

THE AGENTIC READINESS GATE

A diagnostic of your organization's readiness to deploy and govern agentic AI, scored on evidence, one use case at a time.

Eight readiness checks, with what levels 1, 3 and 5 look like. Keep this side for reference. Score on the back.

AI: you govern the output. Agents: you govern the system. These eight measure whether you can.

Levels: Only 1, 3 and 5 are described. Tick the level 3 statements you can show: none 1, some 2, all 3. All of level 3 plus some of level 5: 4. All of level 3 and 5: 5. A statement that cannot apply to your design counts as shown; write why.

VALUE | DECIDES WHETHER TO SCALE A gap here gets the project cancelled. It does not lower your ceiling.

1 BUSINESS OUTCOME

How clearly is its purpose defined, owned and measured?

- 1 Exists because agents matter. No number named, no owner.
- 3 ☐ Named owner and target, usually cost or cycle time, ☐ with a baseline measured before build.
- 5 The outcome is sustained and evidenced over time, whatever the aim, and owned like any operating commitment.

8 UNIT ECONOMICS AND IMPACT

How well do you know what each accepted case costs, and what it delivers?

- 1 Neither. Cost surfaces on an invoice. Value is asserted, not measured.
- 3 ☐ Cost per accepted case tracked, including review, corrections, retries and support, ☐ with spend and step limits, ☐ and impact measured against the area 1 target.
- 5 Unit economics shape design before build: which model, how many steps, where a person reviews. Impact measured consistently enough to compare use cases.

CONTROL | SETS YOUR CEILING A gap here can hurt someone. Your ceiling is your lowest score in areas 2 to 7.

2 RISK APPETITE

How firmly are the actions it may take on its own set, and enforced by the system?

- 1 No stated position. Whoever builds it decides what it may do alone.
- 3 ☐ A written rule, by type of action (read, draft, change, irreversible), for which actions need a person's sign-off, applied to every new use case. ☐ Enforced through permissions and approval steps, not only instructions to the agent.
- 5 Tiered by reversibility, with standing approval only for routine actions, reviewed as the agent's track record and the models change, and teams can cite it unprompted.

3 WORKFLOW DESIGN AND RECOVERY

How well are its handoffs and its recovery designed around how it fails?

- 1 Agent added to the existing process. Checkpoints sit where the old steps were, not where this agent makes mistakes. No agreed way to stop it or undo what it did.
- 3 ☐ Handoffs documented, including when the agent must stop and pass to a person. ☐ A documented way to stop it and reverse or contain what it did.
- 5 Checkpoints placed where this agent's errors actually occur, based on its track record. Stopping it, rolling back its actions and revoking its access are rehearsed.

4 HUMAN-IN-THE-LEAD

How reliably would the final human check catch an error?

- 1 Oversight is a sign-off box. The signer lacks the time and the source material.
- 3 ☐ Named qualified reviewers ☐ who see the agent's sources and trail, ☐ with authority to stop the workflow.
- 5 Oversight itself is measured: override and edit rates tracked, known errors planted in controlled tests of reviewers, never in live records. Depth matches consequence.

5 DATA PROVENANCE

How easily can you trace what it used, including what it remembers?

- 1 Provenance reconstructed by hand when someone asks, if at all.
- 3 ☐ Retrieved statements carry their source and version. ☐ Reachable sources are governed and documented.
- 5 Provenance automatic end to end, including which version applied at the time. Retrieval scoped so it cannot surface what it should not. What it keeps between tasks is governed like any other source.

6 AGENT IDENTITY AND ENTITLEMENTS

How tightly is its access scoped, and how fully are its actions traced to it?

- 1 Runs under a shared account or a person's full credentials. Nobody could separate its actions from a person's.
- 3 ☐ Own identity, or a person's access narrowed to the task and marked as the agent's. ☐ Actions logged well enough to reconstruct. ☐ The tools, connectors and other agents it can call are approved and listed. ☐ Any code it writes runs in isolation. ☐ Calls to and from other agents are authenticated.
- 5 In access review and decommissioning like any privileged identity, with a named owner. If it reads outside content, it cannot send or change data without a check.

