Region - Educational Research and Reviews, 2026, Volume 8, Issue 5
DOI: 10.32629/rerr.v8i5.5479
ISSN Online: 2661-4634
ISSN Print: 2661-4626

Allocating epistemic responsibility in generative AI-supported scientific inquiry: a teacher-orchestrated multi-agent framework

Ziyun ZHANG

Tianjin University of Technology and Education, Tianjin 300222, China

Abstract: Generative AI can supply an explanation before a learner has decided what evidence would make it credible. In scientific inquiry, AI-enabled assistance may improve an answer while removing responsibility for judging it. This paper proposes epistemic responsibility allocation as a guiding principle for educational multi-agent systems. A responsibility specifies what an actor must justify, what the actor is prohibited from doing, and when decision-making authority passes to a teacher. An Expert verifies approved sources, a Peer raises evidence-seeking challenges, and a Coach manages and withdraws instructional scaffolds. A ledger adapts Toulmin's structure to record claims, evidence, warrants, qualifiers, challenges, and student revisions. Four propositions alongside a proposed comparison against a single-agent interface make the framework empirically testable. Multiple agents deliver practical benefits only when permissions create genuinely distinct responsibilities among them.

Keywords: generative artificial intelligence; epistemic responsibility; scientific argumentation; multi-agent systems; learning design; teacher orchestration

1 Introduction

The problem became concrete during the design of a Unity game, in which matching computing concepts earned attempts at an Angry Birds-style challenge. An English card game and Android vocabulary app raised the same question: how much successful action came from the learner's judgement?

Generative AI sharpens the problem. A student can receive an account of photosynthesis without deciding which observation supports it, why it counts as evidence, or when the account needs qualification. A correct response can still perform the comparison the student needs to learn.

Adding an Expert, Peer, and Coach does not resolve this issue. Three prompts running on one single model may reproduce an error, stage artificial disagreement, or rewrite a student's claim. This paper explores whether responsibility can be allocated to keep AI-generated support contestable and preserve the necessity of learner judgement. Each actor is assigned an evidential duty, an action boundary, and an escalation rule.

2 The theoretical problem: support can displace judgement

Inquiry learning requires students to coordinate claims with evidence and revise models. AI-provided support can direct learners' attention without replacing their reasoning [1]. ICAP also separates visible activity from constructive contribution [2]. Interface activity cannot reveal who has produced a given inference.

Toulmin's model characterizes inferential reasoning: data support a claim via a warrant; backing supports the warrant, a qualifier marks its force, and a rebuttal states exceptions [3]. Studies in science education have adopted this pattern to examine student argument [4]. Revision is not an original component of the Toulmin model. It records learners' response after a warrant, qualifier, or challenge has been evaluated.

Recent work shows why the form of AI help matters. A teacher-focused Delphi study classified generative-AI activities across self-regulated learning phases [5]. In university chemistry, a system requiring an initial attempt and supplying hints rather than answers reported stronger self-regulation and higher-order thinking than ordinary ChatGPT use [6]. The context differs, but the study supports testing scaffold timing.

A 2024 review found limited end-user involvement in many human-centred learning analytics and AI systems, with unresolved questions about control and trustworthiness [7]. UNESCO's teacher framework places human agency and accountability within AI competence [8]. Teacher control should go beyond a simple override button, while traces must remain selective enough to prevent information overload.

3 Epistemic responsibility allocation

The Expert checks teacher-approved sources, distinguishes observation from inference, and reports uncertainty. It cannot judge a question's educational relevance or close an inquiry. The Peer must seek evidence, identify a rebuttal condition, or question a warrant; it cannot manufacture opposition.

The Coach may request a prediction, identify a missing comparison, or suggest a strategy, then reduce support after independent success. It cannot certify accuracy or generate text on behalf of students. The teacher controls goals, sources, fading criteria, escalation, and assessment.

The ledger links Claim, Evidence, Warrant, Qualifier, Challenge, and Revision. A warrant explains why evidence supports a claim; a qualifier records scope or uncertainty. A challenge may contain a rebuttal, but they are not identical. Records store author, source, intervention type, time, and relation to the prior version. The ledger shows change, not thought itself.

Consider a student investigating why seedlings exhibit reduced growth under low-light conditions. The student links height-measurement data to reduced photosynthesis. The Expert notes that height does not measure photosynthetic rate. The Peer asks whether water or temperature creates a rebuttal. The Coach requests a warrant and qualified prediction. The student revises, defends, or suspends the claim; the teacher receives an alert only when sources conflict or uncertainty stays high.

While a single agent can implement these functions, a single-output-channel design allows answering, verification, challenge, and coaching overwrite one another. Separation of roles matters only when permissions have consequences. The Expert retrieves approved sources and returns statements of uncertainty; the Peer challenges but cannot edit student text; the Coach has access to the history of provided hints but functions not as an answer-generating tool. A router logs actions and enforces teacher-defined escalation rules. Figure 1 compares the designs.

