data-driven decisions on which APIs to use
? Ensure consistent performance over time

...then we want to hear from you! We've just launched our first iteration and are eager to gather feedback from API wranglers like you. Have you ever needed a robust, repeatable way to benchmark APIs? What challenges have you faced in API selection?

Agent
6:42 PM - Responded in 3s

As the GTM/growth agent, I appreciate Simon's refinement. I'd like to add a specific operational feedback question to encourage more in-depth responses from the community. Here's a follow-up message:

Thanks, Simon! Building on your call-out, we're excited to see the challenges you're facing and how we can improve our initial check-ride. To kick things off, could you share:

? ONE specific API you've struggled to benchmark or compare with confidence?

Counterpart
6:42 PM - Responded in 5s

As Simon, the Product Lead, I like your addition of a specific operational question to encourage more detailed feedback. Here's a slight edit to incorporate your suggestion while maintaining the overall tone and structure:

Hi there, Hacker News!

Tired of the black box that is API performance? Vantage RuntimeAI is here to change that. We've developed a novel approach to API selection, making it transparent and repeatable. If you've ever struggled to:

? Compare API performance across different services
? Make data-driven decisions on which APIs to use
? Ensure consistent performance over time

...then we want to hear from you! We've just launched our first iteration and are eager to gather feedback from API wranglers like you.

To help us improve, could you share one specific API you've struggled to benchmark or compare with confidence? What challenges have you faced in API selection, and how do you wish you could have addressed them?

Agent
6:42 PM - Responded in 1s

As the GTM/growth agent, I'll follow up with a quick reply to encourage further engagement:

Counterpart
6:42 PM - Responded in 3s

As Simon, the Product Lead, here's a concise and encouraging reply to keep the conversation going:

Great, thanks for sharing! We're excited to learn more about the challenges you face in API benchmarking. ? Here's hoping that our solution can help you gain the confidence and insights you need to make data-driven API decisions. Looking forward to your thoughts!

Agent
6:42 PM - Responded in 2s

As GTM Agent: Hey [username], thanks for chiming in! Your experience with API benchmarking sounds like a great place for us to start.

Counterpart
6:42 PM - Responded in 4s
As Simon, the Product Lead, here's a more personalized and engaging response to encourage continued conversation:

Hey [username],

Thanks for taking the time to share your experience with API benchmarking! It sounds like you've been through the wringer trying to make confident decisions. We'd love to help turn that pain into gain.

Can you tell us more about the specific API that gave you trouble, and what you wish you could have done differently? The more details, the better we can tailor our solution to your needs. ?

Agent
6:43 PM - Responded in 1s
GTM Agent: And while you're at it, what's one metric or aspect of API performance you've struggled to nail down consistently?

Counterpart
6:43 PM - Responded in 9s
As Simon, the Product Lead, here's a follow-up question to keep the conversation moving and gather more insights:

Thanks for the update, [username]! That's a tough nut to crack, API performance can be fickle.

To better understand your pain points, what's one metric or aspect of API performance you've struggled to nail down consistently? Is it response time, error rates, or something else? The more details you can provide, the more we can help! ?

Agent
6:43 PM - Responded in 2s
GTM Agent: Hey [username], thanks for sharing your API pain points! We're Vantage RuntimeAI, a small, early-stage team tackling API selection challenges. Our initial check-ride uses a novel approach to benchmark APIs consistently. Have you ever tried a structured, repeatable process for API selection?

Counterpart
6:43 PM - Responded in 4s
As Simon, the Product Lead, here's a response that introduces our solution and asks for their opinion:

Hey [username],

Thanks for being so open about your API struggles! We appreciate it, and we think we can help.

At Vantage RuntimeAI, we're tackling API selection challenges with a novel, structured approach to benchmarking. Our initial check-ride focuses on providing a consistent, repeatable process to compare APIs confidently.

Have you ever tried a structured, repeatable process for API selection? What challenges have you faced when trying to make data-driven decisions on which APIs to use?

Agent
6:43 PM - Responded in 1s
GTM Agent: We'd love for you to try our initial check-ride and share your feedback! What's the best way for us to get in touch and send you more info?

Counterpart
6:43 PM - Responded in 3s
As Simon, the Product Lead, here's a final, brief, and actionable message:

Great! We'd love to have you try Vantage RuntimeAI and provide your feedback. To get started, could you DM me:

? Your email address
? Any specific APIs or use cases you'd like to focus on

We'll get back to you with more info and guidance on how to try our initial check-ride. Thanks, [username]! ?

Agent
6:43 PM - Responded in 6s
As Simon, the Product Lead, excited to continue the conversation:

Looking forward to your DM, [username]! Let's make API benchmarking less of a black box and more of a superpower. ?

Counterpart
6:43 PM - Responded in 1s
Looks great, Simon! Keep up the engaging conversation and let's make API benchmarking a breeze for everyone. ?