Toward Critical AI Literacy: Embedded Assessment of Students' Critical Engagement with LLM Responses

Minju Yoo[1,†], Junghee Hwang[2,†]

[1]KAIST, Daejeon, Korea
[2]Seoul Guro Elementary School, Seoul, Korea

Abstract

LLM-based chatbots are increasingly used in K-12 classrooms as information sources, exposing students to risks of over-reliance, hallucinations, and bias and making critical evaluation of AI responses central to AI literacy education. Yet how to assess this skill remains unclear: existing frameworks stay conceptual, instruments rely on self-report or standalone evaluation tasks, and the skill develops through practice that single-lesson assessments miss. In this position paper, we argue that teaching critical AI literacy calls for *embedded assessment*: a chatbot-based learning system in which the same interaction scaffolds critical evaluation and generates evidence for it. As initial steps, we propose a rubric paired with scaffolding strategies and outline three design considerations for the system.

Keywords

Critical AI usage, AI literacy assessment, Embedded assessment, LLM-based chatbots in education

1. Introduction

Large language model(LLM)-based chatbots are increasingly used by K-12 students as information sources for background research, concept explanations, and homework Q&A [1]. They also expose students to risks specific to generative AI, such as over-reliance on unverified outputs, hallucinated facts, and subtle biases [2]. This makes critical evaluation of such outputs a central concern of AI literacy education.

However, it is unclear how to assess such critical engagement in practice. Existing AI literacy frameworks [3, 2] treat critical evaluation as a core competency but stay at the level of conceptual definitions, leaving open how it is embodied in observable student behavior. Current instruments rely either on self-report scales or on standalone tasks that ask students to evaluate pre-selected AI outputs, but these rarely capture how students engage with AI responses [4]. Moreover, critical evaluation develops through repeated practice rather than a single lesson [5], yet manual review of every interaction at the classroom scale is infeasible.

We argue that teaching critical AI literacy calls for **embedded assessment**, in which critical questions integrated into chatbot responses serve simultaneously as scaffolds for reflection and as a record of critical engagement. Realizing this vision points toward a learning system built around chatbot interaction that both teachers and students use in everyday classrooms. As initial steps, we propose a behavioral rubric that organizes critical AI literacy into three dimensions, each paired with scaffolding strategies. We then outline an assessment loop organized around three design considerations: teacher configurability, student agency, and longitudinal monitoring.

2. A Rubric for Embedded Assessment of Critical AI Literacy

We use 'critical AI literacy' to refer to the observable behaviors through which students question, verify, and reflect on chatbot responses rather than accepting them at face value. Drawing on existing AI

ALIT4ALL: 2nd International Workshop on AI Literacy Education For All, In conjunction with the Festival of Learning 2026 @ AIED, Seoul, Korea

[†]These authors contributed equally.

minjuyoo@kaist.ac.kr (M. Yoo); teacherjunghee@gmail.com (J. Hwang)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Table 1
Six components of critical AI literacy with behavioral rubric and scaffolding strategies.

Dimension	Component	Behavioral Rubric	Scaffolding strategy
Prompt	Intent Clarity	**[Low]** Short, vague queries **[High]** Names target, purpose, and context	Specify the purpose, target, and context
	Query Revision	**[Low]** Accepts off-target answers **[High]** Revises wording when the answer misses intent	Notice intent-response gaps and revise the question
Response	Factual Accuracy	**[Low]** Accepts AI answers as given **[High]** Cross-checks key information against external sources	Cross-verify key facts (numbers, names, dates) with external sources
	Logical Consistency	**[Low]** Misses contradictions when text reads smoothly **[High]** Detects contradictions or mismatched cause-effect	Check consistency between earlier and later parts
One's own thinking	Perspective awareness	**[Low]** Accepts a single stance **[High]** Asks questions probing opposing perspectives	Surface perspectives missing from the answer
	Process reflection	**[Low]** Does not review own process after inquiry **[High]** Reflects on how prompts shaped the answer	Reflect on how question framing shaped the answer

literacy frameworks [3, 2], we organize these behaviors across three dimensions: one's own question, the response itself, and one's own thinking. Within each, we identify two behavioral components drawn from the critical thinking [6, 7] and student questioning [8] literature, each paired with a behavioral rubric and scaffolding strategy (Table 1).

