

The Capstone Rubric

How a capstone is reviewed and credentialed

Version 1.1 · June 2026

The Capstone Rubric is the standard against which a participant's capstone is reviewed — at Session 4 of the AI for Managers program, and at Session 6 of the AI for Champions program. It applies to both tracks. A Manager's capstone is a working AI prototype for a real task on their team; a Champion's capstone is a function-level AI solution. The artifacts, and the standard, are the same.

The rubric is built directly on the Seven-Stage Process — the method participants use to design, build, and deploy their solution. Each criterion corresponds to a stage's defined artifact. The rubric does not introduce new requirements; it states the standard each artifact must reach.

HOW THE RUBRIC WORKS

Seven scored criteria. Each capstone is reviewed against seven criteria, drawn from the seven stages of the capstone method. Each criterion is rated at one of three levels.

Three levels. Below Standard — the artifact is missing, or too vague to act on. Meets Standard — the artifact is complete, specific, and defensible; this is the bar. Strong — the artifact is not only complete but well-judged, and could serve as a reference example for future participants.

The credential standard. A capstone credentials when it reaches Meets Standard on all seven criteria. The criteria are not averaged. The seven stages form a chain: a capstone with a strong value case but no named risk owners is not partially deployable — it is not deployable. Every criterion must hold.

The gating criterion. Criterion 4 — The Capstone — is the gating criterion. A Below Standard on any other criterion can be carried as a named 90-day commitment, with the credential still awarded. A Below Standard on Criterion 4 is disqualifying: there is no credential until it reaches Meets Standard. Criterion 4 is where the working solution itself lives; the rest of the rubric frames it or protects it.

Review, not a pitch contest

The Session 4 / Session 6 review is structured against this rubric — it is not a presentation competition. The output of the review is a clear picture: which criteria are green, which are not, and what specifically closes each gap. A Below Standard on a non-gating criterion converts directly into one of the participant's three 90-day commitments.

The Seven Criteria

Each criterion below shows the standard at all three levels. Meets Standard is the bar a capstone must clear on every criterion to credential.

#	Criterion	Below Standard	Meets Standard	Strong
1	Workflow Definition	A category, not a workflow (“improve marketing”). No clear beneficiary. “Why now” cannot be stated.	One concrete workflow. The beneficiary is named in a sentence. The “why now” is defensible in plain language.	The workflow is sharply scoped and the “why now” connects clearly to a real business pressure or opportunity.
2	Value Case	“Faster” or “better” with no numbers. No baseline. The improvement can only be described, not measured.	A baseline number, a target number, and a measurement date. Estimated figures are acceptable; absent figures are not.	Baseline and target are well-grounded, and the KPI clearly reflects what the workflow is meant to improve.
3	Tool & Approval	Tool unnamed, or its approval status is unknown or undocumented.	The AI tool is named, its approval status is documented, and a one-line statement explains why it fits the workflow.	Tool fit is genuinely well-reasoned against the workflow's real needs, not chosen by default or convenience.
4	The Capstone — inputs, prompt, human oversight GATING CRITERION	No working prototype, or the prompt lives only in one person's head. Inputs unnamed or of doubtful quality. No defined human review.	A working prototype. (a) Inputs are named, owned, and assessed as good enough quality for the workflow. (b) A detailed prompt built on the full 6-step framework — written down and reusable. (c) Defined human oversight rules: who reviews, what they check, how often, with authority to edit, reject, or escalate.	All three elements are not just present but well-designed: clean inputs, a precise reusable prompt, and human oversight placed deliberately where it changes the outcome.

#	Criterion	Below Standard	Meets Standard	Strong
5	Evaluation — tested before go-live	Untested, or “we’ll keep an eye on it.” No defined test cases.	Before go-live, the capstone has been run against a golden-source set of known test cases, with failures logged. The participant has identified where golden-source testing is insufficient — edge cases, unusual or adversarial inputs — and stated what additional testing those require, or made a defensible case that none is needed.	The test approach is thorough: golden-source coverage is strong and the gaps beyond it have been deliberately probed, not just acknowledged.
6	Risks & Mitigants	Generic risks (“bias,” “hallucination”) with no specific mitigant, or risks with no owner.	The top three to five failure modes specific to this workflow are named, each paired with a mitigant and a named owner.	Failure modes are sharply specific to the workflow, and the mitigants are practical and genuinely owned.
7	Data Governance	No data-handling answer, or “probably fine.” No data-owner sign-off.	A one-paragraph data-handling statement for the workflow, signed off by the data owner — even where the answer is “no protected data involved.”	The data-handling statement is precise about inputs, vendor handling, and default posture, with a clear sign-off.

