

Turn your two lowest areas into two actions

Keep this with your card. The evidence below is what level 3 usually looks like; your use case may need more or different evidence. Owners are typical, not prescribed.

HOW TO WRITE ONE ACTION

- **1. Read the next described level.** Level 3 if you scored 1 or 2; level 5 if you scored 3 or 4. The statement you could not tick is your action.
- **2. Write it as a thing that will exist.** "Improve governance" is a theme, not an action.
- **3. One owner, one date.** A name, not a team.
- **4. Evidence you would see, not be told.** A document, a log, a test result, a permission list.

Example, Team A, area 6: Give the analytics agent its own service identity, with read access to approved data marts only and every query logged as the agent's. Owner: head of data platform. Date: before the wider rollout. Evidence: the access list and a query log showing the agent's identity.

WHAT LEVEL 3 EVIDENCE USUALLY LOOKS LIKE

Area	Evidence at level 3 usually looks like	Usually owned by	A first action from 1 or 2
1 Business outcome	A one-page brief: named owner, a target in cost or cycle time, and a baseline measured before build.	The business owner of the workflow	Measure today's baseline for the target before building further.
2 Risk appetite	A written rule by type of action (read, draft, change, irreversible) saying which need a person's sign-off, and permissions and approval steps in the system that match it.	The business owner, with quality or risk	Write the rule for this use case; have the top line enforced in permissions, not in the prompt.
3 Workflow design and recovery	A documented handoff: when it stops and passes to a person. A documented, tested way to switch it off and recall or reverse what it did.	The process owner, with engineering	Write the stop-and-recall procedure and run it once.
4 Human-in-the-lead	Named reviewers qualified for the work, with time budgeted for review, a view of the agent's sources and trail, and authority to stop the workflow.	The function lead	Name the reviewers and give them the sources and trail view.
5 Data provenance	Every material claim carries its source and version. A list of the sources it can reach, each governed and documented.	The data owner or data governance	List what it can reach; remove anything not governed.
6 Identity and entitlements	Its own identity or narrowed access, logged as the agent. A list of approved tools and connectors (including MCP servers, skills and plug-ins). Code it writes runs in isolation. Calls between agents authenticated, if it makes any.	IT security, identity and access	Move it off a person's login onto its own identity with least-privilege access.
7 Evaluation and monitoring	A test set of your own cases with known right answers, pass thresholds agreed first, and cases built to mislead it. Production runs traced and scored against them.	Engineering, with QA or validation	Build a test set from past cases with known answers; agree the thresholds before the next release.
8 Unit economics and impact	Cost per accepted case, including review, corrections, retries and support. Spend and step limits. Impact measured against the area 1 target.	A finance partner, with the business owner	Track cost per accepted case for one month, review time included.

AREA 7: WHAT TO MEASURE

Pick the measures that fit what the agent does, and agree the pass thresholds before you see results.

Task success rate	share of workflows completed correctly without escalation
Precision and recall (F1)	for classification, extraction and decisions
Retrieval accuracy	correct documents or facts retrieved against a known set
Semantic similarity	generated output against a reference answer
LLM as judge	outputs scored against gold standards, with the judge itself checked
Bias detection	differences in outcomes across user groups
Hallucination rate	unsupported or incorrect claims
Calibration error	whether its confidence matches how often it is right

Source: McKinsey, One year of agentic AI: Six lessons from the people doing the work, 12 Sep 2025 (evaluation types). Descriptions shortened.

RULES TO KEEP IN MIND

- **A U is a finding.** Write who would know. Three or more U's, and your first action is finding out.
- **Cannot apply?** A statement that cannot apply to your design (it writes no code; it calls no other agent) counts as shown. Write why.
- **Plans go beside the score.** Score what is in place today; write planned controls next to it.
- **Ties.** Take the area that blocks what the agent will do. If still tied: 1, 5, 2, 3, 6, 4, 7, 8.
- **The ceiling moves one step at a time.** Fix the lowest, and the next lowest becomes your ceiling. Re-score in 90 days, or sooner if the agent's job changes.
- **Scaling goes in stages.** A ceiling of 3 says nothing stops you; it does not prove reviewer capacity or cost at the next scale. Check both at each step.

NO USE CASE YET? PICK WORK THAT PASSES FOUR TESTS

- **1.** It needs judgment on messy inputs: documents, text, exceptions.
- **2.** It happens often, and you can measure today's effort.
- **3.** A mistake would be caught and undone before it reaches a customer, a patient, a regulator or a production system.
- **4.** Someone would own the result.

Then check it needs an agent: does it act on its own along the way, and does it pick its own next step? Steps known and inputs structured? Code or rules will do, and cost less.

The Agentic Readiness Gate, instrument v2.6. A directional self-assessment, not a validated audit. © 2026 Vivek Hans, Genitive.AI. Licensed CC BY 4.0. genitive.ai/agentic-readiness-gate. DOI 10.5281/zenodo.23021021.