AI Researchers
Embedding Responsible AI in Research

AI READINESS TOOLKIT

UC6-01 Use case 6: Peer review and refereeing

AI referee agents for pre-submission review

A doctoral researcher stress-tests her own manuscript before it reaches her supervisor

AT A GLANCE

A doctoral researcher runs three independent AI referee agents over her own manuscript before sending it to her supervisor. One checks the conceptual contribution, one the methods and evidence, and one reads it as a journal editor would. The case supplies all three prompts, a decision log and a verification rule. The student and supervisor decide what to change and when the manuscript is ready.

TYPE
Case study

FORMAT
PDF

CONTRIBUTING PARTNER
FEP, University of Porto

WHO IT IS FOR
Doctoral researchers, postdocs and the supervisors who meet them

FRAMEWORK BARRIERS
Protecting expertise, The rules, This task

HUMAN-ONLY BOUNDARY

The AI agents may surface weaknesses, questions and revision options. The PhD student and supervisor verify the evidence, decide what to change, retain authorship, and decide when the manuscript is ready to circulate or submit.

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.

The situation

Mariana is a second-year PhD student at the University of Porto. She has a full draft of her first article: the introduction is written, the empirical section runs, and the tables are finally in place. In seven days she wants to send the paper to her supervisor, not as a perfect manuscript, but as a serious draft that can receive useful guidance.

She has one worry. Because she has lived with the paper for months, she may no longer see its weak points. A colleague can read the abstract, but not the methods. Her supervisor will help, but Mariana wants to arrive prepared: with the main vulnerabilities already identified, the strongest claims checked, and a short list of decisions that genuinely need supervisory judgement.

This use case gives Mariana three independent AI referee agents. Each agent reads the same authorised draft from a different angle. None of them decides whether the paper is good, and none of them rewrites the manuscript. Their job is to help Mariana see what she should verify before asking her supervisor for higher-level feedback.

The use case in brief

A doctoral researcher uses three independent AI reviewer roles to stress-test an author-owned manuscript before sending it to a supervisor, co-author, seminar discussant or target journal. The three roles are deliberately separate: one reviews the conceptual contribution, one reviews the methods and evidence, and one reviews the manuscript as an editor or referee would.

The value is not that the AI becomes a referee. The value is that the student obtains three structured, traceable readings before the human conversation begins. The student then checks the reports, rejects weak suggestions, and prepares a short revision memo for the supervisor.

WHEN THIS IS USEFUL

> The student has a complete draft, not only an idea or an outline.
>
> The student owns the manuscript, or has explicit co-author permission to use the selected AI environment.
>
> The aim is to identify risks and questions before human supervision, not to outsource judgement.
>
> The draft can be processed without breaching confidentiality, data protection, funder, journal or institutional rules.

THREE INDEPENDENT AGENTS

AGENT	MAIN QUESTION	OUTPUT
1. Conceptual referee	Is the research question, contribution, mechanism and argument internally coherent?	Argument map and P0 to P2 conceptual findings
2. Methods referee	Do the data, design, identification, reporting and robustness support the claims?	Design reconstruction and evidence checks
3. Editor and referee	Will a reader understand the claim, structure, novelty and limitations?	Claim-consistency audit and likely objections

INDEPENDENCE RULE

Run the three prompts in three fresh conversations. Do not paste the report from one agent into another. Compare the three reports only after all three are complete.

EXPECTED OUTPUTS

Three separate AI reports: conceptual, methods, and editor or referee.

A one-page student revision memo: accepted issues, rejected issues, evidence to verify, and questions for the supervisor.

A decision log showing what the student changed, checked, postponed or brought to supervision.

Workflow for the doctoral researcher

STEP	ACTION	STUDENT DECISION
1	Prepare a clean version of the manuscript and remove restricted material.	If permission or platform approval is unclear, stop.
2	Fill the context block inside each of the three prompts.	Use the same context and manuscript version for all agents.
3	Run Agent 1, Agent 2 and Agent 3 in separate fresh conversations.	Save each report separately before reading across them.
4	Read the three reports and mark each high-priority finding as accept, reject, defer or check.	Do not revise from AI output until the evidence is verified.
5	Prepare a short memo for the supervisor.	Bring only the issues where human judgement is needed.

