SOFIYA

Student- and faculty-first approach to AI-mediated dialogue

June 2026

Executive Summary

Generative AI has made it easy to produce polished answers. That is precisely why education needs tools designed around a different purpose: for the development and demonstration of human thinking.

Sofiya is a dialogic learning environment, supporting faculty to create bounded assignments, choose a pedagogical partner role, provide the relevant materials and expectations, and invite students into a sustained conversation. The AI asks; the student explains, defends, revises, and connects. The resulting transcript is both a learning process and a visible record of reasoning.

This inversion matters. A fluent model can generate plausible language without possessing a student's knowledge, experience, commitments, or responsibility for judgment. Sofiya therefore treats AI as an interlocutor, not as an epistemic authority. It is designed to create productive questions and intellectual friction while keeping cognition, authorship, and accountability with people.

CENTRAL IDEA

Sofiya is the AI that asks the questions. Students do the thinking. Faculty design the learning and remain the authority.

The external evidence is promising but conditional. A 2026 systematic review of 67 empirical studies found that scaffolded, inquiry-oriented uses of ChatGPT were associated with metacognitive regulation, argumentative reasoning, and idea generation, while unstructured use was associated with cognitive offloading and superficial engagement (Li et al., 2026). The lesson is not that AI is inherently good or bad for learning. The lesson is that pedagogy, task structure, and the distribution of cognitive work matter.

Sofiya's philosophy follows from that evidence: design the role of AI narrowly; make the student's reasoning the object of attention; make feedback formative; and reserve consequential judgment for faculty. The project is student-first because it protects the learner's agency and intellectual work. It is faculty-first because technology serves the instructor's pedagogical intentions rather than displacing them.

1. The Problem with Answer-First AI

Most general-purpose generative AI is optimized to respond. Ask for an explanation, summary, argument, code sample, or draft, and the system attempts to supply one. That can be useful in many settings. In education, however, it can quietly transfer the very work a learner needs to practise from the learner to the tool.

This creates more than an academic-integrity problem; it also creates a pedagogical problem. When a student receives a finished explanation before attempting one, accepts a polished argument without testing it, or submits text whose reasoning they did not build, task completion can become separated from learning. The output may look accomplished while revealing little about the student's actual understanding.

The current research supports caution about simple claims. Li and colleagues' review found a consistent contextual divide: structured tasks requiring verification, justification, comparison, and revision were associated with stronger cognitive engagement; unstructured summarization and essay-writing uses more often produced overreliance and surface engagement. The review's authors also stress that the underlying studies were methodologically heterogeneous, mostly short-term, and often non-experimental. Their synthesis identifies patterns, not a universal causal law (Li et al., 2026).

The essay debate makes the same issue visible from the assessment side. Corbin and colleagues (2026) argue that a polished final artifact can no longer be assumed to transparently represent the intellectual process that produced it. Their proposed response is structural: redesign assessment to make process, judgment, and dialogue visible, rather than relying only on prohibitions or attempting to detect AI-generated prose.

DESIGN IMPLICATION

The relevant question is not simply whether AI appears in a course. It is who performs the explanation, evaluation, justification, and revision that the learning outcome requires.

2. A Different Role for AI

AI as Interlocutor, Not Authority

Sofiya assigns AI a narrower and more educationally defensible role. The system is not there to know on the student's behalf. It is there to sustain a structured encounter in which the student must make their own understanding intelligible.

Wegerif and Casebourne (2026) propose a dialogic foundation for educational AI: technology can expand the space of dialogue without replacing human thought. Their account is conceptual, not an effectiveness trial, but it provides a useful design orientation. Learning is treated as participation in a living conversation, not the receipt of static information. The learner encounters another perspective, explains, listens, reconsiders, and contributes.

