

AIPM AI Project of the Year Award: Judge Guide

Edition 1 (2026). Wave B, release candidate. Status: pending fold-in of the subgroup comments due 2026-07-15.

This judge guide is a **derived document**. It derives from, and is subordinate to, the operating rules in `operating-rules.md` (the single source of truth, the SSOT). Where anything in this guide could be read to conflict with the operating rules, **the operating rules prevail** and this guide is corrected. Section references in the guide (Rules Section N) point to the operating rules. Nothing here is decided until the founders' forum (Ricardo Vargas and Antonio Nieto-Rodriguez) ratifies it; every design choice below is the ambassadors' recommendation for the forum to ratify.

House-style note: this text avoids the em-dash and the en-dash by design. Commas, parentheses, and colons carry the same work, and ranges are written with the word "to" (for example, 1 to 10). Where a number is not fixed by the operating rules or the facts of the award, this guide states the rule at the principle level and marks the exact value as an operational parameter to be set and published before judging.

1. Purpose and how this guide relates to the operating rules

1.1 Purpose. This guide equips every juror, and the organizers who assemble the panels, to run scoring the same way across all regions and in the global round. It operationalizes the judging design that the operating rules fix: the two-tier model (Rules Section 7), the six-criterion weighted rubric (Rules Section 8), the cut-off (Rules clauses 2.6, 7.2, and 8.6), the Tier-2 fresh score (Rules clauses 7.6 and 8.7), the calibration requirement (Rules clause 7.7), the integrity rules (Rules Section 9), and the finalist feedback report (Rules Section 11).

1.2 What this guide adds. The operating rules state the composition principle for panels and point to this guide for the detail (Rules clause 7.8). This guide supplies that detail: how the committee forms, the judge-selection criteria at both tiers, the scoring walkthrough, the criterion-specific evidence examples, the calibration protocol, and the structured comment capture that Wave D assembles into finalist reports.

1.3 Rules-win note. This guide never restates a rule in a way that changes it. Weights, criteria, the anchor bands, the cut-off mechanism, the promotion counts, the calendar, and the integrity rules all live in the operating rules; this guide references and applies them. If a reader finds a divergence, the reader treats the operating rules as governing and reports the divergence to the working team so this guide is aligned (Rules clauses 13.1 and 13.6).

1.4 Derived scope. This guide governs the human judging process. The specification for the shared scoring sheet that jurors enter scores into is a separate derived document, `scoring-sheet-spec.md`, which itself derives from these rules and this guide.

2. Committee formation

This section is the enforceable machinery for how the judging committee comes into being. It is written as numbered clauses so the organizers can run it as a checklist and so a juror can see exactly how a panel was assembled.

2.1 The call for judges

2.1.1 The committee is formed through an open call for judges, not by closed invitation only. The call is seeded from the judge segment of the Step-1 interest-and-intent lists (the pre-announcement kit captured a nominate, judge, or amplify intent), and is supplemented by direct outreach to AIPM ambassadors in each macro-region (Rules clause 7.9.1 opens the interest-and-intent list).

2.1.2 The call states, up front and in writing: the tier structure (Tier 1 regional, Tier 2 global), the time boxes (Tier 1 in October 2026, Tier 2 in November to December 2026, per Rules clauses 7.2 and 7.4), the judge-selection criteria in Section 3 of this guide, the conflict-of-interest rules (Rules Section 9), the confidentiality undertaking (Rules clause 12.3), and the no-contact rule (Rules clause 9.3).

2.1.3 The call is non-commercial. Judging is unpaid and carries no fee in either direction, consistent with the free-end-to-end guardrail (Rules clauses 1.3.2 and 9 preamble: all integrity rules are zero-cost).

2.2 The formation process (six steps)

The committee is formed by the following six steps, in order. A step is not skipped.

2.2.1 Step 1: Call and sourcing. Publish the call for judges (clause 2.1). Draw candidates from the interest-list judge segment and from direct ambassador outreach in each region.

2.2.2 Step 2: Expression of interest. Each candidate submits an expression of interest with a short competency-and-affiliation declaration: the rubric competency domains they can cover (clause 3.1.2), their region link, and the affiliations relevant to conflict screening.

2.2.3 Step 3: Qualification check. The organizers check each candidate against the tier criteria in Section 3. A candidate advances only if they meet the tier they are being considered for.

