"Debate Guru": Honing Public Speaking Skills Among Secondary School Students with AI Tutoring Systems

Zachary Nadan[1], Nathan Goldberg[1], Roozbeh Aliabadi[2], Shuhan Li[2]

[1]Debate Guru AI

[2]ReadyAI

zach.nadan@gmail.com, nathanmauricegoldberg@gmail.com, rooz@readyai.org, shuhan.li@readyai.org

Abstract

AI-supported tools have entered K-12 classrooms in recent years to reshape student learning and skill-building. We are particularly interested in AI's application in literacy subjects, such as English, where students are expected to hone their critical thinking and public speaking skills through AI interactions. This report details the pilot implementation of Debate Guru, an AI-enhanced debate education platform, across two secondary schools with varying instructional contexts. Over the course of a summer school course, educators integrated Debate Guru in one of two ways: 1) by using Debate Guru's complete curriculum; or 2) by combining platform resources with their own instruction such as a literary text. The pilot was implemented with approximately 50 students across Grade 8-11. Findings suggest significant increases in students' confidence, argumentative reasoning, and engagement. Teachers reported high usability and strong pedagogical value, while students responded positively to interactive features and AI feedback integration. The pilot demonstrated preliminary usefulness and scalability of AI tutoring systems in debate classes. Future iterations can benefit from longitudinal studies, technical refinements, and school capacity-building to ensure greater generalizability of results and deeper knowledge acquisition.

Introduction

The integration of artificial intelligence (AI) into education has become a defining feature of the 21st-century classroom. From personalized tutoring systems to real-time language feedback and intelligent writing assistants, AI is reshaping the way students learn and how educators teach (Kasih et al., 2024). Especially after digital learning has become a growing trend, a significant body of research and educational practices has emerged regarding AI-supported literary subjects (Kasih et al., 2024; Labadze et al., 2023). AI's impacts on skill-based disciplines such as reading and writing are evident; personalization allows for timely, specific, and actionable feedback that accelerates students' learning process (Kasih et al., 2024; Labadze et al., 2023). Although some students and scholars maintain skepticism regarding AI ethics and self-motivated learning (Labadze et al., 2023), AI tools like chatbots offer educators versatile means of course design, tailored pedagogies, and assessments (Labadze et al., 2023). The bulk of relevant research focuses on writing-intensive courses (Kasih et al., 2024). However, few studies exist addressing the application of AI to more sophisticated skillsets, such as debate and public speaking. While one recent study (Zhang, 2024) explored the successful intervention of AI in college-level civic education, upscaled AI integration in debate-based classrooms with younger students has historically been underrepresented.

Debate Guru responds to the emerging opportunity of broadening AI-supported debate training among secondary school students. This initiative aims to leverage AI for maximizing students' skills in evidence-based reasoning, critical thinking, and argumentation under pressure. Originally designed as an AI-powered online platform for debate education, Debate Guru blends intelligent tutoring systems (ITS) with argument mining and natural language processing (NLP) to support students as they practice and refine their argumentation skills. The platform's core features include AI-generated feedback on speeches and writing, real-time rebuttal challenges, interactive speaking drills, and a built-in curriculum aligned with research-backed pedagogical goals. More than a tool for generating arguments, Debate Guru is intended to serve as a full-scale instructional support system for developing civic literacy, critical thinking, and persuasive communication.

This experience report presents the findings from a pilot implementation of Debate Guru across two school settings. The pilot involved students in 8th-11th grade and encompassed both a full curriculum-based integration and a literature-adapted instructional model. We conducted pre-post surveys, along with classroom observations and interviews, to holistically document the in-class dynamics and assess the preliminary results. Drawing from both qualitative and quantitative data, we evaluate student outcomes,

teacher feedback, and classroom-level implementation strategies. In doing so, we aim to contribute to emerging conversations about AI in K-12 education and offer practical insights into how AI can enrich student debate and communication skills at scale. To guide our inquiry, we propose the following research questions:

- RQ1: To what extent does Debate Guru impact students' debate-related competencies, AI engagement, and self-efficacy?
- RQ2: What are the differences, if any, in learning outcomes between the two implementation methods?
- RQ3: What design elements in Debate Guru contribute to students' learning experience and outcomes?