That orientation also requires epistemic humility. Quattrociocchi, Capraro, and Perc (2025) argue in a preprint that the fluent outputs of large language models can conceal structural differences between machine text generation and human judgment. Whatever one's position on their broader philosophical account, the practical warning is sound: linguistic plausibility should not be mistaken for warranted knowledge. Sofiya therefore keeps the model away from final authority. Course materials constrain the interaction; analysis points back to the transcript; and faculty remain responsible for interpretation and feedback.

How Refusal Boundaries Work

Sofiya's most critical technical behavior is its refusal to complete work that belongs to the student. This is enforced at two levels: the system prompt, which specifies what the partner role does and does not do, and a trained boundary layer that resists circumvention even when students probe, reframe, or pressure the system.

Faculty-defined assignment materials and role parameters are passed to the model as a bounded context. The partner draws only on that context and asks questions arising from it. When a student attempts to elicit a direct answer, the system is designed to redirect: to ask what the student thinks, to request an example, or to return the question to the student. Sofiya's pilot testing will measure the rate at

which this boundary holds, and any failure modes will be documented and addressed before broader deployment.

Productive Friction

Good dialogue does not merely make learning pleasant. It creates occasions when an explanation fails to land, a claim meets a counterexample, an assumption becomes visible, or evidence proves weaker than expected. These moments of productive friction are educationally valuable because the student must respond. Sofiya is designed to elicit that response rather than resolve the difficulty for them.

Productive friction is operationalized through a set of partner behaviors that faculty can calibrate. The partner can be instructed to request a concrete example before accepting an abstract claim, to surface an obvious counterexample after a student states a generalization, to ask for evidence when a claim is made without support, and to express genuine confusion when an explanation is incomplete rather than filling the gap. What the partner is explicitly not permitted to do is supply the missing content, endorse an answer without pressing it, or move on before the student has responded.

Faculty control the intensity of this friction. A first-year introductory course may call for a partner that asks clarifying questions but does not press counterarguments. A graduate seminar may call for one that challenges every unsupported claim. The disciplinary framing and level of challenge are assignment-level choices, not platform defaults.

3. Pedagogical Foundations

Partner Roles Faculty Can Choose

Sofiya offers a set of defined partner roles, each embodying a different pedagogical relationship. Faculty select the role that matches their learning goal, then frame and name it for their discipline and context. The following roles are currently in the design specification.

- **Peer Learner.** The partner takes the position of a curious student who does not yet understand the material. The learner being assessed plays the teacher: explaining, giving examples, fielding confusion, and demonstrating understanding through the act of making it intelligible to someone else. The partner asks what a definition means, how its parts relate, why it matters, and whether the student can apply it. It does not supply the missing content. This role draws from learning-by-teaching research (Roscoe & Chi, 2007; Fiorella & Mayer, 2013) and is well suited to concept mastery, procedural explanation, and introductory content across disciplines.
- **Debate Coach.** The partner takes a position opposed to the student's and presses the case for it. The student must defend their argument, respond to the strongest objection, qualify claims that are too broad, and engage with counterevidence. The partner's challenge is directed at the argument, never the person, and it does not seek to win. Its purpose is to keep inquiry open long enough for the student to examine the quality of their own reasoning. Faculty define the disciplinary context, the assigned materials, the task, and the standards of evidence. This role is suited to argumentation, essay preparation, ethical reasoning, and any task where students must hold a position under scrutiny.
- **Oral Examiner.** The partner conducts a structured oral examination modeled on a viva voce or professional review. It moves systematically through the student's work or knowledge domain, asking for justification, pressing on gaps, and requiring the student to connect ideas across the topic. Unlike the debate coach, it does not take a position; it evaluates. This role is suited to thesis defense preparation, professional licensure review, capstone projects, and any context where students must demonstrate comprehensive understanding under questioning.
- **Skeptical Reviewer.** The partner reads the student's submitted work or listens to their explanation and responds as a critical but fair peer reviewer would: identifying claims that need more support, noting logical gaps, asking why a particular approach was chosen, and pressing

the student to engage with literature or evidence they may have overlooked. This role is suited to research methods, writing-intensive courses, and professional programs where students must demonstrate that their work can withstand scrutiny.