7 EVALUATION AND MONITORING

How well can you show it works, and how quickly would you know if it stopped?

- 1 You would find out from a complaint or an audit finding.
- 3 ☐ Before go-live, a test set of your own cases with known right answers and agreed pass thresholds, including cases built to mislead it. ☐ In production, runs traced and scored against them.
- 5 Any change to model, prompt, tools or permissions, data or memory, or the process it serves reruns the test set before release. Production failures are added to it. Changes you did not make are detected.

The Agentic Readiness Gate, instrument v2.6. Front: reference. A directional self-assessment, not a validated audit: a one-level difference can be noise. A full assessment is led by experts and draws on several sources. Re-run it as things change.

© 2026 Vivek Hans, Genitive.AI. Licensed CC BY 4.0. genitive.ai/agentic-readiness-gate. DOI 10.5281/zenodo.23021021.

Score what is in place today. On your own, in silence.

Use case you are scoring (one agent or several: write the job) _____

What the agent does, and what a person decides _____

Stage ☐ Idea · ☐ Pilot · ☐ Production · *Not built yet? Score what is in place today for the design you would deploy. Planned controls go beside the score, not in it.*

Does it need an agent? Does it act on its own along the way (run a query, call a tool, change a record), even if a person reviews the result? No: ☐ AI, not an agent. Yes: who picks the next step? Steps you wrote: ☐ A fixed workflow. It decides, from what it finds: ☐ An agent.

A fixed workflow or an agent: score it. Steps known and inputs structured? Code or rules will do, and cost less.

Score from the front. Tick the level 3 statements you can show: none 1, some 2, all 3. All of level 3 plus some of level 5: 4. All of level 3 and 5: 5. Cannot apply to your design? It counts as shown; write why. Torn between two? Take the lower.

Don't know? Mark U. It counts as 1 for your ceiling, marked unverified. Write who would know. Three or more U's, and your first action is finding out.

All eight apply. How much evidence is enough depends on the use case, its stage and what the agent does.

AREA	CIRCLE ONE	ONE FACT, OR THE QUESTION YOU WOULD NEED ANSWERED	WHO WOULD KNOW
VALUE			
1 Business outcome	1 2 3 4 5 U		
8 Unit economics and impact	1 2 3 4 5 U		
CONTROL			
2 Risk appetite	1 2 3 4 5 U		
3 Workflow design and recovery	1 2 3 4 5 U		
4 Human-in-the-lead	1 2 3 4 5 U		
5 Data provenance	1 2 3 4 5 U		
6 Agent identity and entitlements	1 2 3 4 5 U		
7 Evaluation and monitoring	1 2 3 4 5 U		

YOUR CEILING _____ *Your lowest score in areas 2 to 7. Not the average. A U there counts as 1: write “1 (unverified)”. It holds for this use case as written. It says what not to do yet; it never certifies that a rollout is safe.*

VALUE Area 1 _____ Area 8 _____ *Either below 3, or a U? Do not scale, even if your ceiling allows it.*

YOUR DECISION ☐ Proceed in stages · ☐ Narrow what the agent does · ☐ Keep testing offline · ☐ Use a simpler option

YOUR TWO LOWEST AREAS _____ and _____ *At least one must be a control area. If tied, take the one that blocks what the agent will do; if still tied: 1, 5, 2, 3, 6, 4, 7, 8.*

YOUR TWO ACTIONS

For each area, read the next described level: level 3 if you scored 1 or 2, level 5 if you scored 3 or 4. That is where your action must get you. You may swap one for a higher-scoring area that blocks what the agent will do, if you write down why. Fixing your lowest area moves the ceiling to the next one: re-run in 90 days to find it. The action helper shows what evidence usually looks like.

AREA	ACTION	OWNER	BY WHEN	EVIDENCE IT IS DONE

ROOM RESULT, ANONYMOUS. *Tear off along the line. Circle every area at your ceiling score, and your U count. Pass it to the aisle. No names.*

Ceiling area(s)	2	3	4	5	6	7	U's I marked	0	1 to 2	3 or more
-----------------	---	---	---	---	---	---	--------------	---	--------	-----------