Dialogic AI Scaffolding: A Proof-of-Concept Protocol to Foster Critical Thinking and Metacognition

Marta Valentini[1,*], Alberto Montresor[1]

[1]Department of Computer Science and Engineering, University of Trento, Italy

Abstract

This paper presents a pedagogically grounded protocol for integrating generative AI into educational settings as a dialogic partner, rather than a source of ready-made answers. Drawing on Bruner's theory of education and dialogic instructional models, the protocol leverages Socratic questioning and "Devil's Advocate" strategies to promote metacognition, critical thinking, and learner awareness. A four-phase protocol guides students from task comprehension to reflective self-assessment through structured AI interaction. A pilot implementation in a university-level programming course showed that students perceived the protocol as helpful in supporting learning, problem-solving, and self-awareness, despite the higher cognitive demands. These findings suggest that dialogically structured AI interactions can foster slow, reflective learning and support the development of learning-to-learn skills.

Keywords

AI in education, dialogic inquiry, metacognition, critical thinking

1. Introduction

The emergence of generative artificial intelligence (AI) poses significant pedagogical questions. While AI holds considerable promise for enhancing learning through personalization, immediate feedback, and adaptive support [1], its uncritical use risks reducing learning to the passive consumption of answers. In such cases, students are deprived of opportunities to actively construct knowledge and engage in reflective thinking [2]. The pedagogical challenge, therefore, lies not merely in adopting AI tools, but in designing learning experiences where AI enhances learning itself, contributing meaningfully to learners' cognitive and metacognitive development. In educational settings, generative AI must be deployed differently than in task-oriented or productivity contexts. Rather than delivering pre-packaged solutions, it should provide structured scaffolding that supports inquiry, invites interpretation, and sustains reflection. Particularly during the learning phase, it is essential to adopt methods that promote deeper engagement with knowledge and foster learning-to-learn capacities.

This study takes as its starting point the premise that well-established pedagogical methods for fostering metacognition and critical thinking can be enacted through an AI system that operationalizes their principles in a structured, scalable way.

As a concrete implementation of this approach, we propose a protocol in which AI is integrated into classroom dialogue as a cognitive partner–enhancing, rather than replacing, pedagogical interaction. Our approach is grounded in Bruner's theory of education as a cultural and dialogic process [2], in which learners construct understanding through structured exchanges, guided discovery, and recursive reflection. Within this view, dialogue is not merely a vehicle for cognitive probing; it is a site of learning in itself–a space where knowledge is co-constructed and understanding emerges through interaction.

Building on this foundation, we develop a protocol of *AI-mediated dialogic inquiry*, informed by the instructional principles articulated by Collins and Stevens [3].

The novelty of our contribution lies in the formalization of this protocol: a principled integration of dialogic educational roles into AI behavior, with the goal of fostering metacognition, critical thinking,

2nd Workshop on Education for Artificial Intelligence (edu4AI 2025, https://edu4ai.di.unito.it/), Co-located with ECAI 2025, the 28th European Conference on Artificial Intelligence which will take place on October 26, 2025 in Bologna, Italy

*Corresponding author.
marta.valentini-1@unitn.it (M. Valentini); alberto.montresor@unitn.it (A. Montresor)
0009-0006-7310-1891 (M. Valentini); 0000-0001-5820-8216 (A. Montresor)

and learner self-awareness. Beyond task support, it helps students learn how to use AI in reusable, controllable ways, via concrete and repeatable routines. Our central research questions are:

- *How can a protocol based on AI-mediated dialogic inquiry be designed to foster learning?*
- *How is the proposed protocol perceived by university students in terms of supporting metacognitive awareness and critical thinking?*

To explore these questions, we design an AI peer agent that alternates between two dialogic functions:

- *Socratic questioning*: a disciplined sequence of open-ended questions designed to help learners surface assumptions, differentiate what is known from what is not, and critically examine their reasoning [4].
- *Devil's Advocate interventions*: counter-arguments and alternative perspectives intentionally introduced to challenge learners' positions and strengthen their argumentative resilience [5].

This paper presents a first proof-of-concept implementation of the protocol in a university-level programming course. Qualitative exploratory data were collected from pre- and post-intervention questionnaires and AI–student interaction logs. These findings will inform a subsequent implementation in an introductory course, with more systematic data collection and evaluation.

