

Document, Verify, Explain: A Transparent Accountability Framework for Equitable Generative AI Use in Computer Science Education

Angel Rivera and Unnati Shah

Department of Computer Science
Utica University
Utica, NY

Abstract

The rapid adoption of generative AI tools in Computer Science (CS) education has created a tension between their potential to support learning and growing concerns about academic integrity, equity, fairness, and erosion of core skills. Attempts to prohibit or police AI use have proven difficult to enforce, inconsistently applied, and costly in instructional effort, often amplifying inequities arising from unequal prior experience, access, or confidence in AI tools. The main objective of this paper is to present a transparent AI accountability framework that integrates generative AI into CS courses in a structured, auditable, and equitable manner, enabling consistent assessment while promoting responsible use. The framework is built on three principles: explicit expectations for AI use, structured documentation and reflection, and mechanisms for student accountability. This paper presents the framework's design and reports on its initial feasibility through pilot applications in an introductory programming course and an upper-level CS course. While exploratory in nature, these case studies demonstrate how the framework structures AI-supported problem solving with required logging, verification, and oral explanation, scaffolding responsible AI use for students with diverse preparation levels in the introductory course. In the upper-level course, it supported AI-assisted design, testing, visualization, and formal verification of complex systems. Across both cases, students demonstrated stronger engagement with reasoning, validation, and explanation, while faculty experienced reduced enforcement burden. The results provide a foundational proof-of-concept for scalable, transparent AI integration in CS curricula, offering a structured alternative to detection-based approaches.

Introduction

Generative Artificial Intelligence (AI) tools, particularly large language models (LLMs) such as ChatGPT, GPT-4, Claude, and LLaMA, are rapidly transforming Computer Science (CS) education. These tools offer unprecedented support for coding, debugging, and problem solving, but their growing presence also challenges traditional instructional practices, assessment methods, and learning dynam-

ics. Left unstructured, AI use can obscure student understanding, complicate evaluation, and introduce inequities, making it difficult for instructors to ensure that learning outcomes are met fairly and effectively. Without careful integration, classrooms risk a situation where AI use is uneven, opaque, and potentially harmful to both learning and fairness.

Why Generative AI Is a Challenge in Practice

The increasing prevalence of generative AI in higher education has substantially reshaped teaching and learning practices, particularly in CS. AI tools offer clear benefits, including conceptual support, accelerated problem solving, and assistance with testing and debugging. However, research consistently shows that unstructured AI use undermines learning quality, transparency, and academic integrity. Students frequently over-rely on LLMs, accept incorrect or hallucinated outputs, work backward from complete solutions, and engage in less planning, verification, and conceptual reasoning (Prather et al. 2024; Joshi et al. 2024; Hak et al. 2025; Zviel-Girshin 2024). Even when short-term productivity improves, these gains often come at the cost of conceptual understanding and reduced engagement with legitimate learning supports (Andleeb, Kantorski, and Carver 2025; Ramirez Osorio et al. 2025).

Recent studies converge on a clear insight: the problem is not the presence of LLMs, but the lack of instructional structure, transparency, and accountability. Park and Ahn (2024) show that students both value and distrust ChatGPT due to hallucinations, opacity, and perceived learning erosion, highlighting the absence of pedagogical scaffolding typically provided by intelligent tutoring systems. Intervention studies suggest that when students are required to document, challenge, verify, and reflect on AI-generated output, learning outcomes improve particularly for complex, higher-order tasks (Farinetti and Cagliero 2025; Jamie, HajiHashemi, and Alipour 2025).

Equity and Hidden Advantages

Beyond learning quality, unregulated AI use raises equity concerns. Students enter courses with varying levels of preparation, access to technology, and prior exposure to AI tools. Without guidance, students with stronger backgrounds or informal AI experience gain hidden advantages, while

others may avoid AI due to fear of policy violations or rely on it uncritically. This can widen learning gaps. These inequities are often invisible to instructors. Explicit structures for documenting and evaluating AI-assisted work are necessary to ensure that differences in performance reflect true understanding rather than unequal access or skill. Achieving equitable AI integration requires shared expectations, transparency, and consistent standards across coursework.

