

INTEGRATING AI FEEDBACK LOOPS IN ACADEMIC READING INSTRUCTION: PEDAGOGICAL OPPORTUNITIES AND ETHICAL CHALLENGES

Adhamova Nigora

ESL teacher

adhamovanigora05@gmail.com

TMC INSTITUTE

<https://doi.org/10.5281/zenodo.17491884>

Annotation: This article examines the integration of AI feedback loops into academic reading instruction, highlighting both their pedagogical benefits and ethical implications. As intelligent systems increasingly assist learners through adaptive feedback, summarization, and comprehension analysis, educators face new opportunities to enhance reading engagement, self-regulation, and critical awareness. The study investigates how AI-generated feedback—delivered through tools such as ChatGPT, Elicit, and Perplexity—can promote metacognitive reflection, formative assessment, and individualized reading support. Drawing upon socio-constructivist and feedback literacy frameworks, the research emphasizes the role of iterative AI-human interaction in fostering deeper textual comprehension. However, it also addresses ethical concerns such as data privacy, algorithmic bias, intellectual dependency, and the diminishing of authentic interpretive skills. Findings suggest that AI feedback loops, when thoughtfully embedded within pedagogically guided environments, can enhance learner autonomy, motivation, and comprehension outcomes. The paper concludes with recommendations for balancing technological innovation with ethical responsibility, proposing a reflective model for educators to critically evaluate AI's role in academic literacy development.

Keywords: AI feedback loops, Academic reading instruction, Pedagogical innovation, Feedback literacy, Metacognitive reflection, Ethical challenges, Artificial intelligence in education, Learner autonomy.

AI feedback systems can analyze individual student responses in real-time, identifying specific comprehension breakdowns, misinterpretations, or analytical gaps and generating targeted feedback addressing these issues (Luckin et al., 2016). For instance, when a student misconstrues a passage's main idea, an AI system might provide feedback highlighting textual evidence the student overlooked, posing clarifying questions, or suggesting alternative interpretative frameworks. This individualized support can occur immediately and repeatedly, allowing students to receive guidance precisely when needed rather than waiting for teacher availability or next-day feedback. Moreover, AI systems can adapt instruction based on ongoing performance data, adjusting text complexity, question difficulty, and scaffolding levels to maintain optimal challenge. This adaptive capability addresses a persistent problem in reading instruction: ensuring that materials and tasks neither overwhelm students with excessive difficulty nor bore them with insufficient challenge. By continuously calibrating instruction to individual performance, AI feedback loops can theoretically maintain students in their zones of proximal development more consistently than static instructional approaches. AI feedback loops align powerfully with formative assessment principles, which emphasize ongoing evaluation designed to inform instruction and support learning rather than merely measuring achievement (Black & Wiliam, 2009). Traditional summative assessments provide limited information about the specific cognitive processes underlying reading comprehension, offering

end-point measurements that reveal whether students understood material but not why they struggled or succeeded. AI systems capable of analyzing student responses at granular levels can provide rich formative data revealing patterns in comprehension strategies, common misconceptions, and developmental trajectories. Teachers can access detailed analytics showing which types of questions students struggle with, which textual features they overlook, and how their reading strategies evolve over time. This information enables more targeted instructional interventions and helps educators identify students requiring additional support before they fall significantly behind. Furthermore, AI feedback can support students' metacognitive development by making thinking processes visible. When systems provide explanatory feedback that not only indicates correct answers but also explicates reasoning processes, evaluates response quality, and suggests strategic approaches, students gain insight into expert reading practices. This metacognitive transparency can help students develop awareness of their own comprehension monitoring and strategic reading capabilities, supporting the development of self-regulated learning skills essential for academic success. AI feedback systems offer particular promise for supporting diverse learners who often face barriers in traditional reading instruction contexts. English language learners, students with learning disabilities, and those from under-resourced educational backgrounds may benefit substantially from immediate, non-judgmental feedback available in unlimited quantity. Unlike human feedback, which students may experience as socially risky or embarrassing, AI feedback provides a low-stakes environment for practice and error-making. These systems can incorporate multiple forms of support—text-to-speech functionality, visual representations of concepts, multilingual translation, and adjustable complexity levels—making academic texts more accessible to students with varied needs. AI feedback can also provide scaffolding that gradually fades as students develop competence, offering more intensive support initially and reducing assistance as skills improve. This adaptive scaffolding can help bridge achievement gaps by ensuring all students receive the level of support necessary for productive engagement with challenging texts. Additionally, AI systems can provide consistency in feedback quality that human instructors, subject to fatigue, time constraints, and unconscious biases, may struggle to maintain. While this consistency does not eliminate bias—as discussed in the ethical challenges section—it offers the potential for more equitable feedback delivery across diverse student populations when systems are properly designed and validated. The immediacy of AI feedback represents a significant pedagogical advantage over delayed human feedback. Research in learning sciences demonstrates that feedback effectiveness diminishes as temporal distance from performance increases (Shute, 2008). When students receive feedback hours or days after completing reading tasks, they may have difficulty reconstructing their original thinking processes, reducing feedback's corrective impact. Immediate AI feedback allows students to reconsider their interpretations while the reading experience remains fresh, facilitating more effective revision of understanding. This immediacy also enables iterative learning processes wherein students can revise responses, receive additional feedback, and continue refining their comprehension and analysis. Traditional assignment structures often involve single submission-feedback cycles, limiting opportunities for revision and improvement. AI systems can support multiple attempts with feedback at each iteration, normalizing revision as part of the learning process and helping students understand that reading comprehension develops