THE SUPERVISOR MEMO

After reviewing the three reports, Mariana prepares a brief note for her supervisor containing the paper title and central claim, the three strongest AI-identified risks, what she has already checked, what she thinks should change, and two or three questions where she wants supervisory judgement.

Example opening sentence. “Before our meeting, I ran three independent AI referee-style checks on my own draft. I have not accepted the outputs as authoritative, but they helped me identify three points I would like your judgement on: the scope of the contribution, the identification argument, and whether the abstract overstates the evidence.”

DECISION LOG

FINDING	DECISION	EVIDENCE CHECKED	SUPERVISOR QUESTION
	Accept / Reject / Defer / Check		
	Accept / Reject / Defer / Check		
	Accept / Reject / Defer / Check		

QUALITY CRITERIA

CRITERION	WHAT GOOD LOOKS LIKE
Traceability	Every major finding points to a manuscript location or a named piece of missing evidence.
Calibration	The agent separates observed text, inference, missing information and confidence.
Usefulness	The report helps the student prepare a better conversation with the supervisor.
Human control	No agent gives a publication decision, rewrites silently, or substitutes supervisor judgement.

The three prompts

Use the same manuscript and the same completed context block in each prompt. Replace all bracketed fields. Run the prompts separately, save the outputs separately, and compare them only afterwards.

PROMPT 1. AGENT 1, CONCEPTUAL AND THEORETICAL REFEREE

You are Agent 1: an independent conceptual and theoretical referee. You are reviewing my own authorised manuscript before I send it to my PhD supervisor. Do not decide whether the paper should be submitted. Do not rewrite the manuscript. Your task is to identify conceptual risks and questions I should verify before supervision.

Context: discipline [DISCIPLINE]; paper type [PAPER TYPE]; provisional title [TITLE]; target audience or outlet [TARGET]; central claim [CENTRAL CLAIM]; theory or literature area [THEORY / LITERATURE]; revision window [REVISION WINDOW]; main worry [MAIN WORRY].

Review only the supplied manuscript and this context. Do not invent references, journal rules, theory gaps or facts that are not in the manuscript. For each important issue, report: location; observed manuscript evidence; your inference; missing information; why it matters; suggested author action; verification needed; confidence (high, medium or low). Rank issues as P0 = threatens the central claim, P1 = material but not fatal, P2 = useful later.

Output: (1) a one-paragraph argument map; (2) the three to five most important conceptual findings; (3) claims or references I must verify manually; (4) questions I should ask my supervisor; and (5) one possible sharper contribution statement, labelled as a suggestion, not a rewrite.

PROMPT 2. AGENT 2, METHODS AND EVIDENCE REFEREE

You are Agent 2: an independent methods and evidence referee. You are reviewing my own authorised manuscript before I send it to my PhD supervisor. Do not decide whether the paper should be submitted. Do not rewrite the manuscript. Your task is to identify methodological, empirical, reporting and reproducibility risks I should verify before supervision.

Context: discipline [DISCIPLINE]; paper type [PAPER TYPE]; provisional title [TITLE]; target audience or outlet [TARGET]; central claim [CENTRAL CLAIM]; data or evidence [DATA / EVIDENCE]; method or design [METHOD / DESIGN]; revision window [REVISION WINDOW]; main worry [MAIN WORRY].

Review only the supplied manuscript and this context. Do not assume access to data, code, appendices, ethics documents or robustness checks unless they are supplied. For each important issue, report: location; observed manuscript evidence; your inference; missing information; why it matters; suggested author action; verification needed; confidence (high, medium or low). Rank issues as P0 = threatens validity, ethics or the central claim, P1 = material but not fatal, P2 = useful later.

Output: (1) reconstruction of the research design or evidence chain; (2) the three to five most important methods and evidence threats; (3) inconsistencies between text, tables, figures and claims; (4) minimum feasible checks within [REVISION WINDOW]; and (5) items that require data, code, ethics or source verification before I discuss the draft with my supervisor.