- **Client or Stakeholder.** The partner plays the role of a non-expert with a real stake in the student's work: a patient in nursing, a planning client in architecture, a user in software engineering, a juror in law. The student must explain their reasoning, recommendations, or decisions in terms the client can understand and defend choices when the client asks follow-up questions. This role develops communication, professional judgment, and the ability to translate disciplinary expertise for a non-specialist audience.

These roles share a structural commitment: the partner asks and the student does the intellectual work. The specific behaviors, the tone, and the disciplinary framing vary by role and by the faculty member's configuration. Faculty may rename any role for their context.

Learning by Teaching

The Peer Learner role draws from learning-by-teaching and peer-tutoring research. Roscoe and Chi's (2007) review describes how explaining and questioning can promote reflective knowledge-building through self-monitoring, integration, and elaboration. It also supplies an important caveat: tutors often fall into knowledge-telling, delivering information without reorganizing or deepening it. Later experimental work likewise found benefits from preparing to teach and actually teaching, while showing that the mechanisms and durability of those benefits depend on the activity design (Fiorella & Mayer, 2013).

The Peer Learner role addresses the knowledge-telling problem through follow-up questions. A student cannot simply recite a definition and leave. The partner asks what the definition means, how its parts relate, why it matters, and whether the student can apply it. The pedagogical value lies less in performance before an artificial audience than in the student's repeated effort to make ideas coherent.

Socratic Challenge and Argument

The Debate Coach and Skeptical Reviewer roles probe the evidence, surface counterarguments, test generalizations, and press the student to respond to the strongest objection. Faculty define the disciplinary context, the assigned materials, the task, and the standards of evidence. The partner's behavior is then bounded by those choices.

A 2025 qualitative case study illustrates both the promise and the limits of AI-supported questioning. Brazao and Tinoca studied 25 students working in five groups over a 15-week course. The authors collected 136 weekly records and reported a frequency analysis of 37 coded questions; critical-engagement questions were the most frequent category at 30%. The study had no control group or pre/post measure of critical-thinking competence. It therefore shows that advanced question types occurred in the dialogic setting, not that the AI caused a gain in critical thinking (Brazao & Tinoca, 2025). Sofiya's own educational effects must likewise be tested rather than presumed.

Assessment as a Process of Learning

Sofiya treats the conversation as more than a route to an answer. The conversation is the learning event. The student externalizes reasoning, encounters resistance, and revises in response. The transcript preserves that process as evidence faculty can inspect.

This is consistent with a broader shift from assessment of learning toward assessment for and as learning. In a systematic review of 93 empirical higher-education articles, Banihashem and colleagues (2025) found that learning analytics most often described where learners currently stood, with less attention to the formative question of how they could move forward. Sofiya's intended contribution is not another opaque prediction about a student. It is a traceable account of what the student said, where the dialogue exposed uncertainty, and what the faculty member may wish to address next.

4. Students First

Student-first design is not a claim that the learner should control every part of an assessment. It means the system is judged by whether it protects and develops the learner's agency, dignity, and capacity to think.

- **Cognitive ownership.** The student explains, argues, supplies examples, and responds to challenge. The AI is constrained from completing the task.
- **Process visibility.** The student's path through an idea matters, including uncertainty, revision, and recovery, not only the polish of a final product.
- **Formative orientation.** Student-facing feedback is intended to identify strengths, gaps, and next steps. AI does not pronounce a grade or verdict.
- **Transparent expectations.** Students know the partner's role, the assignment boundaries, what information is recorded, and how faculty will use the transcript.
- **Trust before surveillance.** Sofiya is not a proctoring system. It rejects webcam monitoring, browser lockdown, automated misconduct findings, and unreliable AI-text detection.
- **Room for difference.** Partner behavior and accessibility must be evaluated with multilingual students, disabled students, neurodivergent students, and students from varied cultural and disciplinary contexts. Fluency in a dominant academic register must not become a proxy for understanding. Where a student's explanation is substantively correct but linguistically informal, the partner is designed to engage with the substance rather than the surface register. Faculty retain final judgment on what counts as adequate demonstration of understanding in their discipline.