Gaps in Current Research and Practice

Current approaches to AI in CS courses typically fall into two categories: unrestricted access (Prather et al. 2024; Joshi et al. 2024; Hak et al. 2025; Zviel-Girshin 2024; Ramirez Osorio et al. 2025) and ad hoc restrictions (Park and Ahn 2024; Joshi et al. 2024; Zviel-Girshin 2024). Both approaches are limited:

- **Unrestricted access** leaves AI use opaque and unassessed, potentially harming learning and fairness.
- **Ad hoc restrictions** require significant enforcement effort and may disadvantage some students.

There is a clear gap in course-level frameworks that make AI use transparent, auditable, pedagogically guided, and instructionally accountable. Addressing this gap is essential for enabling responsible and equitable AI use while supporting deeper learning outcomes.

Proposed Transparent AI Accountability Framework

The primary objective of the proposed framework is to enhance learning outcomes in CS while maintaining fairness and ethical responsibility. The framework establishes three key components:

1. Clear expectations for AI-assisted work.
2. Structured reflection exercises.
3. Mechanisms for student self-assessment and peer accountability.

At the outset, faculty define explicit expectations for AI use, including permitted tools, scope of assistance, required validation, and assessment criteria. These expectations are communicated through standardized rubrics, documentation templates, and accountability artifacts, providing consistent structure across students and assignments. Reflection and accountability are critical evaluation layers; instructors assess AI use statements, ethical checklists, and responsibility attestations to evaluate reasoning, validation strategies, and ethical decision-making. Self-assessments and peer accountability, in the form of documentation logs, attribution traces, and peer reviews, provide traceability and ownership of the final outputs.

To improve clarity and reproducibility, instructors implement the framework through a structured workflow: providing rubrics and reflection templates, scheduling baseline and AI-assisted tasks, reviewing student documentation, and conducting guided check-ins to ensure compliance and consistency. This approach allows replication across courses while maintaining oversight standards.

Assignments under this framework progress in four phases. A non-AI baseline activity allows instructors to verify foundational understanding before AI-supported work begins, reducing hidden advantages and supporting equitable participation. Then, the execution of the AI-assisted task. During this phase, faculty guide learning through structured constraints rather than surveillance. Students document AI interactions, specifying accepted, modified, or rejected outputs and providing validation evidence such as test cases or formal verification results. Faculty evaluation focuses on completeness, correctness, and quality of student judgment, rather than mere presence of AI use. The AI-assisted task must be defensible in nature as testing and verification alone do not imply understanding of how that task works.

Next, demonstration-based assessments including oral walkthroughs, live modifications, and targeted questioning, provide direct evidence of mastery and ensure assessment integrity in AI-supported environments. Finally, structured feedback enables calibration and iteration for both students and instructors. This continuous evaluation loop reduces enforcement fatigue, supports consistent grading, and allows the framework to be replicated across courses, levels, and institutions with minimal modification. Each component is designed to transform the use of AI from a potential shortcut into a structured learning tool. By documenting AI contributions and embedding reflective practices, the framework encourages students to critically evaluate suggestions, verify correctness, and engage in ethical decision-making. This approach addresses faculty concerns regarding unregulated AI use and provides a replicable model for responsible technology integration.

Figure 1 illustrates the framework. The proposed framework integrates generative AI into CS coursework through continuous faculty evaluation rather than post hoc enforcement. Oversight is embedded across all stages, ensuring AI use remains structured, auditable, equitable, and replicable across courses and institutions. Rather than attempting to detect or prohibit AI usage, the framework aligns instructional design, assessment, and accountability mechanisms with demonstrated student understanding and responsible professional practice.

Case Studies: Applying the Proposed Transparent AI Accountability Framework

To evaluate the practical impact of the transparent AI accountability framework, we applied it in two courses at different levels of the curriculum: an introductory-level programming course and an upper-level CS course. Together, these cases illustrate how the same core principles of explicit expectations, documentation, reflection, and accountability, can be adapted to distinct pedagogical contexts while preserving learning objectives and assessment integrity.