2.2.4 Step 4: Conflict-of-interest declaration up front. Before any assignment is made, each qualified candidate declares conflicts of interest (own entry, employer, client, or any materially connected project), per Rules clause 9.1. Declaration precedes assignment; it is never retrofitted after a juror has seen an entry.

2.2.5 Step 5: Panel assembly. The organizers assemble each regional panel and the global jury to hit the size band and the competency mix in clause 2.3, screening the declared conflicts so that no juror is seated on a panel that will judge a connected entry.

2.2.6 Step 6: Publish the named panels. The organizers publish the panel composition and the conflict-of-interest policy before or at judging (clause 2.3.5).

2.3 Panel assembly rules

2.3.1 Size. Each panel is about seven jurors, within the founders' 5-to-10 range (Rules clause 7.8). No panel is seated below five jurors: below five, a single recusal would gut the panel. If a panel cannot reach five qualified, conflict-clear jurors from the ambassador bench, the shortfall is filled with criterion-eligible external jurors for that region (clause 3.1.1).

2.3.2 Mixed, not all-academic. Each panel is a mixed group, not an all-academic one, drawn from AIPM ambassadors and external experts (Rules clause 7.8).

2.3.3 Competency coverage. The assembled mix must, between its members, cover the five competency domains: project management and delivery, AI governance, ethics and risk, business value, and sector representation (Rules clause 7.8). No single juror needs all five; the panel as a whole must span all five.

2.3.4 Conflict declared up front. Conflicts are declared before assignments are made (clause 2.2.4), so a panel is never assembled around an undisclosed connection.

2.3.5 Named-panel publication. The panel composition and the conflict-of-interest policy are published (before or at judging). This is a deliberate contrast with the anonymous-jury anti-pattern: AIPM names its jurors and publishes its COI policy rather than hiding the bench.

3. Judge-selection criteria per tier

A candidate is qualified against the tier for which they are being considered. The criteria below are the design the pack constructs; they are consistent with the operating rules and keep any unset threshold at the principle level. Where this guide says a bar is qualitative, no numeric threshold is set, and none is to be invented.

3.1 Tier 1: regional panel juror

A Tier-1 juror must satisfy all five criteria below.

3.1.1 Regional link. The juror is an AIPM ambassador of that region, or is linked to that region by affinity rather than residence alone (assignment is by affinity, not only residence, per Rules clause 7.3), or is a criterion-eligible external juror serving that region where the ambassador bench is thin.

3.1.2 Competency. The juror has demonstrable expertise in at least one of the five rubric competency domains (project management and delivery, AI governance, ethics and risk, business value, or sector), selected so that the assembled panel covers all five (clause 2.3.3).

3.1.3 Integrity. The juror passes conflict-of-interest screening (no own entry, employer, client, or materially connected project, per Rules clause 9.1), signs a confidentiality undertaking (Rules clause 12.3), and accepts the no-contact rule (Rules clause 9.3).

3.1.4 Calibration. The juror completes the calibration exercise before any real scoring (Section 7 of this guide; Rules clause 7.7).

3.1.5 Availability. The juror commits to the October 2026 time box (Rules clause 7.2).

3.2 Tier 2: global jury

The global jury comprises the ex officio members, the ambassadors who pass the global criterion, and, only if Decision #8 is ratified, a sponsor juror (Rules clause 7.5).

3.2.1 Ex officio. Ricardo Vargas and Antonio Nieto-Rodriguez (Rules clause 7.5).

3.2.2 Ambassadors who pass the global criterion. An ambassador qualifies for the global jury by meeting the Tier-1 bar (competency, integrity, and calibration, per clauses 3.1.2 to 3.1.4) plus a higher bar: demonstrated cross-domain judgment (the ability to weigh outcomes, responsible-AI governance, adoption, solution fit, delivery rigor, and community contribution against one another) and a strong calibration outcome (close alignment with the reference scoring in the calibration exercise). This higher bar is qualitative and is assessed by the organizers; no numeric threshold is set for it.

3.2.3 Availability. A global juror commits to the November to December 2026 time box (Rules clause 7.4) and to attending the live finalist pitch (per Rules clause 7.4.2).

3.2.4 Conflict of interest. No global juror may judge a materially connected global finalist. A global juror with any such connection recuses from the affected finalist per Rules Section 9.