These commitments also shape the tone of the interaction. A debate partner can be rigorous without becoming hostile. A peer learner can admit confusion without shaming the student. Difficulty should arise from the intellectual work, not from a system performing authority.

5. Faculty First

Faculty-first design means that AI does not define the curriculum, invent the learning outcomes, or decide what counts as achievement. The faculty member orchestrates the learning environment.

- **Assignment-driven, not open-ended.** Faculty define the question, materials, scope, expected duration, and criteria for review.
- **Pedagogical roles, not generic personas.** Faculty select an interaction pattern suited to the learning goal, then frame and rename the role for the discipline and context.
- **Curated knowledge context.** Faculty choose the materials relevant to the assignment. The system does not treat the model's general training data as the course authority.
- **Evidence anchored to the transcript.** Structured analysis should identify moments worth reviewing and link each observation to what the student actually said.
- **Human judgment is final.** Faculty may accept, modify, or reject any AI-supported observation. Sofiya is not an auto-grader.
- **Pedagogical work remains portable.** The project specification supports exportable learning packs based on Common Cartridge, including assignment design, materials, role selection, rubric structure, licensing, and attribution, but excluding student records.

Sofiya may reduce the effort required to locate important moments in a long transcript, but efficiency is not the primary educational claim. The purpose is to make sustained, individualized dialogic work feasible where faculty could not personally conduct a full oral defense or studio-style critique with every student.

FACULTY AUTHORITY	*AI can organize attention. It cannot assume responsibility for a consequential educational judgment.*

6. The Sofiya Learning Cycle

The current design translates the philosophy into a bounded faculty-student workflow.

- **Design.** Faculty define the learning purpose, assignment prompt, relevant materials, interaction role, and review criteria.
- **Orient.** Students see what the role will do, what it will not do, what the task expects, and how the conversation will be used.
- **Engage.** The student explains or defends. The partner asks for clarification, evidence, examples, connections, or responses to counterarguments. It refuses to supply work the student is meant to perform.
- **Preserve.** The transcript records the unfolding reasoning. The record is not treated as a perfect window into cognition, but it provides richer process evidence than a decontextualized answer alone.
- **Review.** Sofiya prepares a structured, transcript-anchored analysis of strengths, gaps, and moments worth attention. Faculty inspect the transcript and exercise their own judgment.
- **Respond.** Faculty return formative feedback and guidance to the student. The student-facing experience emphasizes development and next steps rather than an automated score.

Different disciplines can frame the same underlying pedagogy differently. A Peer Learner might be presented as a patient in nursing, a client in architecture, or a junior colleague in software engineering. A Debate Coach might become opposing counsel in law or a skeptical grant reviewer in research methods. The Oral Examiner maps naturally to thesis defense preparation, professional licensure practice, or medical case review. The disciplinary framing belongs to faculty; the safety and refusal boundaries remain platform responsibilities.

7. Feedback That Supports Judgment

Feedback is not effective merely because it is fast or personalized. Students must interpret it, decide whether it is credible, and act. Two recent studies underscore why Sofiya's analysis should support human attention rather than bypass it.

Alsaiari and colleagues (2026) conducted a 12-week randomized study with 329 students assigned to directive, metacognitive, or hybrid AI-generated feedback. Hybrid feedback led to the highest revision rate, but confidence and final resource quality did not significantly differ across conditions. The study was conducted in one course with high baseline performance, and it did not distinguish surface edits from deeper integration. More revision cannot automatically be equated with more learning.