Introductory Level Programming CS Course

In an introductory-level programming course, the framework was implemented in a practicum lab following a POGIL-based class session and a pair-programming lab

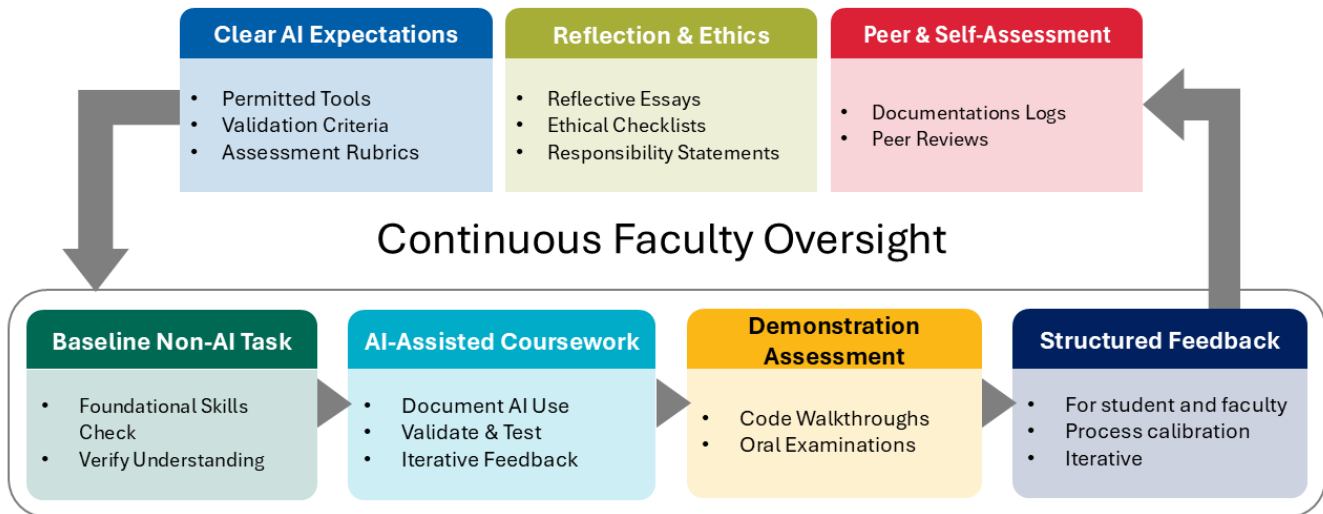


Figure 1: Proposed Transparent AI Accountability Framework: AI-Assisted Activities and Continuous Improvement Loop

activity on indefinite loops that were conducted without AI assistance the prior week. The practicum lab was explicitly structured as an AI partnership activity with an accountability-focused rubric. Students first completed a manual warm-up problem without AI assistance, then engaged in a guided workflow in which they used a generative AI tool as a partner to solve additional, more complex loop problems.

To ensure transparency and student ownership, students maintained simple logs of accepted, modified, and rejected AI suggestions; documented test cases; completed an ethics and accountability checklist; and signed a certification of accountability stating that they had reviewed, verified, and took full responsibility for all submitted work. Deliverables included the program files, a short-form AI Use Statement, and a brief oral walkthrough in which students explained loop logic, demonstrated how they detected and fixed an AI error, and performed a live modification under instructor questioning.

Upper-Level CS Course

In this upper-level CS course on Operating Systems concepts, the framework was implemented following a pilot study with five students, which informed refinements applied to a current class of seven students. The course focused on system-level assignments that required students to integrate AI support responsibly while solving complex process scheduling and concurrency problems. Students first completed baseline activities without AI assistance, allowing faculty to verify foundational understanding and establish equitable starting conditions. They then engaged in AI partnership workflows, structured according to the transparent accountability framework and guided by standardized rubrics.

To ensure transparency and student ownership, students

maintained detailed logs of AI interactions, specifying accepted, modified, or rejected outputs. They documented test cases, used formal verification tools, and completed reflective essays describing AI contributions, validation steps, and ethical decision-making. Deliverables included fully functional system simulations, interactive Gantt charts, AI documentation, reflective essays, and demonstration-based assessments where students explained their design choices and performed live modifications under instructor questioning.