2. State of the art

Recent research has raised concerns about the use of AI in education when it replaces, rather than supports, student learning [6, 7]. In such cases, AI risks promoting shallow engagement, hindering the development of core skills [8] and depriving learners of opportunities to construct knowledge through inquiry, exploration, and reflection. At the same time, a growing body of work highlights the potential of AI to enhance learning when embedded within thoughtfully designed instructional frameworks. In these contexts, AI has been shown to foster deeper understanding, critical thinking, and metacognitive awareness [9]. Holmes et al. [10] have introduced the notion of *AI-augmented learning*, in which AI strengthens, rather than supplants, the teacher's role, acting as a catalyst for active learning.

We situate our approach within the tradition of *dialogic inquiry* [11], which views learning not as the passive reception of information, but as a process of meaning-making through structured dialogue. In this perspective, informed by sociocultural and constructivist theories of learning, dialogue is not merely a tool for instructional delivery but a site where knowledge is co-constructed, negotiated, and transformed [12]. Accordingly, the goal is not to have AI transmit content, but to support reflective conversations that help students articulate, challenge, and refine their understanding, while also developing practical habits for using AI in reusable, student-controlled ways.

To this end, our protocol integrates two complementary dialogic strategies: *Socratic questioning* and *Devil's Advocate interventions.* These are well-established pedagogical techniques for fostering critical thinking and self-awareness [13]. Socratic questioning involves a disciplined sequence of probing, open-ended questions that encourage learners to clarify assumptions and evaluate their reasoning. Devil's Advocate interventions, by contrast, challenge students' initial positions through counter-arguments or alternative perspectives, stimulating epistemic conflict and argument elaboration [5].

Recent empirical work suggests that AI systems embodying these roles can support deeper engagement than traditional, answer-oriented models. For example, Favero et al. [4] demonstrate that a Socratic chatbot can significantly enhance students' critical reflection compared to standard conversational agents; similar results are reported in [14].

An important feature of our approach is its attention to the regulation of cognitive load and pacing. Learning is not instantaneous: it requires time for uncertainty, error, and revision. Our protocol accommodates this by activating the AI only after the learner has engaged autonomously with the task, thus supporting slower, more deliberate thinking processes [15].

In doing so, our work contributes to emerging efforts aimed at formalizing pedagogically grounded interactions with AI to foster metacognitive awareness and critical thinking [16, 17]. We align with a broader pedagogical stance that views AI not as a replacement for human instruction, but as a means of deepening and extending established educational practices [18].

3. Protocol

The protocol, structured into four distinct phases, guides students from their initial engagement with the task through to metacognitive reflection (and, where applicable, collective sharing). Its primary objective is to foster students' ability to formulate high-quality questions, engage in critical reasoning, develop metacognitive awareness, and build autonomy in approaching new problems–through a structured interaction with an AI agent. Figure 1 summarizes the dialogic protocol we used: students first attempt the task independently and document their reasoning, then select one of two guided AI interventions (Socratic or Devil's Advocate), and finally complete a critical review form (GF) before submitting. This visual provides the end-to-end flow we analyze in the following section. The protocol can be applied individually or in small groups.

Phase	Core Student Activity	Illustrative interaction
0 – Task Assignment	Review the GF task brief, identifying how the project must be extended through structural refactoring.	Task excerpt: reorganize the package hierarchy; introduce an interface that exposes the core methods; rename the legacy class to remove ambiguity; move shared logic into an abstract base class; and implement a specialised concrete subclass.
1 – Question Generation & Refinement	Generate an initial question set, iterate with AI-driven refinements, and pick the best using the mini-rubric.	*S → AI: "What is an abstract class? How can I refine this question or what else should I ask?"* *AI → S: "Try these prompts: How does an abstract class differ from an interface? When is an abstract class preferable to interfaces alone?"*
2 – Response Drafting & AI Feedback	Develop a preliminary solution and ask the AI for Socratic and/or Devil's Advocate questions.	S → AI: *"Here are my answers. Play devil's advocate—challenge my assumptions and surface any weaknesses."* AI → S: *"You say an interface is effectively an abstract class, but that isn't accurate. Can you explain the difference and typical use cases for each?"*
3 – Validation & Metacognitive Reflection	Answer guiding questions, engage in metacognitive reflection, and label good vs. poor prompt examples.	GF: *"After the AI session, did you revise your answers? Did you ask yourself: What visibility should the fields have? And the methods? How can you improve your interaction with the AI? Review good vs. poor prompts to improve future interactions."*

Figure 1: Example application of the protocol to a CS2 refactoring task (GF = Google Form; S = Student, AI = ChatGPT 4o). Quoted lines are excerpts from the AI–student interaction log.