Reading the table: Criterion 4 is shaded as the gating criterion. A capstone must reach the Meets Standard column on all seven rows to credential — a Below on rows 1, 2, 3, 5, 6, or 7 becomes a 90-day commitment; a Below on row 4 blocks the credential until resolved.

Beyond Credentialing: The Observability Commitment

The seven criteria above are all verifiable on the day of review — each artifact is either in the Capstone Dossier or it is not. Observability is different. It is a commitment about the future, and it cannot be demonstrated at the showcase because nothing has yet changed. For that reason it is not one of the seven scored criteria. It is, however, required.

What observability is. A capstone does not live in a stable world. Three things drift underneath it over time: the input data changes, the AI model is upgraded by the vendor, or the prompt is edited. Observability is a documented, owned process that states plainly: when any of those three things changes, the capstone is re-evaluated before it is trusted again. Pre-launch evaluation (Criterion 5) asks “is this good enough to launch.” Observability asks “will we know when it stops being good enough.”

HOW OBSERVABILITY IS HANDLED

- **A required dossier component.** The Capstone Dossier is incomplete without a documented observability process. The dossier names the three change triggers — data, model, prompt — and an owner responsible for the re-evaluation. An incomplete dossier cannot be presented for review.
- **Not scored at the review.** Because the process cannot be shown to have worked at the showcase, it is not assigned a Below / Meets / Strong rating. Its presence in the dossier is required; its operation is not scored at credentialing.
- **Evidenced at the 60-Day Review.** The 60-Day Deployment Review is where the observability process is evidenced — the sponsor-facing session at roughly day 60 examines whether the process held up: what changed, whether the capstone was re-evaluated, and what happened next.

This separation is deliberate. Every scored criterion on this rubric can be verified at the moment of review. Observability — a forward-looking commitment — is held to a standard that fits what it is: required in the dossier, and proven later, where it can actually be proven.

Worked Example — Criterion 4 at Each Level

Criterion 4 is the gating criterion and the most demanding. The example below shows the same capstone — a workflow that drafts first-pass responses to internal IT support tickets — described at each of the three levels, so a reviewer can calibrate.

Below Standard

“We use AI to help with IT tickets.” There is a prompt, but it is pasted fresh each time by whoever is on duty and varies between people. Inputs are “the ticket” — their source and quality are not examined. A person reads the output before it is sent, but no one has said what they are checking for or whether they may rewrite it.

Meets Standard

A working prototype drafts the response. Inputs are named — the ticket text and the relevant knowledge-base article — each with an owner, and both are confirmed current and reliable. The prompt is written out using all six steps of the framework, saved, and reused by everyone on the team. A named reviewer checks every draft for factual accuracy and tone before it is sent, and has explicit authority to edit or reject it.

Strong

Everything at Meets Standard, and: the prompt has been refined through several iterations and includes a worked example of a good response; input quality is actively maintained, with a check that flags stale knowledge-base articles; and the human review point is placed precisely — full review on tickets touching billing or security, lighter review on routine password resets — so oversight effort is matched to risk.

The distinction that matters most is between Below and Meets. The Below example is not lazy — a person is reviewing the output, a prompt does exist. It fails because the prompt is not reusable, the inputs are unexamined, and the review has no defined standard. Meets Standard is reached when all three elements are specific, owned, and repeatable.

USING THE RUBRIC

- **Before the review.** The participant assembles the Capstone Dossier — the seven stage-artifacts plus the observability process — and brings it to the review.
- **At the review.** Each of the seven criteria is rated Below / Meets / Strong against the standards in this document. The review produces a clear record, not a single score.
- **To credential.** Meets Standard on all seven criteria. A Below on Criterion 4 must be resolved first. A Below on any other criterion is recorded as one of the participant's three 90-day commitments.
- **After the credential.** The 60-Day Deployment Review revisits the capstone against its 90-day commitments and evidences the observability process in operation.