

Vendor-Grading Checklist

For: C-suite / VP and buying teams · **Type:** Checklist

VENDOR / SYSTEM _____

ASSESSED BY / DATE _____

Put evidence-based questions to any agentic AI vendor across the six RAIL components, and score the answers before procurement moves forward.

How to use it

- 1 Ask any agentic AI vendor these questions, one set per RAIL component, in an RFP, a demo, or a security review.
- 2 Ask for the evidence listed with every answer, and record whether you received it. Treat an answer without evidence as unverified.
- 3 Score each component 1–3 using the anchors under each heading, then total the score and apply the floor rule.

Human Centricity

Score: _____ /3

1 generic, role-agnostic assistant

2 some role tailoring; human checkpoints informal

3 built role-by-role; human ownership of consequential action enforced

Q1. Was this built role-by-role from real workflows, or is it a general-purpose assistant configured for us?

Evidence to request: the role and workflow it was designed around.

Vendor's answer: _____

Evidence received: ☐ Yes ☐ No

Q2. Where does a human review or approve before the agent takes a consequential action?

Evidence to request: the checkpoint, shown in the product.

Vendor's answer: _____

Evidence received: ☐ Yes ☐ No

Value Creation

Score: _____ /3

1 value asserted; no baseline or metric

2 metrics tracked, but no clean before/after

3 measured against the human baseline; ROI mapped to a number

Q1. What is the before/after number, measured against how our team does this work today?

Evidence to request: a baseline and a result from a comparable customer.

Vendor's answer: _____

Evidence received: ☐ Yes ☐ No

Q2. How will we measure value during our own pilot, and who owns that measurement?

Evidence to request: the metrics and reporting provided.

Vendor's answer: _____

Evidence received: ☐ Yes ☐ No

Orchestration

Score: _____ /3

1 system-bound; hand-offs manual

2 some connections; hand-offs brittle or context lost

3 coordinates across systems, agents, and approved data; guardrails upheld at every hand-off

Q1. Which of our systems, data sets, and other agents does it work across, and what happens at each hand-off?

Evidence to request: an architecture view of a live deployment.

Vendor's answer: _____

Evidence received: ☐ Yes ☐ No

Q2. How does it carry context and guardrails from one agent or step to the next?

Evidence to request: a demo of a multi-step task end to end.

Vendor's answer: _____

Evidence received: ☐ Yes ☐ No

Governance

Score: _____ /3

1 no named owner; no in-system guardrails

2 named owner; guardrails at the top level only

3 named owner and escalation path; guardrails enforced at agent and sub-agent level

Q1. Are guardrails, SOPs, and disclaimers enforced in the system at both the agent and sub-agent level, or only at the top?

Evidence to request: the configuration itself, beyond a policy document.

Vendor's answer: _____

Evidence received: ☐ Yes ☐ No

Q2. Who is the named owner on your side, and what is the escalation path when something goes wrong?

Evidence to request: a named contact and a documented escalation process.

Vendor's answer: _____

Evidence received: ☐ Yes ☐ No

Security

Score: _____ /3

1 no clear data boundary; no one can switch it off

2 standard protections; revocation manual and untested

3 data protected, IP contained, least-privilege access, tested deploy-and-revoke

Q1. Where does our data go, and is any of it used to train models outside our environment?

Evidence to request: data-flow documentation and contract terms.

Vendor's answer: _____

Evidence received: ☐ Yes ☐ No

Q2. Can we revoke a single agent's access without disrupting the rest, and have you tested it?

Evidence to request: the access controls and a record of a revocation test.

Vendor's answer: _____

Evidence received: ☐ Yes ☐ No

Testing & Certification

Score: _____ /3

1 a black box; outputs not logged or graded

2 basic logging and spot checks; no re-certification

3 a glass box; full audit trail, graded and re-certified on a schedule or trigger

Q1. Can we see what the agent did and why, for any given output?

Evidence to request: the audit trail and observability tools.

Vendor's answer: _____

Evidence received: ☐ Yes ☐ No

Q2. How is the agent's performance graded, and how often is it re-certified?

Evidence to request: the test standard and the re-certification schedule or trigger.

Vendor's answer: _____

Evidence received: ☐ Yes ☐ No

Total: _____ / 18

Floor check: Governance = 3? ☐ Yes ☐ No · Security = 3? ☐ Yes ☐ No

Tier: ☐ High (16–18) ☐ Medium (12–15) ☐ Low (below 12)

■ A score below 3 on Governance or Security caps the result at Low, regardless of the total.

NEXT If the vendor moves forward, scope the pilot with the **Agent JD Template**, and re-score the system with the **RAIL Self-Grading Scorecard** once it's running.