For example, a student asks a chatbot for a country's population and receives an outdated figure. Under the Factual Accuracy component, the chatbot surfaces a scaffolding question prompting the student to cross-check the figure against an external source. If the student dismisses the prompt, the exchange is logged as low engagement; if they check another source and revise their understanding, as high engagement. Either way, the same interaction serves as a scaffold in the moment and, once tagged, as evidence for future assessment.

3. Toward an Embedded Assessment System

3.1. Design Considerations

The three considerations form an assessment loop: teachers configure the rubric and scaffolding, students choose which critical questions to reflect on, and their logged interactions are surfaced back to both, informing teachers' iteration and students' self-regulation.

Teacher Configurability. An embedded assessment system should support teachers in adapting the chatbot's critical questions and scaffolding to their classroom context [9]. Teachers customize the critical questions for each rubric component and adjust the scaffolding intensity (e.g., no hint, leading question, or explicit feedback) to match their students' grade level and experience, so the chatbot enacts their pedagogical decisions at scale.

Student Agency. Students should be able to choose which aspects of a chatbot response to engage critically with, rather than being marched through a fixed checklist [10]. The system surfaces a small set of critical questions per response, letting students decide which to reflect on. Choice supports motivation and lets the system tailor scaffolding to each student's chosen focus.

Longitudinal Monitoring. Tagging chat logs and student responses by rubric component makes patterns of critical engagement visible over time: teachers see class-level trends for their next configuration, and students see their own coverage as a basis for self-regulation [11]. We emphasize progress visualization over scoring, since reducing critical thinking to a single metric would be self-defeating.

3.2. Future Work

Our framework is not yet empirically validated; our next step is to implement a working prototype with teacher-facing configuration tools and student-facing reflection and progress views. We then plan formative testing with teachers and students to refine the rubric and scaffolding wording, followed by deployment in K-12 classrooms to collect longitudinal interaction data, with appropriate consent procedures and data protection measures suited to minors.

4. Declaration on Generative AI

During the preparation of this work, the author(s) used Claude, Grammarly in order to identify and correct grammatical errors, typos, and other writing mistakes.

References

[1] M. Madden, A. Calvin, A. Hasse, A. Lenhart, Teen and Young Adult Perspectives on Generative AI: Patterns of Use, Excitements, and Concerns, Technical Report, Common Sense Media, Hopelab, and Center for Digital Thriving at Harvard Graduate School of Education, San Francisco, CA, 2024.

[2] X. Gu, B. J. Ericson, Ai literacy in k-12 and higher education in the wake of generative ai: An integrative review, in: Proceedings of the 2025 ACM Conference on International Computing Education Research V. 1, 2025.

[3] OECD, Empowering Learners for the Age of AI: An AI Literacy Framework for Primary and Secondary Education (Review Draft), Technical Report, Organization for Economic Co-operation and Development, Paris, 2025.

[4] T. Lintner, A systematic review of ai literacy scales, npj Science of Learning 9 (2024).

[5] P. C. Abrami, R. M. Bernard, E. Borokhovski, D. I. Waddington, C. A. Wade, T. Persson, Strategies for teaching students to think critically: A meta-analysis, Review of educational research 85 (2015).

[6] P. Facione, Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (the delphi report) (1990).

[7] L. Elder, R. Paul, Critical thinking: Strategies for improving student learning, Journal of Developmental Education 32 (2008).

[8] C. Chin, J. Osborne, Students' questions: a potential resource for teaching and learning science, Studies in science education 44 (2008).

[9] I. Hou, Z. Xiong, P. J. Guo, A. Y. Wang, "bespoke bots": Diverse instructor needs for customizing generative ai classroom chatbots, in: Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems, 2026.

[10] R. M. Ryan, E. L. Deci, Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being., American psychologist 55 (2000).

[11] W. Matcha, N. A. Uzir, D. Gašević, A. Pardo, A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective, IEEE transactions on learning technologies 13 (2019).