Literature Review

Intelligent Tutoring System (ITS)

Intelligent Tutoring System (ITS) refers to the computer-based tutoring mechanism that provides feedback, debriefing, and step-by-step instructions based on students' inquiries (VanLehn, 2015). ITS is characterized by electronic modalities, natural language dialogues, simulations, and/or interactive interfaces for problem solving (VanLehn, 2015). Although ITS is commonly believed to underperform human tutors, progress reports have shown that digital systems demonstrate equal effectiveness as individualized human tutoring, particularly for STEM topics (VanLehn, 2015).

ITS has been used for enhanced learning experience in K-12 education. Earlier ITS interventions focused on web-based interface with practice assignments and multimedia modules (Wijekumar, 2006; Wijekumar et al., 2012). More recent systems incorporated AI and machine learning techniques to generate personalized reports for real-time assessments (Ni & Cheng, 2022). Some ITS interfaces also include gamification to sustain student engagement (Wijekumar, 2006). Empirical evidence shows the promising outcomes of ITS-integrated classrooms. One experimental trial with 4th grade students observed immediate and statistically significant improvement in reading comprehension among 4th-grade students, who received 6-7 months of weekly ITS instruction (Wijekumar et al., 2012). ITS is also able to facilitate English learning as a foreign language by identifying knowledge blind spots and offering consistent exercise (Ni & Cheng, 2022).

ITS comes with both opportunities and challenges. While ITS can facilitate self-paced and personalized learning (Singh et al., 2022; Wijekumar, 2006), perceived usefulness and ease of use, as well as consistency across digital system designs, may limit its actual impact on students (Ni & Cheng, 2022; Wijekumar, 2006). The lack of usability jeopardizes learner retention and self-efficacy that hinders knowledge acquisition. Besides technical constraints, content quality also affects effectiveness. Repetitive learning materials, for example, risk disorienting learners (Singh et al., 2022), leading to disengagement with the lesson. Therefore, successful ITS shall not only address technical robustness but also diversify the tasks and assignments (Ni & Cheng, 2022). More importantly, ITS is recommended as a supplementary tool for homework and in-class activities, rather than a replacement of the comprehensive classroom experience (VanLehn, 2015).

Argumentation Mining with AI

Argumentation mining is the process that automatically detects the arguments in a text (Mochales & Moens, 2011). An argumentation mining model is able to identify not just the property of arguments, but also the logical structure within an argument and interactions between multiple arguments (Mochales & Moens, 2011). Relevant research emphasizes discourse analysis and theoretical construction for modeling and interpreting textual data (Habernal & Gurevych, 2017). The state-of-the-art model is built upon classic argumentation theory consisting of 6 distinct components: Claim, Grounds, Warrant, Backing, Qualifier, and Rebuttal (Habernal & Gurevych, 2017). Experiments with legal texts presented the prospect of AI-backed mining techniques (Mochales & Moens, 2011). Other experiments with web-based discourse revealed the defining factors of model reliability, such as topics, length, and stylistic choices in a given text (Habernal & Gurevych, 2017). Most recent research looked into different deep learning models' performance in classifying argumentation, yielding varying capacities across multiple natural language processing models (Gunawan & Suhartono, 2023).