3.2.5 Sponsor juror (pending ratification).

PENDING RATIFICATION (Decision #8: Sponsorship posture)

Recommended: owner-funded, no sponsor for edition 1, and therefore no sponsor juror on the global bench. **Alternative:** if a sponsor is taken under the CloudFest sponsor-to-charity model, a sponsor juror from a technology company or a venture-capital firm may join the global jury. Such a juror serves under the published sponsorship-does-not-influence-judging firewall and under the same conflict-of-interest, disclose-and-recuse, confidentiality, and no-contact rules as every other juror, holding no additional weight in any decision (Rules clause 7.5; Decision #8 box in Rules Section 9). The forum decides.

4. The judging process walkthrough

This section describes the juror journey from assignment to feedback capture. The two-tier flow is fixed by the operating rules (Rules Section 7) and is only operationalized here.

4.1 The two-tier flow (referenced from the SSOT). Judging runs in two tiers, regional then global (Rules clause 7.1).

4.1.1 Tier 1 (regional, October 2026). Each region’s panel screens entries for eligibility and scores the eligible entries against the six-criterion rubric (Rules clauses 7.2.1 and 8.1). Each region promotes up to 2 or 3 entries that clear the published cut-off score. The cut-off is a floor of quality, not a seat count: slots are never filled below the cut-off (Rules clauses 2.6, 7.2.2, and 8.6).

4.1.2 Tier 2 (global, November to December 2026). The regional winners (about six to nine entries across the three macro-regions: Americas, EMEA, and Asia-Pacific) pitch live online to the global jury (Rules clauses 7.4.1 and 7.3). The global jury re-scores every global finalist from zero under the same rubric; regional scores do not carry into the global round (Rules clauses 7.6 and 8.7). The global jury then designates the overall winner and the two special recognitions (subject to the qualifying condition for the agentic recognition in Rules clause 3.4).

4.1.3 Calibration and the zero re-score work as a pair. The Tier-2 zero re-score and calibration are two halves of one mechanism (Section 7). The re-score removes the anchoring of the regional scores (regional scores do not carry), and calibration supplies the single shared scale that replaces that anchor, so the global result rests on one consistent reading of the rubric rather than on carried-over regional numbers (Rules clauses 7.6, 7.7, and 8.7).

4.2 The juror journey. Each juror moves through the following stages.

4.2.1 Assignment. After qualification (Section 3) and panel assembly (Section 2), the juror receives their entries. A juror who spots a conflict at this point (for example, recognizing a connected project) declines or recuses immediately, before reading the evidence (Rules clause 9.1).

4.2.2 Calibration. Before scoring any real entry, the juror completes the calibration exercise on the shared worked example (Section 7).

4.2.3 Reading the evidence pack. The juror reads the structured questionnaire, the evidence pack, and the short video for each assigned entry (Rules clause 6.1). The juror assesses what is evidenced, not what is asserted (Section 6, the claims-versus-evidence rule).

4.2.4 Scoring. The juror scores each of the six criteria from 1 to 10 in the shared scoring sheet, guided by the common anchors (Section 6) and the criterion-specific examples (Section 5). The sheet computes the weighted total automatically (see scoring-sheet-spec.md); the juror does no manual arithmetic.

4.2.5 Comment capture. At scoring time, the juror records one strength and one improvement area per criterion in the sheet’s per-criterion comment field (Section 8). This capture is mandatory, not optional.

4.2.6 Promotion (Tier 1) or designation (Tier 2). In Tier 1, the sheet flags whether each entry meets or exceeds the published cut-off, and the panel promotes up to 2 or 3 cut-off-clearing entries (clause 4.1.1). In Tier 2, after the fresh re-score and the live pitch, the global jury designates the overall winner and the two special recognitions (clause 4.1.2).

4.2.7 Feedback hand-off. The per-criterion comments captured for finalists feed the Wave D finalist feedback report (Section 8; Rules Section 11). No further work is asked of the juror to produce that report beyond the comments captured at scoring time.

5. The rubric with criterion-specific evidence examples

5.1 One funnel, one rubric. There is a single submission funnel and a single scoring rubric. All entries are scored against the same six weighted criteria (Rules clauses 3.1 and 8.1). Each criterion is scored from 1 to 10. The weights are published before the call opens (Rules clause 8.1) and are not altered here.