Faculty observations during the pilot revealed several key issues: some students over-relied on AI outputs without critical evaluation, documentation of AI interactions was inconsistent, and reflection on ethical AI use varied. Differences in engagement highlighted the need for clearer expectations, structured reflection prompts, and standardized assessment rubrics. The refined implementation addressed these challenges: students received explicit AI-use guidelines, structured reflection assignments, and enhanced verification and visualization tools.

Faculty evaluation focused on judgment, correctness, completeness, reflection and equitable participation, rather than mere AI usage. Structured rubrics emphasized: AI Interaction Documentation, critical evaluation of AI outputs, technical correctness, reflection and ethical use, visualization and communication, engagement and equity, and problem-solving and adaptation.

Aggregated rubric scores across these dimensions were calculated to provide quantitative evidence of learning impact: students achieved an average of 4.3/5 for AI Interaction Documentation, 4.1/5 for Reflection and Ethical Use, and 4.0/5 for Problem-Solving and Adaptation. These aggregated results reinforce the claims of enhanced engagement, learning, and fairness under the proposed framework.

Observations confirmed that these refinements improved engagement and accountability: students actively partici-

pated in AI-supported problem-solving, critically analyzed outputs, and integrated feedback effectively. Faculty were able to verify equitable participation, ensuring all students had comparable opportunities and preventing hidden advantages.

Through this approach, students developed practical skills in AI-assisted design, verification, and problem-solving, strengthened their understanding of scheduling and synchronization mechanisms, and practiced responsible AI use. The structured oversight and standardized assessment workflow allowed faculty to reliably evaluate engagement, technical correctness, reflection, and equitable participation. These results demonstrate that the framework is robust, replicable, and scalable to larger cohorts, preserving learning outcomes and aligning with academic and ethical standards for AI-assisted learning.

Cross-Case Observations

Across both introductory-level and upper-level courses, the framework produced positive outcomes for both students and faculty despite the differences in course level and technical complexity. Students demonstrated improved understanding of complex topics, actively engaged in problem-solving, and produced work with higher correctness and clarity. Reflective essays provided evidence of students' ethical awareness, critical evaluation of AI suggestions, and responsible integration of AI into their workflows. Faculty observed reduced enforcement fatigue, as the framework shifted the focus from policing AI usage to supporting learning outcomes. Structured accountability mechanisms allowed instructors to objectively assess student performance, ensure fairness even when AI tools were used and concentrate on guiding learning rather than monitoring compliance.

Equity and fairness were central considerations in the framework. By providing structured access to AI tools, all students were able to engage with advanced computational resources, reducing disparities based on prior experience or external access. Documentation of AI contributions and reflective essays ensured transparency, allowing instructors to evaluate student work equitably. Students reported increased confidence in their problem-solving abilities, and faculty noted enhanced learning quality across the cohort. The integration of AI in this structured and transparent manner allowed students to take ownership of their learning, demonstrating accountability while maintaining the integrity of the course objectives.

Implementation Costs, Sustainability, and Scalability

Implementing this framework requires an initial investment in student on-boarding to clarify expectations and documentation standards. However, this represents a strategic reallocation of effort rather than a net increase in labor. By shifting the focus from the unproductive "arms race" of AI detection to the auditable assessment of student reasoning, faculty experience a significant reduction in enforcement fatigue and policing burdens. This transformation allows instructional

time to be spent on guiding learning rather than monitoring compliance.

Sustainability and scalability are achieved using standardized "traceability artifacts," such as interaction logs, test matrices, and accountability certifications. These templates make AI use visible and instructionally consequential across diverse course levels and contexts. In large-enrollment settings, these structured artifacts provide an objective "paper trail" that allows instructional staff to reliably evaluate engagement and technical correctness without the need for exhaustive individual surveillance. This design pattern ensures the framework remains robust and replicable across institutional contexts while preserving the integrity of learning outcomes.

Limitations & Future Work

While this study provides a detailed account of the framework's implementation, we acknowledge the limitations inherent in its exploratory nature, including small sample sizes and the lack of a controlled comparative group. As an initial pilot phase, the primary objective was to evaluate the framework's feasibility and refine the accountability artifacts across different instructional levels, rather than to establish broad statistical generalizability.

