

Verification Kit

Three-layer, cross-model verification with human adjudication, applied to Unlocking the Invisible Mile

Use this kit to verify any AI-assisted work before it is released. It implements Section 6 of ADAM and AI (v1.0) and adapts the author's reflexive, cross-model benchmarking method. Layer 1 freezes the work. Layer 2 asks several leading models to evaluate it independently. Layer 3 compares their evaluations and records a person's verdict on every issue. The result is a one-page summary that any colleague can read.

How to use this kit

1. Freeze the work as a numbered version. Record its checksum, the systems that produced it and the dates.
2. Give Prompt V1 and the work to at least three leading models, one at a time and in separate sessions. Give one of them the adversarial variant.
3. Save each evaluation. Record each model's name, version and date in the table in Section 3.
4. Give Prompt V2 and all the evaluations to a model that was not an evaluator.
5. For every consolidated issue, check the source yourself and complete the adjudication record (Section 4).
6. Record errors by type and the time spent (Section 5), then write the one-page summary (Section 6).
7. Correct the work, release it under a new version number and keep this kit with the project.

1. Layer 2 evaluation prompt

V1. Layer 2 evaluation prompt (give to each evaluator model with the artifact)

You are an independent reviewer under the Agentic AI Checkpoint Protocol (Phase 5: verify and trace). Evaluate the attached work. Do not rewrite it.

1. Quantitative: count the cited works and classify each as primary/official, technical/scholarly, news/trade, encyclopedia, travel/aggregator/vendor/commercial, or malformed/unresolvable. List the key numeric claims and the source cited for each.
2. Identity: for each DOI, state whether it resolves and whether title, authors and year match. Note any author or institution you cannot confirm.
3. Claim-source fit: rate each key claim as supported, weakly supported, unsupported or cannot verify. Flag claims whose source type does not fit them.
4. Method transparency: could the method be reproduced from the text? What is missing?
5. Qualitative: rate 1-5 for evidence quality, method transparency, internal consistency, clarity and fitness for the intended audience.
6. Level: at what professional level is the work written? One paragraph.
7. Limits: list what you could not check and why. Do not invent sources.

[Adversarial variant: your only task is to find where the work's integrity fails. Report every weakness you can substantiate.]

2. Layer 3 meta-analysis prompt

V2. Layer 3 meta-analysis prompt (give to a different model with all evaluations)

You will receive independent evaluations of one work from several AI models. Compare them: which issues all evaluators raised, which some raised, which only one raised; where ratings differ; which judgments seem too generous or too harsh. Produce a comparison table and a consolidated list of issues ranked by severity, each with a recommended correction. Add no claims about the work that the evaluations or the work itself do not support.

3. Evaluators and their ratings

One evaluation has been completed. Add a row for each further evaluator, and record your own rating in the last row. Ratings use the 1–5 scales in Prompt V1.

Evaluator (model, version)	Date	Adversarial?	Evidence	Transparency	Consistency	Clarity	Level of the work (one line)
Claude (Anthropic)	Sept 2026	No	source audit only	–	–	–	Source-type audit and six flags; see Section 4
Human adjudicator		–					

4. Human adjudication record

The first six rows are the issues flagged by the first evaluation. Add the issues consolidated in Layer 3. For each, open the source or recompute, then record your verdict and action.

#	Issue	Raised by	Your check (source opened, DOI resolved, recomputed)	Verdict (confirmed / partly / not)	Action (correct / remove / keep with note)	Min.
1	Island distances cited to a route-planning aggregator	Claude				
2	Airport transfer times cited to a transfer company's blog	Claude				
3	Soft-ground conditions cited to a design-competition website	Claude				
4	The ROI weights, scales and per-factor scores are not stated in the main report	Claude				
5	Several cost figures appear as estimates without a stated derivation	Claude				
6	One reference is a search-engine link rather than the document itself	Claude				

#	Issue	Raised by	Your check (source opened, DOI resolved, recomputed)	Verdict (confirmed / partly / not)	Action (correct / remove / keep with note)	Min.

5. Error and time log

Error type	Count	Examples
Citation errors (missing, wrong or unresolvable)		
Claim–source mismatch		
Instructions lost between sessions		
Silent loss of content between versions		
Numeric or code errors		
Graphics or layout defects		
Identity errors (author, institution, DOI metadata)		

Time	Hours
Creating the work	
Layer 2 evaluations	
Layer 3 meta-analysis	
Human adjudication and correction	

6. One-page verification summary

V4. One-page verification summary

Work and version: [title, version, checksum] Tier: [T] Date: [date]
 What was checked: [claims, citations and DOIs, numbers, code, identities]

How: [evaluators and versions; adversarial auditor; human checks performed]
What was found: [issues confirmed, partly confirmed, not confirmed]
What changed: [corrections made; items removed; notes added]
What remains uncertain: [open questions and why]
Checked rate: [n of n material claims] Sign-off rate: [n of n irreversible acts]
Signed off by: [name, ORCID id, date]

From ADAM and AI: Agentic Directed AI Modeling (ADAM) and the Agentic AI Checkpoint Protocol (AACP), Version 1.0 (2026). CC BY 4.0.