5.2 The six criteria. The “what judges assess” and “typical evidence” columns are the SSOT rubric wording (Farhad concept note Section 7) and are not altered.

#	Criterion	Weight	What judges assess	Typical evidence
1	Outcomes, benefits and impact	25%	Measurable value, before and after evidence, benefits realized, organizational or societal impact.	Baseline and actual metrics, benefits report, ROI, public value, operational improvements.
2	Responsible AI governance and Human-in-Command	20%	Ethics, risk controls, accountability, transparency, privacy, security, bias controls, and human oversight.	AI governance artifacts, risk or ethics review, privacy review, audit trail, human-approval model.
3	Adoption and change management	15%	Real operational embedding, user uptake, stakeholder trust, training, and sustained use beyond launch.	Adoption metrics, training materials, change plan, stakeholder feedback, usage data.
4	AI solution fit and innovation	15%	Appropriate and effective use of AI for the problem, originality where relevant, and fit-for-purpose design.	Architecture summary, model or tool rationale, performance evidence, solution demo.
5	Project management and delivery rigor	15%	Problem definition, governance, planning, delivery discipline, risk and issue management, and lessons learned.	Charter, plan, schedule, risk log, governance cadence, delivery artifacts.

#	Criterion	Weight	What judges assess	Typical evidence
6	Learning and contribution to the community	10%	Transferable lessons and the quality of the story the case can contribute to AIPM knowledge sharing.	Lessons learned, case-study draft, reusable practices, reflection on what others can adopt.
	Total	100%		

5.3 Design principle. The rubric rewards durable change and real adoption over model novelty. Criteria 1, 2, and 3 carry 60% of the weight combined, so a technically impressive AI solution cannot win without credible evidence of outcomes, responsible governance, and real adoption (Rules clause 8.2).

5.4 Criterion-specific examples of how the anchors apply. The common 1-to-10 anchors in Section 6 apply to every criterion. What changes per criterion is what counts as evidence. The banded reads below show how those anchors read for each criterion. They are illustrative guidance, not additional rules, and they do not add or change any threshold.

5.4.1 Criterion 1 (Outcomes, benefits and impact). A 9 to 10 shows a stated baseline, an actual result measured against it, and benefits that are realized rather than projected (for example, a benefits report tied to the baseline metric). A 5 to 6 asserts a result and offers a figure, but with no baseline to compare it to. A 1 to 2 claims impact (“saved significant cost”) with no number and no source.

5.4.2 Criterion 2 (Responsible AI governance and Human-in-Command). A 9 to 10 shows the governance artifact plus the human-approval model actually operating (an audit trail, a documented risk or ethics review, a point where a human signs off). A 5 to 6 names a policy but does not show it in use. A 1 to 2 asserts “the model is fair and safe” with no risk review, no bias control, and no oversight point. This criterion also anchors the Excellence in Responsible AI recognition (Rules clause 3.3).

5.4.3 Criterion 3 (Adoption and change management). A 9 to 10 shows sustained use beyond launch (usage data over time, a change plan that was executed, stakeholder feedback). A 5 to 6 shows the solution went live but shows uptake only at launch. A 1 to 2 shows deployment with no evidence anyone used it; note that deployment alone is not adoption (Rules clause 4.1.6).

5.4.4 Criterion 4 (AI solution fit and innovation). A 9 to 10 justifies why this AI approach fits this problem, with performance evidence. A 5 to 6 describes the solution but not why it was the right tool. A 1 to 2 invokes a fashionable technique with no rationale and no fit to the problem. Remember the design principle: novelty alone does not carry an entry.

5.4.5 Criterion 5 (Project management and delivery rigor). A 9 to 10 shows the delivery spine (charter, plan, risk log, governance cadence) and lessons learned. A 5 to 6 shows some artifacts but thin risk or governance discipline. A 1 to 2 offers a narrative of delivery with no supporting artifact.

5.4.6 Criterion 6 (Learning and contribution to the community). A 9 to 10 offers transferable, reusable lessons and a case-study-ready story others can adopt. A 5 to 6 reflects on the project but keeps the lessons local and non-transferable. A 1 to 2 offers no reflection beyond “it went well”. This criterion is where the community-first purpose of the award is scored.