Future work will leverage the refinements identified in this pilot to conduct a large-scale empirical study. This will include a controlled comparison between the accountability framework and unrestricted AI use, utilizing quantitative metrics to measure impacts on student learning outcomes, technical proficiency, and perceptions of academic integrity across larger, more diverse cohorts.

Conclusion

The structured integration of AI into undergraduate CS courses can enhance learning, strengthen ethical awareness, and reduce faculty burden when accompanied by a transparent accountability framework. The rapid and largely unstructured adoption of generative AI in undergraduate CS has produced well-documented challenges: over-reliance, acceptance of hallucinated outputs, erosion of planning and verification practices, opacity in student work, and growing enforcement burdens for instructors. Prior work shows that neither prohibition nor unrestricted access resolves these issues, instead calling for structured, transparent, and pedagogically grounded integration mechanisms, which are often absent from current course designs.

Experience in introductory-level and upper-level CS courses demonstrate that a transparent AI accountability framework can directly address these gaps by making AI use visible, auditable, and instructionally consequential. Structured integration shifts student behavior away from solution outsourcing toward critical evaluation, verification, and explanation, precisely the practices that unstructured LLM use can undermine. By requiring documentation, reflection, and explicit ownership of AI-assisted work, the framework mitigates AI-driven shortcuts, reduces faculty policing effort, and preserves assessment integrity. More broadly, this work contributes a replicable course-level design pattern for inte-

grating AI responsibly, sustaining learning, equity, and foundational skill development in STEM education.

References

- Andleeb, S.; Kantorski, B.; and Carver, J. 2025. Chatgpt in introductory programming: Counterbalanced evaluation of code quality, conceptual learning, and student perceptions. In *Proceedings of the 26th ACM Annual Conference on Cybersecurity & Information Technology Education, SIGCITE '25*, 180–186. New York, NY, USA: Association for Computing Machinery.
- Farinetti, L., and Cagliero, L. 2025. A critical approach to chatgpt: An experience in sql learning. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1, SIGCSETS 2025*, 318–324. New York, NY, USA: Association for Computing Machinery.
- Hak, J.; Lam Johnson, N.; Amoozadeh, M.; Alipour, A.; and Chattopadhyay, S. 2025. Observing without doing: Pseudo-apprenticeship patterns in student llm use. In *Proceedings of the 25th Koli Calling International Conference on Computing Education Research, Koli Calling '25*. New York, NY, USA: Association for Computing Machinery.
- Jamie, P.; HajiHashemi, R.; and Alipour, S. 2025. Utilizing chatgpt in a data structures and algorithms course: A teaching assistant's perspective. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, CHI EA '25*. New York, NY, USA: Association for Computing Machinery.
- Joshi, I.; Budhiraja, R.; Dev, H.; Kadia, J.; Ataullah, M. O.; Mitra, S.; Akolekar, H. D.; and Kumar, D. 2024. Chatgpt in the classroom: An analysis of its strengths and weaknesses for solving undergraduate computer science questions. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1, SIGCSE 2024*, 625–631. New York, NY, USA: Association for Computing Machinery.
- Park, H., and Ahn, D. 2024. The promise and peril of chatgpt in higher education: Opportunities, challenges, and design implications. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, CHI '24*. New York, NY, USA: Association for Computing Machinery.
- Prather, J.; Reeves, B. N.; Leinonen, J.; MacNeil, S.; Randrianasolo, A. S.; Becker, B. A.; Kimmel, B.; Wright, J.; and Briggs, B. 2024. The widening gap: The benefits and harms of generative ai for novice programmers. In *Proceedings of the 2024 ACM Conference on International Computing Education Research - Volume 1, ICER '24*, 469–486. New York, NY, USA: Association for Computing Machinery.
- Ramirez Osorio, V.; Zavaleta Bernuy, A.; Simion, B.; and Liut, M. 2025. Understanding the impact of using generative ai tools in a database course. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1, SIGCSETS 2025*, 959–965. New York, NY, USA: Association for Computing Machinery.
- Zviel-Girshin, R. 2024. The good and bad of ai tools in novice programming education. *Education Sciences* 14(10).