Phase 0: Task Assignment: The instructor introduces the task–for instance, the implementation of a code component–along with general guidelines and task-specific instructions. During this initial phase, students become familiar with the objectives of the activity and with the operational format they will follow in the subsequent phases.
Phase 1: Question Generation and Refinement: Students independently generate an initial set of questions, using metacognitive transfer techniques if needed to stimulate divergent thinking. They then submit these questions to the AI with prompts such as "What other questions should I consider?" or "How can I improve my questions?". From the AI's suggestions, students select the most relevant questions and refine them using a rubric designed to assess their relevance and depth.
Phase 2: Task Execution and AI Feedback: Students respond to the final set of questions and work independently to complete the task. At this point, the AI intervenes in one or both of the following ways: using a Socratic approach, it poses guiding questions to support the logical progression of student reasoning; using a Devil's Advocate approach, it challenges assumptions through provocative questions designed to expose weaknesses and stimulate further reflection. Students deliberately choose one of these guided AI interventions; the choice is not randomized but aligned with their perceived need. This fosters a metacognitive process that strengthens students' self-awareness and their awareness of AI use.
Phase 3: Validation and Metacognitive Reflection: A structured Google Form presents guiding

questions focusing on both the content of the task and the interaction with the AI, that encourage students to reflect on the overall effectiveness of the process and to identify and correct any conceptual misunderstandings. Finally, they select examples of "good" and "poor" prompts and justify their classification from a metacognitive perspective in a shared document. Instructors did not pre-validate AI turns; chat logs were monitored only to verify protocol flow and extract illustrative snippets.

4. Pilot implementation of the protocol

Context and setting *Programmazione 2* (CS2) is an advanced course in object-oriented programming in Java, delivered during the 2nd semester of the 1st year of the Computer Engineering degree at the University of Trento. Alongside lectures, students attend in-person lab sessions. The protocol was implemented during lab session 2, involving approximately 86 students organized into 29 groups of 2–3 members each. Two additional sessions (lab sessions 6 and 7) applied the same protocol, but involved only a limited number of students and did not yield analyzable data; hence, they are not discussed here.

Learning tools Each student group was provided with: (i) a dedicated instance of ChatGPT-4o, supplied by the university and configured to centrally record interaction logs—this traceability will allow for a later analysis of student–AI interaction and identification of pedagogical refinements; (ii) a structured Google Form (GF) guiding students through the activity step by step: each phase of the protocol (question generation, AI feedback, metacognitive reflection, etc.) included instructions and optional fields to record questions, answers, and observations; (iii) a short evaluation rubric used to assess the quality of the questions (both student-generated and AI-suggested) and to support decisions about which to explore or discard; (iv) a collaborative document in which students recorded examples of "good" and "poor" prompts, along with a metacognitive justification of their choices.

Procedure After a brief theoretical introduction emphasizing the value of reflective–rather than passive–use of AI, students were presented with the four-phase protocol and the supporting digital tools. They then began working on the first of seven tasks comprising the lab activity, managing their time autonomously to complete the remaining ones. Figure 1 illustrates the application of the protocol to Task 1 (Refactoring of the `Block` class hierarchy) as an example. After completing the first task, students applied the same sequence to the six remaining tasks.

5. Data Collection and Descriptive Results

In line with our research questions on fostering deeper learning and supporting metacognitive awareness and critical thinking, the data analysis conducted in this study was descriptive and exploratory, consistent with the proof-of-concept nature of our proposal.