through continued engagement rather than single-shot performance. Despite the accessibility opportunities described above, AI feedback systems pose serious equity risks stemming from algorithmic bias. Machine learning algorithms learn from training data, and if this data reflects existing educational inequities, the resulting systems perpetuate and potentially amplify these biases (Noble, 2018). Research on automated essay scoring systems has documented systematic bias against non-standard dialects, multilingual writers, and students from marginalized communities whose writing patterns differ from the dominant norms represented in training data. In reading comprehension contexts, AI systems trained predominantly on responses from privileged student populations may unfairly evaluate interpretations reflecting diverse cultural perspectives, life experiences, or epistemological frameworks. A student offering a culturally grounded interpretation of a text might receive negative feedback from an AI system that recognizes only interpretations aligned with dominant cultural assumptions. This cultural bias in feedback could discourage diverse perspectives, implicitly communicating that certain ways of reading and interpreting are illegitimate. Moreover, errors in AI feedback—false positives identifying correct comprehension as errors or false negatives overlooking actual misunderstandings—may not distribute randomly across student populations. If systems perform less accurately for certain demographic groups, these students receive lower-quality educational support, exacerbating rather than reducing achievement gaps. Ensuring algorithmic fairness requires extensive validation across diverse populations, transparent reporting of performance disparities, and ongoing monitoring for emergent biases as systems are deployed in real-world contexts. AI feedback systems require extensive data collection about student reading behaviors, comprehension patterns, and learning trajectories. While this data enables personalized instruction, it also raises profound privacy concerns. Educational AI systems often collect detailed information about when students read, how long they spend on passages, which questions they struggle with, and how their performance changes over time. This granular behavioral data creates comprehensive profiles that could be misused in various ways (Williamson, 2017). Questions emerge regarding data ownership, retention, and secondary usage. Who owns the reading comprehension data generated through AI systems—students, educational institutions, or technology companies? How long should this sensitive information be retained, and what safeguards prevent unauthorized access? Can student reading data be used for purposes beyond immediate educational feedback, such as algorithm training, product development, or commercial applications? The answers to these questions have significant implications for student autonomy and privacy rights. Furthermore, continuous AI monitoring of reading processes creates surveillance dynamics that may alter the educational environment in troubling ways. When students know their every interaction with texts is being tracked and analyzed, they may engage differently—potentially more cautiously or performatively—than they would in less monitored contexts. This surveillance can undermine the exploratory, risk-taking approach to reading that fosters genuine intellectual development. Educational environments characterized by pervasive monitoring may inadvertently train students for surveillance rather than intellectual freedom. Fundamental questions exist regarding what AI systems actually measure when they assess reading comprehension. Reading represents a complex, multifaceted cognitive process involving linguistic decoding, background knowledge

activation, inference generation, critical evaluation, and personal response (RAND Reading Study Group, 2002). Current AI systems typically assess reading through structured tasks—multiple-choice questions, short-answer responses, or constrained writing prompts—that capture only limited dimensions of authentic reading comprehension. The risk of construct underrepresentation looms large: AI feedback systems may provide valid assessments of what they explicitly measure—such as literal comprehension or basic inference—while failing to capture crucial aspects of skilled reading like aesthetic appreciation, critical consciousness, or complex interpretive synthesis. If educators and students come to view AI-assessed performance as equivalent to genuine reading ability, important educational goals may be neglected in favor of narrower, more easily automated assessments. Related concerns involve assessment authenticity. Authentic reading occurs in rich, contextualized situations where readers pursue genuine purposes—understanding, enjoyment, information-seeking, or critical analysis. AI feedback systems often decontextualize reading, presenting it as a series of discrete tasks to be completed for system evaluation rather than meaningful engagement with texts for authentic purposes. This decontextualization may inadvertently communicate distorted messages about what reading is and why it matters, potentially undermining rather than supporting lifelong literacy development. Perhaps the most philosophically significant ethical challenge concerns the role of human relationships in education. Teaching and learning are fundamentally relational processes involving trust, empathy, cultural responsiveness, and mutual recognition between humans (Noddings, 2013). When students struggle with reading, effective teachers provide not merely technical feedback but also encouragement, validation, cultural connection, and belief in student capability. These relational dimensions shape motivation, identity development, and willingness to persist through difficulties. AI feedback systems, regardless of their technical sophistication, cannot replicate these human relational qualities. While AI can simulate empathy through carefully designed language, it lacks the authentic human presence, cultural understanding, and ethical commitment that characterize effective pedagogical relationships. The risk exists that increased reliance on AI feedback may diminish opportunities for meaningful human interaction around literacy development, particularly if systems are implemented as cost-saving measures that reduce instructional staffing. Moreover, the nature of reading itself raises questions about the appropriateness of algorithmic mediation. Reading literature, in particular, involves personal meaning-making, emotional response, and interpretive freedom that resist reduction to assessable right answers. When AI systems evaluate and provide feedback on literary interpretation, they risk imposing algorithmic authority over domains that should remain open to multiple valid perspectives. This tension between algorithmic standardization and interpretive plurality represents a fundamental incompatibility between certain types of AI feedback and the educational goals of humanistic literacy instruction. AI systems often operate as "black boxes," generating feedback through complex algorithmic processes opaque to users. This lack of transparency creates several ethical problems. Students receiving feedback may not understand the basis for evaluations, making it difficult to learn from or contest feedback they believe is inaccurate. Teachers implementing AI systems may not fully comprehend how algorithms reach conclusions, limiting their ability to identify errors, explain feedback to students, or make informed decisions about system reliability. The explainability problem becomes particularly