6. Scoring anchors and the claims-versus-evidence rule

6.1 The 1-to-10 anchors (verbatim). The same 1-to-10 anchor scale applies to every criterion (Rules clause 8.4). It is reproduced here without alteration.

Band	Level	Description
1 to 2	Minimal or unsupported	Mostly assertion, little or no evidence, major gaps against the criterion.
3 to 4	Notable gaps	Some relevant content, but evidence is weak, incomplete, or not yet operationalized.
5 to 6	Adequate	Meets a reasonable baseline, evidence exists, but impact or maturity is limited.
7 to 8	Strong	Well-evidenced and credible, shows strong practice and transferable lessons.
9 to 10	Exceptional	Outstanding evidence, durable value, responsible practice, and high learning value for the community.

6.2 Claims versus evidence. Judges must distinguish claims from evidence. Unsupported assertions score in the 1 to 2 band and are not credited (Rules clauses 8.5 and 9.4). A confident narrative with no artifact behind it does not earn a higher band; the artifact, the metric, or the demonstrated practice is what moves a score up. When a juror is tempted to reward a strong-sounding statement, the test is simple: could an auditor verify it from the evidence pack? If nothing backs it up, it scores as minimal or unsupported.

6.3 Apply the anchors as written. The anchors, not a juror’s private scale, govern the score. Calibration (Section 7) exists so that a 9 out of 10 means the same thing across every panel and in the global round.

7. The calibration protocol

7.1 Calibrate before real scoring. Before scoring any real entry, every panel scores a shared, de-identified worked example against the rubric and the anchors, then discusses divergence to align. This is fixed by the operating rules (Rules clause 7.7) and is a qualification criterion for every juror (clauses 3.1.4 and 3.2.2).

7.2 Cross-panel calibration. All regional panels use the same shared worked example, so that a 9 out of 10 means the same thing across the three macro-regions (Rules clause 7.7). Calibrating on a common example is what keeps the regional cut-off comparable across regions.

7.3 Global-jury calibration. The global jury calibrates before the Tier-2 fresh scoring begins. Because Tier 2 re-scores every global finalist from zero and regional scores do not carry (Rules clauses 7.6 and 8.7), the global jury aligns on the anchors afresh before the re-score, on the same shared basis. Calibration and the zero re-score work as a pair: the re-score removes the anchoring of the regional scores, and calibration supplies the single shared scale to replace it, so the from-zero re-score starts aligned rather than adrift.

7.4 Divergence discussion. Calibration is not only scoring the example; it is discussing why scores diverged and aligning on the reading of the anchors, so that later real scores are applied consistently. The goal is not that everyone lands on the same number, but that everyone agrees on what evidence justifies which band.

7.5 A worked calibration example (illustrative and de-identified). Use this to see how the anchors, the claims-versus-evidence rule, and the weighting fit together. The entry below is fictional and stripped of identifying detail; the numbers illustrate the method, not any real result, and they add no threshold.

Illustrative entry: “an AI triage assistant for a public service desk.” The questionnaire claims a large reduction in handling time and says the tool is “trusted and widely adopted.” The evidence pack contains: a baseline handling-time metric and a post-launch actual (Criterion 1); a documented risk review and a human-approval step before any automated response is sent (Criterion 2); usage data for the first launch week only, with no later data (Criterion 3); an architecture summary and a rationale for the model choice, with light performance evidence (Criterion 4); a charter and a risk log but no governance cadence (Criterion 5); and a short lessons-learned note with one reusable practice (Criterion 6).

A calibrated panel might reason as follows.

- **Criterion 1:** baseline plus actual is real evidence, but benefits realization is partial. Place it Strong (7 to 8), not Exceptional, because the durability of the benefit is not yet shown.
- **Criterion 2:** governance artifact plus a working human-approval step is strong, evidenced practice. Place it Strong (7 to 8).
- **Criterion 3:** launch-week usage only, with nothing showing sustained use. The “widely adopted” claim is not backed beyond launch, so it is largely a claim. Place it Adequate (5 to 6), and note the missing sustained-use evidence.
- **Criterion 4:** rationale plus light performance evidence is a reasonable baseline. Place it Adequate to Strong (6 to 7).
- **Criterion 5:** charter and risk log exist, but the governance cadence is missing. Place it Adequate (5 to 6).