Pozdniakov and colleagues (2026) analyzed 7,670 instances of AI assistance used by 433 students during peer-feedback work. Seventy-nine percent of the assistance emphasized pedagogical strengths, yet only 9%, or 690 instances, were followed by revision. The study did not include a comparison of learning-resource quality with and without assistance. The result is a useful warning: well-formed AI feedback can be ignored, misunderstood, or simply unnecessary.

These findings do not undercut the value of Sofiya's analysis loop; they clarify what it is for. The goal is not to generate feedback that students act on autonomously. The goal is to give faculty a structured, transcript-anchored account they can use to inform their own feedback, identify students who need follow-

up, and make formative judgments more efficiently. The faculty member, not the system, is the feedback author.

Sofiya's approach is therefore deliberately hybrid. AI can help surface relevant moments and formulate provisional observations. Faculty bring disciplinary knowledge, knowledge of the course, ethical responsibility, and an understanding of the student and context. Student-facing feedback is published through that human judgment.

This orientation also responds to stakeholder research. Zeivots and colleagues (2026) combined interviews and focus groups with 46 participants and a survey completed by 203 people. Across students, educators, employers, accreditors, and policy actors, quality assessment required balancing feedback, integrity, student experience, authenticity, purpose, and technology. Students placed particular value on detailed feedback, while educators described the difficulty of providing it at scale. Sofiya addresses that tension without claiming that automation itself resolves it.

8. Responsible and Verifiable by Design

Grounding Helps; Verification Remains Necessary

Faculty-selected materials can reduce the distance between a conversation and the intended curriculum. Retrieval-augmented generation can help a model draw on those materials, but grounding is not a guarantee of truth. In one 2026 prototype study, hybrid retrieval reached 90% hit@5 on 70 manually written questions about a Bulgarian UML manual. In a separate manual evaluation of only 21 generated responses, 23.81% contained unsupported additions (Toskova et al., 2026). These figures are not Sofiya benchmarks. They illustrate why source grounding is useful and why the model must still not be treated as a final evaluator.

Privacy, Integrity, and Accountability

The current specification adopts data minimization, role-based access, auditability, retention controls, and separation of student records from shareable faculty learning packs. The following table summarizes the current status of these commitments.

Commitment	Status	Notes
Data minimization	Design commitment	Student transcript data collected only for the purposes of the assignment. No data used to train the underlying model.
Role-based access controls	Design commitment	Faculty see transcripts for their assignments; students see their own. Institutional admins configurable by institution.
Retention controls	Design commitment	Institutions and faculty set retention periods. Data deleted on schedule. Student records excluded from shareable learning packs.
Auditability	Design commitment	All model interactions logged with timestamps. Faculty can inspect

Commitment	Status	Notes
		the full transcript, not only the AI summary.
No automated misconduct findings	**Confirmed design constraint**	Behavioural signals are non-determinative prompts for faculty attention only. Sofiya does not infer guilt or impose penalties.
No AI-text detection	**Confirmed design constraint**	Sofiya does not use probabilistic authorship detection. Integrity evidence is the conversation itself.
Privacy impact assessment	Pending	Subject to institutional review and applicable law before deployment.

Academic integrity follows the same restraint. The strongest evidence in Sofiya is intended to be the conversation itself: whether the student can explain, connect, defend, and respond. Any behavioural signal is, at most, a non-determinative prompt for faculty attention. Sofiya does not infer guilt, impose an automated penalty, or present probabilistic authorship claims as fact.

This posture aligns with the direction of British Columbia's Draft Version 1 Model Principles and Guidelines for the Use of Artificial Intelligence, which emphasizes human-centred and pedagogy-centred use, human accountability, transparency, digital literacy, accessibility, information governance, and caution about unreliable AI-detection systems (B.C. Ministry of Post-Secondary Education and Future Skills, 2026). The provincial resource is itself a draft and voluntary guidance; alignment should not be represented as government approval or endorsement.