Data were collected from two anonymous questionnaires administered before and after the lab activity. The pre-questionnaire, completed by 86 students at the beginning of the course, focused on students' previous experience with AI tools, their beliefs about AI's usefulness, and some metacognitive aspects. The post-questionnaire focused on students' subjective evaluation of the protocol and the learning experience. Chat logs were not used for data analysis but only to monitor the flow of interaction and extract illustrative examples (see Figure 1). The post-questionnaire was administered after the final session, near the end of the course, when participation was limited. Only 17 students completed the form. As items were context-specific and self-reported, and the post sample was small, results are exploratory and should be interpreted with caution.

Although not designed for direct comparison with the post-questionnaire, the pre-questionnaire provided useful context. About 71% (61/86) of students reported daily use of AI tools, and 86% (74/86) rated positively their familiarity[1]. AI was considered useful primarily for identifying bugs in code (77%, 66/86) and generating test cases (77%, 66/86). In contrast, its usefulness for *learning* to debug or test was rated lower (56%, 48/86 and 52%, 45/86 respectively). Open-ended responses mainly framed AI as a

[1]On a 5-point scale, we considered score 4 and 5 as positive

practical support tool, with few references to metacognitive or reflective uses. This context informed the interpretation of the post- data, highlighting how the protocol may have expanded students' perception of AI toward reflective use.

Regarding the post-questionnaire, the data suggest a positive evaluation of the protocol. Students found the strategies helpful for improving their awareness and problem-solving skills. Specifically, 53% (9/17) and 59% (10/17) rated the Socratic questioning phase and the Devil's Advocate phase positively, respectively; 47% (8/17) appreciated the metacognitive transfer phase, and 47% (8/17) valued the final review phase. These results indicate a perceived added value from structured AI interaction in terms of clarity, awareness, and reasoning stimulation.

Regarding perceived learning outcomes, 65% (11/17) of students agreed that the protocol helped them identify errors, 65% (11/17) reported it improved their testing abilities, and 82% (14/17) said it helped them better understand core Java concepts. Despite the increased cognitive effort required, students rated the learning benefits as worthwhile. Open-ended feedback acknowledged the length of the activity but emphasized its usefulness in supporting deeper conceptual understanding.

On critical thinking, 53% (9/17) of students agreed that they critically evaluated the AI's questions and reflected more deeply on the task because the AI offered questions instead of direct answers.

In terms of metacognition, 65% (11/17) reported that the protocol helped them reflect on their own learning process, and 59% (10/17) stated that the Socratic and Devil's Advocate phases helped them question their assumptions and hypotheses. These findings suggest that guided AI interaction, when structured through an appropriate protocol, can promote reflection on learning processes–supporting not only content acquisition but also the development of learning-to-learn strategies.

6. Conclusions and Future Work

This paper introduced a four-phase protocol that reframes AI from an answer provider to a thinking companion capable of enacting established pedagogical methods. We examined how such a protocol, grounded in dialogic interaction, can foster students' critical thinking and metacognitive awareness.

In this preliminary implementation, the protocol appeared to prompt students to reason critically about both the task at hand and the AI's responses. The analysis of the post-questionnaire answers suggests that the protocol was perceived as a meaningful support for learning.

The protocol promotes learning by requiring students to generate and refine their own questions (activation), engage with targeted challenges that induce cognitive conflict, and reflect explicitly on their learning process (metacognition). It fostered deliberate, slow thinking, which students valued despite the additional time required. The four phases of the protocol structure cognitive engagement and transforms the AI into a dialogic scaffold. As highlighted in the literature, passive use of AI can reduce students' cognitive effort but leads to impoverished learning; in contrast, structured AI interaction requires greater investment but promotes growth and metacognitive awareness.

However, several limitations must be acknowledged. The study was conducted in a single disciplinary context, involved a small sample limiting the generalizability of findings. Moreover, the analysis relied partly on self-reported data and did not include objective measures of critical thinking and metacognition.

These limitations guide our future research agenda. A follow-up study in an introductory programming course will implement the protocol multiple times and incorporate validated rubrics to assess metacognitive awareness and critical thinking. Interaction logs will be analyzed through both qualitative and quantitative methods, focusing on the evolution of dialogic practices over time.