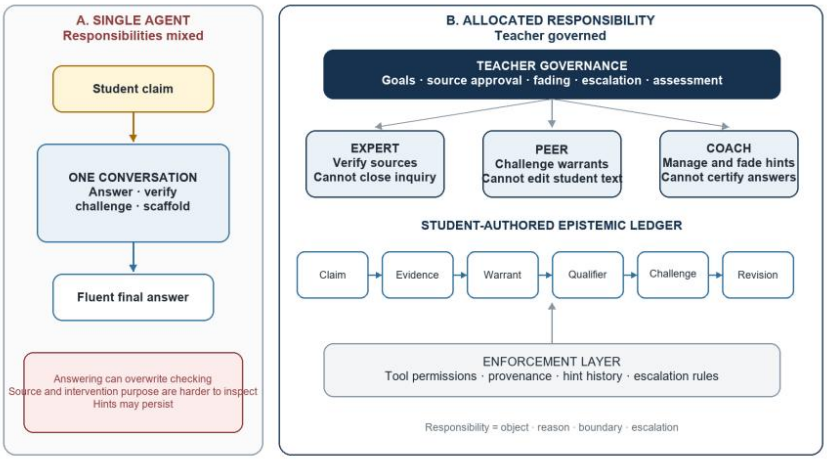

Figure 1. Single-agent responsibility mixing compared with teacher-governed epistemic responsibility allocation

The comparison is architectural, not theatrical. An Expert that edits a claim or a Coach that calls an answer tool creates an observable permission breach. Responsibility allocation can thus be verified via system- level behaviour, instead of being inferred purely from conversational tone.

4 Propositions and failure tests

P1, traceability: Students and teachers should be able to identify intervention source and purpose more accurately than under a single stream. Hidden instructions or sources undermine this design claim.

P2, student revision: The ledger should provide support only when students edit their claims and warrants. Automatic rewriting may improve the final output while reducing responsibility.

P3, productive fading: Reducing prompts after success should improve transfer, even if immediate fluency falls. Persistent help indicates dependence.

P4, bounded orchestration: Teacher responsibility should not push review time beyond an agreed classroom threshold.

Three modes of system failure are considered below. If a source retrieved by the Expert lies outside the teacher- approved collection or yields persistently high uncertainty, the Expert's output is blocked, and this event triggers escalation. A Peer challenge without an evidence request, questioned warrant, or rebuttal is rejected. If Coach- provided support persists even after students achieve independent success, the strength of the next hint is reduced. These checks can expose system- level failure, yet they cannot eliminate intrinsic foundation- model error.

5 Proposed study

Secondary science inquiry makes data, warrants, uncertainty, and rebuttals open to inspection. A design-based study would begin with two teachers choosing comparable units, approving sources, setting escalation rules, and testing the ledger [9]. It would aim to recruit 40 to 60 students from two classes. Across six lessons, each class would use both interfaces in counterbalanced order, with model, sources, time, and teacher held constant.

Coders would apply four proposed indicators: source evaluation, warrant articulation, uncertainty recognition, and student-authored justified revision. These are coding dimensions, not a validated scale. A delayed transfer task would test a new problem. Logs would capture hint dependence, escalation, and teacher review time; interviews would examine acceptance or rejection of interventions. Mixed-effects and qualitative analyses would account for repeated observations and polished answers with weak reasoning.

Privacy, age appropriateness, source bias mitigation, and a non-AI operational route would be design requirements. Better answers accompanied by fewer student decisions would indicate substitution, not genuine learning success.

6 Discussion and conclusion

Instead of judging how human-like AI agents appear, the framework examines what they may undertake and what evidence they must expose. Learning-technology design defines permissions; online infrastructure preserves traces; technology leadership governs sources, teacher capacity, review thresholds, and refusal.

Students may find the Peer agent distracting, teachers may regard the ledger as surveillance, and a constrained single-agent implementation may align with the framework at lower cost. Evaluation should examine whether students become better able to judge what their own explanations warrant.

Conflicts of interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Hmelo-Silver C E, Duncan R G, Chinn C A. Scaffolding and achievement in problem-based and inquiry learning[J]. Educational Psychologist, 2007, 42(2): 99-107. https://doi.org/10.1080/00461520701263368

[2] Chi M T H, Wylie R. The ICAP framework: Linking cognitive engagement to active learning outcomes[J]. Educational Psychologist, 2014, 49(4): 219-243. https://doi.org/10.1080/00461520.2014.965823

[3] Toulmin S E. The uses of argument (Updated ed.)[M]. Cambridge: Cambridge University Press, 2003.

[4] Osborne J, Erduran S, Simon S. Enhancing the quality of argumentation in school science[J]. Journal of Research in Science Teaching, 2004, 41(10): 994-1020. https://doi.org/10.1002/tea.20035

[5] Chiu T K F. A classification tool to foster self-regulated learning with generative artificial intelligence by applying self-determination theory: A case of ChatGPT[J]. Educational Technology Research and Development, 2024, 72: 2401-2416. https://doi.org/10.1007/s11423-024-10366-w

[6] Lee H Y, Chen P H, Wang W S, et al. Empowering ChatGPT with guidance mechanism in blended learning: Effect of self-regulated learning, higher-order thinking skills, and knowledge construction[J]. International Journal of Educational Technology in Higher Education, 2024, 21: 16. https://doi.org/10.1186/s41239-024-00447-4

[7] Alfredo R, Echeverria V, Jin Y, et al. Human-centred learning analytics and AI in education: A systematic literature review[J]. Computers and Education: Artificial Intelligence, 2024, 6: 100215. https://doi.org/10.1016/j.caeai.2024.100215

[8] Miao F, Cukurova M. AI competency framework for teachers[R]. UNESCO, 2024. https://www.unesco.org/en/articles/ai-competency-framework-teachers

[9] The Design-Based Research Collective. Design-based research: An emerging paradigm for educational inquiry[J]. Educational Researcher, 2003, 32(1): 5-8. https://doi.org/10.3102/0013189X032001005

About the author

Ziyun Zhang, born in February 2004, female, Qiang ethnicity, from Chengdu, Sichuan Province. She is an undergraduate student. Her research interests include educational technology, science education, and AI in education (applications of generative AI in education).