9. Research Evaluation

Sofiya's pilot rollout is designed to thoroughly test the mechanisms described in this paper:

- **Depth of explanation.** Do students move from definition and recall toward causal explanation, example, connection, and application?
- **Quality of argument.** Do students respond to strong counterarguments, qualify claims, and use evidence more carefully over the conversation?
- **Revision and metacognition.** Can students identify what changed in their thinking and why?
- **Faculty usefulness and trust.** Are transcript anchors accurate, relevant, and efficient to verify? When do faculty reject the analysis?
- **Equity and accessibility.** Do interaction patterns create different burdens or benefits across language background, disability, discipline, and prior knowledge?
- **Reliability and safety.** How often does the partner supply prohibited answers, introduce unsupported claims, challenge inappropriately, or fail to follow the intended role?
- **Transfer.** Does performance improve on a later task completed without the dialogic partner?

Pilot reporting should include denominators, comparison conditions where feasible, uncertainty, null findings, implementation failures, and changes made in response. Conversation completion, faculty satisfaction, and system uptime matter, but they are not substitutes for educational outcomes.

10. A Human Purpose for Educational AI

Sofiya begins with a modest proposition that leads to a substantial design choice: in education, the most valuable contribution an AI system can make may not be a better answer. It may be a better occasion for a learner to think.

That occasion must be designed. It requires a bounded role, a meaningful task, a faculty-defined context, and enough intellectual resistance that the student cannot simply pass through unchanged. It also requires restraint: the model must not be confused with a knower, the transcript must not be confused with the whole person, and analysis must not be confused with judgment.

Students and faculty come first because education is a human relationship organized around growth, knowledge, and responsibility. Sofiya's purpose is to strengthen that relationship: to give every student more opportunity to explain and defend what they know, and to give faculty a clearer, verifiable view of the thinking that occurred.

SOFIYA	*The AI asks. The student thinks. The faculty member guides and judges.*

Evidence Register

This register reports the design, sample, finding used in this white paper, and the boundary placed on the claim. It is included to make the paper's empirical statements easier to audit.

Source	Design and sample	Finding used and boundary
Li et al. (2026)	PRISMA systematic review of 67 empirical higher-education studies published from 2022 to April 2025.	Structured, inquiry-oriented uses were associated with metacognitive, argumentative, and creative gains; unstructured uses were associated with offloading and superficial engagement. The evidence base was heterogeneous and largely short-term, so the review does not establish a universal causal effect.
Alsaiari et al. (2026)	12-week randomized controlled study, 329 students in one design and programming course.	Hybrid directive-plus-metacognitive feedback prompted the most revisions. Confidence and final resource quality did not significantly differ across conditions. Revisions were not classified as surface versus deep, and transfer beyond the course was not tested.
Pozdniakov et al. (2026)	433 students; 7,670 uses of AI assistance in peer-feedback work in one undergraduate course.	Seventy-nine percent of assistance instances emphasized pedagogical strengths; only 9% (690) were followed by revision. The study did not compare learning-resource quality with and without the assistance, so uptake is not evidence of learning impact.
Brazao & Tinoca (2025)	Qualitative single-course case study: 25 students, five groups, 136 weekly records over 15 weeks; 37 questions coded in the reported frequency analysis.	Critical-engagement questions were the most frequent coded category (30%). With no control group or pre/post critical-thinking assessment, the study describes questioning patterns and does not show that AI caused improvement.