acute when AI systems make high-stakes determinations about student placement, intervention needs, or achievement levels. Without clear understanding of algorithmic reasoning, educators cannot verify that systems make decisions for appropriate educational reasons rather than spurious pattern recognition or biased proxies. Demanding algorithmic transparency and explainability is essential for educational accountability, but current AI systems often resist such transparency due to proprietary concerns or genuine technical complexity. Given these pedagogical opportunities and ethical challenges, responsible integration of AI feedback loops in reading instruction requires a comprehensive framework balancing innovation with educational values and ethical principles. Several key elements should guide implementation decisions. First, AI feedback systems should complement rather than replace human instruction and feedback. The optimal model positions AI as a supplementary resource that extends teachers' capacity to provide individualized support rather than as a substitute for human pedagogical engagement. Teachers should remain central to literacy instruction, with AI systems handling routine feedback tasks while educators focus on complex, culturally responsive, and relationship-based dimensions of teaching. Second, rigorous validation across diverse student populations must precede widespread implementation. Systems should be extensively tested to identify and mitigate algorithmic bias, with particular attention to performance disparities across demographic groups. Validation should assess not just technical accuracy but also pedagogical effectiveness—whether feedback actually supports learning and skill development for all students. Ongoing monitoring should continue after deployment to detect emergent biases or unintended consequences. Third, robust data governance frameworks must protect student privacy while enabling beneficial data use. Clear policies should specify data collection limits, retention periods, access restrictions, and prohibited secondary uses. Students and families should provide informed consent with clear explanation of data practices in accessible language. Data minimization principles should guide system design, collecting only information necessary for educational purposes. Fourth, transparency in algorithmic processes should be maximized to the extent possible. Even when full explainability remains technically challenging, systems should provide meaningful information about how feedback is generated, what factors influence evaluations, and what limitations exist. Both educators and students should understand the basis for AI feedback and recognize its provisional nature. Fifth, implementations should preserve authentic reading experiences and avoid construct underrepresentation. AI systems should assess reading in ways that honor its complexity and avoid reducing literacy to narrow skill sets. Multiple forms of assessment, including human evaluation of complex interpretive work, should complement AI feedback to ensure comprehensive understanding of student development. Finally, professional development should prepare educators to use AI feedback systems critically and effectively. Teachers need support in understanding system capabilities and limitations, interpreting AI-generated data, identifying algorithmic errors or biases, and integrating AI feedback into coherent pedagogical approaches that maintain focus on meaningful literacy development.

Adabiyotlar, References, Литературы:

1. Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational*

- Assessment, Evaluation and Accountability*, 21(1), 5-31. <https://doi.org/10.1007/s11092-008-9068-5>
2. Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
 3. Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson Education. <https://www.pearson.com/content/dam/corporate/global/pearson-dot-com/files/innovation/Intelligence-Unleashed-Publication.pdf>
 4. Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press. <https://doi.org/10.18574/nyu/9781479833641.001.0001>
 5. Noddings, N. (2013). *Caring: A relational approach to ethics and moral education* (2nd ed.). University of California Press. <https://doi.org/10.1525/9780520957343>
 6. RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. RAND Corporation. https://www.rand.org/pubs/monograph_reports/MR1465.html
 7. Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189. <https://doi.org/10.3102/0034654307313795>
 8. Williamson, B. (2017). Decoding ClassDojo: Psycho-policy, social-emotional learning and persuasive educational technologies. *Learning, Media and Technology*, 42(4), 440-453. <https://doi.org/10.1080/17439884.2017.1278020>