- **Criterion 6:** one reusable practice, lightly reflected. Place it Adequate (5 to 6).

The point of the exercise is the discussion of divergence: if one juror gave Criterion 3 a 9 because the entry said “widely adopted,” calibration surfaces that this was a claim, not evidence, and pulls the panel back to the anchor. The weighted total is then computed by the sheet (weights 0.25, 0.20, 0.15, 0.15, 0.15, 0.10; see `scoring-sheet-spec.md`), and the sheet compares it against the published cut-off (a value set and published before judging, consumed by the sheet, not decided here). No juror does the arithmetic by hand.

8. Structured feedback-comment capture

8.1 Per-criterion capture. The shared scoring sheet has a per-criterion comment field. For each of the six criteria, the juror records one strength and one improvement area (see `scoring-sheet-spec.md`).

8.2 Mandatory at scoring time. Capturing the strength and the improvement area is mandatory, not optional, and it happens at scoring time, not later. This is what lets Wave D assemble the finalist feedback report cheaply and consistently from the comments already on the sheet, consolidated against the published rubric (Rules clauses 11.1 and 11.3). If jurors skip the comment, the feedback report cannot be assembled without going back to the jury.

8.3 Finalists only. Only finalists receive a feedback report. No individual feedback is given to non-finalist nominations (Rules clause 11.2). Jurors still capture comments on every entry they score, because an entry may become a finalist (at Tier 1 the panel does not yet know which entries will clear the cut-off) and because the comments support the audit trail (Section 9 and `scoring-sheet-spec.md`), but the assembled feedback report is issued to finalists only.

8.4 Criteria-anchored, written for the nominee. Each comment is anchored to its criterion and to the anchor band it reflects, so the finalist report reads as strengths and improvement areas against the published rubric rather than as free-form opinion. Because a finalist will read a consolidated version of these comments, the juror writes the strength and the improvement area as specific, evidence-anchored observations, not as one-word verdicts.

9. Integrity reminders for jurors

All integrity rules are published and are zero-cost. This section restates the juror-facing duties and cross-references Rules Section 9, which governs.

9.1 Conflict of interest (auto-decline plus disclose-and-recuse). No juror judges their own entry, employer, client, or any materially connected project. A juror with any such connection declines the assignment automatically; any residual conflict is disclosed and the juror recuses from the affected entry (Rules clause 9.1). In the global round, no juror judges

a materially connected global finalist and recuses per Rules Section 9 (clause 3.2.4 of this guide).

9.2 No contact with judges. During the judging period there is to be no contact with judges by nominees or their representatives. A juror who is contacted reports it rather than engaging; breach results in disqualification of the entry (Rules clause 9.3).

9.3 Confidentiality. Every juror gives a confidentiality undertaking covering the entries and the deliberations they see (Rules clauses 9.5 and 12.3). Entries and evidence packs are not shared or discussed outside the panel.

9.4 Claims versus evidence. Jurors distinguish claims from evidence and score unsupported assertions in the 1 to 2 band (Rules clauses 8.5 and 9.4; Section 6 of this guide).

9.5 Data forwarding. A regional entry's data may be forwarded to the global jury under the nominator's consent-to-forward (Rules clause 12.2); jurors handle forwarded material under the same confidentiality undertaking.

9.6 Audit rights. AIPM reserves the right to verify any information in an entry and to request supporting evidence during evaluation; material misrepresentation disqualifies the entry (Rules clause 6.4). Jurors flag suspected misrepresentation to the organizers rather than adjudicating it alone.

9.7 Governance and ownership. The operating plan is owned by the Community and Growth stream, under lead Eugene De Wee. The edition-1 working team, formed 2026-07-09, is Vitor (Lead), Farhad Abdollahyan (deputy lead), Tooran Khosh, Marcin Nowakowski, Maria Antonieta Huertas, and Mohammed Al Hadeeth, with Fabricio Costa as peer reviewer. Nothing in this guide is decided until the founders' forum (Ricardo Vargas and Antonio Nieto-Rodriguez) ratifies it (Rules clauses 13.2, 13.3, and 13.5).

End of judge guide (edition 1, Wave B release candidate). Derived from operating-rules.md; the operating rules prevail on any conflict. Pending fold-in of the subgroup comments due 2026-07-15 and ratification by the founders' forum.