Source	Design and sample	Finding used and boundary
Banihashem et al. (2025)	Systematic review of 93 empirical higher-education articles on learning analytics and formative assessment, 2011-2023.	Most work emphasized describing where learners are now; less addressed how learners can move forward. The review was not a meta-analysis, relied on one formative-assessment model, and did not use generative AI as a search term.
Zeivots et al. (2026)	Mixed-methods study in business education: 46 interview/focus-group participants and 203 survey completions; 201 included in quantitative analysis.	Stakeholders emphasized the need to balance feedback, integrity, student experience, authenticity, purpose, and technology. The sample was uneven across stakeholder groups and concentrated in business-related fields.
Toskova et al. (2026)	Prototype RAG evaluation using 70 manually formulated UML questions; 21 responses manually evaluated.	Hybrid retrieval reached 90% hit@5, but 23.81% of the small response sample contained unsupported additions. These are prototype results, not Sofiya results; they show why grounding can help but cannot replace human verification.

References

Alsaiari, O., Baghaei, N., Lodge, J. M., Noroozi, O., Gasevic, D., Boden, M., & Khosravi, H. (2026). Directive, metacognitive, or a blend of both? A comparison of AI-generated feedback types on student engagement, confidence, and outcomes. Computers and Education: Artificial Intelligence, 10, 100553. https://doi.org/10.1016/j.caeai.2026.100553

Banihashem, S. K., Gasevic, D., Noroozi, O., Jarodzka, H., Joosten-ten Brinke, D., & Drachsler, H. (2025). Optimizing formative assessment with learning analytics. Review of Educational Research. Advance online publication. https://doi.org/10.3102/00346543251370753

B.C. Ministry of Post-Secondary Education and Future Skills. (2026). Draft Version 1 - Model principles and guidelines for the use of artificial intelligence: An informational resource for the B.C. post-secondary system. https://www2.gov.bc.ca/assets/gov/education/post-secondary-education/institution-resources-administration/digital-learning-strategy/draft_v1_-_model_principles_and_guidelines_for_the_use_of_ai.pdf

Brazao, P., & Tinoca, L. (2025). Artificial intelligence and critical thinking: A case study with educational chatbots. Frontiers in Education, 10, 1630493. https://doi.org/10.3389/feduc.2025.1630493

Corbin, T., Walton, J., Bannister, P., & Deranty, J.-P. (2026). On the essay in a time of GenAI. Educational Philosophy and Theory, 58(3), 198-210. https://doi.org/10.1080/00131857.2025.2572802

Fiorella, L., & Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. Contemporary Educational Psychology, 38(4), 281-288. https://doi.org/10.1016/j.cedpsych.2013.06.001

Li, C., Cui, H., & Hagedorn, L. S. (2026). The cognitive impact of ChatGPT in higher education: A systematic review of critical and creative thinking outcomes. Computers and Education: Artificial Intelligence, 10, 100571. https://doi.org/10.1016/j.caeai.2026.100571

Pozdniakov, S., Brazil, J., Banihashem, S. K., Noroozi, O., Gasevic, D., Sadiq, S., & Khosravi, H. (2026). AI assistance in peer feedback provision: Pedagogically sound, but minimally adopted. Computers & Education, 248, 105591. https://doi.org/10.1016/j.compedu.2026.105591

Quattrociocchi, W., Capraro, V., & Perc, M. (2025). Epistemological fault lines between human and artificial intelligence [Preprint]. https://doi.org/10.48550/arXiv.2512.19466

Roscoe, R. D., & Chi, M. T. H. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. Review of Educational Research, 77(4), 534-574. https://doi.org/10.3102/0034654307309920

Toskova, A., Georgiev, K., & Glushkova, T. (2026). Dialogical learning support in RAG-based e-learning. Information, 17, 418. https://doi.org/10.3390/info17050418

Wegerif, R., & Casebourne, I. (2026). A dialogic theoretical foundation for integrating generative AI into pedagogical design. British Journal of Educational Technology, 57, 639-654. https://doi.org/10.1111/bjet.70026

Zeivots, S., Wright, S., Harris, L., Cram, A., Huber, E., Raduescu, C., White, A., & Brodzeli, A. (2026). Walking the tightrope of quality assessment: Balancing perspectives and priorities of stakeholder groups. Studies in Higher Education. Advance online publication. https://doi.org/10.1080/03075079.2026.2619909