Future work will also involve designing and testing additional interaction protocols grounded in other well-established pedagogical frameworks, to explore how effectively they can be amplified through AI support. These trials will include applications in non-STEM disciplines, examining how different subject domains may shape the suitability and impact of protocol-mediated AI scaffolding. In parallel, we plan to extend the system architecture to build a community of AI agents capable of interacting with each other and with the learner.

The planned extensions converge on a single overarching aim at the heart of education: fostering students' active construction of knowledge and metacognitive awareness. Structured dialogue with AI, governed by pedagogically grounded protocols, can provide a practical path toward achieving this goal.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the author reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] S. S. Lee, R. L. Moore, Harnessing generative AI (GenAI) for automated feedback in higher education: A systematic review, Online Learning Journal 28 (2024) 82–104.

[2] J. Bruner, The Culture of Education, Harvard University Press, Cambridge, MA, 1996.

[3] A. M. Collins, A. L. Stevens, A cognitive theory of interactive teaching, in: C. M. Reigeluth (Ed.), Instructional-Design Theories and Models: An Overview of Their Current Status, Lawrence Erlbaum Associates, Hillsdale, NJ, 1983, pp. 247–278.

[4] L. Favero, J. A. Pérez-Ortiz, T. Käser, N. Oliver, Enhancing critical thinking in education by means of a socratic chatbot, in: Proc. of the International Workshop on AI in Education and Educational Research (AIEER), Santiago de Compostela, Spain, 2024.

[5] C. MacDougall, F. Baum, The devil's advocate: A strategy to avoid groupthink and stimulate discussion in focus groups, Qualitative Health Research 7 (1997) 532–541.

[6] J. O. Wecks, J. Voshaar, B. J. Plate, J. Zimmermann, Generative AI usage and exam performance, arXiv preprint arXiv:2404.19699 (2024).

[7] M. Lehmann, P. B. Cornelius, F. J. Sting, AI meets the classroom: When do large language models harm learning?, arXiv preprint arXiv:2409.09047 (2025).

[8] B. N. Macnamara, I. Berber, M. C. Çavuşoğlu, E. A. Krupinski, N. Nallapareddy, N. E. Nelson, P. J. Smith, A. L. Wilson-Delfosse, S. Ray, Does using artificial intelligence assistance accelerate skill decay and hinder skill development without performers' awareness?, Cognitive Research: Principles and Implications 9 (2024) 46.

[9] S. Chardonnens, Adapting educational practices for generation Z: integrating metacognitive strategies and artificial intelligence, Frontiers in Education 10 (2025).

[10] W. Holmes, M. Bialik, C. Fadel, Artificial Intelligence in Education: Promises and Implications for Teaching and Learning, 1st ed., Center for Curriculum Redesign, Boston, MA, USA, 2019.

[11] G. Wells, Dialogic Inquiry: Towards a Sociocultural Practice and Theory of Education, Cambridge University Press, Cambridge, 1999.

[12] R. Alexander, Towards Dialogic Teaching: Rethinking Classroom Talk, 4th ed., Dialogos, 2008.

[13] N. Le, How do technology-enhanced learning tools support critical thinking?, Frontiers in Education 4 (2019) 126.

[14] D. E. Pitorini, Suciati, Harlita, Students' critical thinking skills using an e-module based on problem-based learning combined with socratic dialogue, Journal of Learning for Development 11 (2024) 52–65.

[15] H. Catalano, I. Albulescu, C. Stan, G. Mestic, A. Ani-Rus, Child-centered approach through slow education principles: A view to child personality development in early childhood, Sustainability 15 (2023) 8611.

[16] B. Yuan, J. Hu, Generative AI as a tool for enhancing reflective learning in students, arXiv preprint arXiv:2412.02603 (2024).

[17] L. Tankelevitch, V. Kewenig, A. Simkute, A. E. Scott, A. Sarkar, A. Sellen, S. Rintel, The metacognitive demands and opportunities of generative AI, in: Proc. of the 2024 CHI Conference (CHI '24), ACM, New York, NY, 2024, pp. 1–24.

[18] H. Fakour, M. Imani, Socratic wisdom in the age of AI: a comparative study of ChatGPT and human tutors in enhancing critical thinking skills, Frontiers in Education 10 (2025) 1528603.