PROMPT 3. AGENT 3, EDITOR AND REFEREE

You are Agent 3: an independent editor and referee. You are reviewing my own authorised manuscript before I send it to my PhD supervisor. Do not decide whether the paper should be submitted. Do not rewrite the manuscript. Your task is to identify communication, structure, positioning and reader-objection risks I should verify before supervision.

Context: discipline [DISCIPLINE]; paper type [PAPER TYPE]; provisional title [TITLE]; target audience or outlet [TARGET]; central claim [CENTRAL CLAIM]; intended contribution [CONTRIBUTION]; revision window [REVISION WINDOW]; main worry [MAIN WORRY].

Review only the supplied manuscript and this context. Focus on title, abstract, introduction, research question, conclusion, tables, figures, terminology, novelty claims, limitation statements, and the match between evidence and claims. Do not invent journal rules, references or reviewer preferences. For each important issue, report: location; observed manuscript evidence; your inference; missing information; why it matters; suggested author action; verification needed; confidence (high, medium or low). Rank issues as P0 = likely to mislead or undermine the central claim, P1 = material but not fatal, P2 = polish for later.

Output: (1) a claim-consistency audit from title to conclusion; (2) the three to five most important editor and referee findings; (3) three likely reader or supervisor objections; (4) low-priority edits to postpone until substance is settled; and (5) a short list of questions I should bring to my supervisor.

AFTER THE THREE AGENTS

The student, not a fourth AI prompt, runs the critical analysis of the reports. The decision depends on evidence, feasibility and supervisory judgement.

HOW MARIANA USES THE OUTPUTS

She checks every P0 and P1 issue against the manuscript, tables, sources, data or code.

She groups overlapping findings but keeps disagreements visible.

She revises only points she can justify, and records what she rejected or postponed.

She sends her supervisor a focused memo rather than three raw AI reports.

Responsible use and transfer

> **Verification rule.** A fluent AI report is not a validated review. Before acting, the student must locate the cited passage, check the evidence where relevant, verify sources and rules, and record a decision with a rationale.

CONDITIONS FOR USE

> Use only the student's own manuscript, or another manuscript for which processing is explicitly authorised.
>
> Confirm co-author consent, institutional rules, data classification, funder obligations and outlet rules before uploading any real draft.
>
> Remove confidential correspondence, restricted appendices, identifiers, licensed material and sensitive data unless their processing is explicitly authorised.
>
> Keep web search and external retrieval off unless there is a clear reason and the student verifies every external claim from an authoritative source.
>
> Disclose AI assistance where institutional, funder, publisher or supervisory rules require it.

RISKS AND CONTROLS

RISK	CONTROL
Automation bias	Use the decision log and require manuscript evidence for every P0 and P1 finding.
False confidence	Separate observed evidence, inference, missing information and confidence.
Confidentiality breach	Stop if permission, data classification or platform approval is unclear.
Weak supervision conversation	Use the AI reports to prepare questions, not to bypass the supervisor.
Premature polishing	Postpone style edits until conceptual and methods issues have been checked.

> **TRANSFER NOTE**
>
> The same three-agent structure can be used by postdocs or faculty, but the doctoral version is especially valuable because it improves the quality of the next supervision meeting while keeping responsibility and judgement with people.

About this resource

TITLE	AI referee agents for pre-submission review
TOOLKIT REFERENCE	UC6-01, use case 6: Peer review and refereeing
DEVELOPED BY	FEP University of Porto, Fábio Duarte
SUGGESTED CITATION	AI Researchers (2026). AI referee agents for pre-submission review. AI Readiness Toolkit, UC6-01. FEP University of Porto, Fábio Duarte. airesearchers.eu
QUALITY ASSURANCE	Reviewed by the AI Researchers consortium for quality, categorisation and alignment with the AI Engagement Framework. External peer review by senior academics is under way, and the item may be updated as a result.
VERSION	1.0, November 2026
PROJECT	AI Researchers: Embedding Responsible AI in Research. Erasmus+ Cooperation Partnership in Higher Education, KA220-HED-66835D2B. October 2025 to March 2028.
CONSORTIUM	University of Lodz (coordinator), FEP University of Porto, UPF Barcelona School of Management, ACEEU, NIPI (OPI PIB), European E-learning Institute and Momentum Marketing Services

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.

AI Researchers
Embedding Responsible AI in Research