In practice, argumentation mining is used for sentiment analysis and educational purposes. As part of a structured repository, argumentation mining helps users query theme-based arguments and facilitate interoperability among various agents (Abbas & Sawamura, 2011). When combined with other classifiers and mining techniques, the model can analyze and retrieve the most relevant information regarding the subject matter (Abbas & Sawamura, 2011). Application in an agent-based educational environment is proposed to deepen users' analytical, decision-making, and communication skills (Abbas & Sawamura, 2011). Argumentation mining can also benefit teachers. One model trained on scientific writings from 2472 students is able to detect the semantic patterns and classify key arguments in specific domains (Xing et al., 2020). It also recognizes sentiments such as uncertainty. This model has the potential to assist teachers with evaluation and informed tutoring (Xing et al., 2020).

AI in K-12 Classrooms

Early intervention programs aimed to prepare young students for a digitized future. In an evaluative study, students worked with Teachable Agents (TA) that entails machine learning techniques, allowing students to train a computer agent for learner support (Chin et al., 2010). Comparative analysis of two classroom sessions revealed that the incorporation of TA provided additional values to students' scientific understanding, without hindering the intake from unplugged regular curriculum (Chin et al., 2010). Integrating TA in K-12 classrooms forges students' enduring understanding and provides them with the prerequisite knowledge for continuous learning (Chin et al., 2010).

With the rapid development of AI, recent endeavors pertain to guidelines and best practices of AI education in K-12 settings. Several articles discussed the opportunities and challenges in adopting AI for K-12 education. In a literature review, Pargman et al. (2025) pointed out that although AI can enhance personalization and predictive analytics, ethical issues need to be addressed such as consent, data privacy, and risks of misuse. Another review discussed the future directions of using generative AI (GenAI) as a learning analytics tool (Khosravi et al., 2025). The authors noted both the usefulness of GenAI as a collaborative agent for student learning and the pressing needs of learner-centered tools (Khosravi et al., 2025). Teachers also expressed various perspectives on GenAI, citing merits of information retrieval and efficiency improvements, as well as concerns of misinformation and overreliance (Chen et al., 2024). Grade-specific approaches to AI in schools were also highlighted (Chen et al., 2024).

On-the-ground implementation of K-12 AI curriculum evidences effective methods in AI literacy. In one remote-learning AI literacy program using conversational agents, high school students saw significant internationalization of AI competencies, including AI mechanisms, data analysis, and critical appraisals of AI (Van Brummelen et al., 2021). Said study also revealed the imperativeness of teacher preparedness and student-centered environment (Van Brummelen et al., 2021). Specifically, hands-on activities with AI maximizes interactivity and thus augments students' engagement. This further strengthens students' learning outcomes (Van Brummelen et al., 2021). While topics of AI ethics require further scaffolding, dialogue- and activity-driven lessons can deepen K-12 students' literacy and computational thinking skills (Van Brummelen et al., 2021).

Implementation Contexts

The Debate Guru platform was deployed in two distinct instructional contexts that offered insight into its adaptability and impact.

School A: Combination with Literary Texts

School A operated within a district facing significant socioeconomic challenges, with a median household income around $77,171 and a poverty rate of 13.6%. Although these figures may seem moderate, they reflect a context where many families experience ongoing material insecurity. Educational research links such conditions to limited access to enrichment opportunities, fewer supplemental academic supports, and reduced exposure to formal public speaking or structured argumentation. Thus, adopting Debate Guru at School A was more than an instructional choice; it was a potential intervention in a setting where students likely had minimal prior experience with competitive or academic debate.

Debate Guru was integrated into a summer-school English course focused on a literary analysis unit. The anchor text was Anton Chekhov's *The Bet*, selected for its concise narrative and rich ethical themes, such as justice, capital punishment, and the psychological impact of isolation. Recognizing the platform's flexibility, the instructor added the 1988 film, *Stand and Deliver,* for students to additionally analyze. This allowed students to engage in a comparative ethical analysis, weighing different moral frameworks, narrative resolutions, and character motivations.

Within this two-format framework, the teacher crafted AI-mediated prompts designed to deepen thematic engagement. These prompts moved beyond basic comprehension, functioning as deliberate provocations: Was the banker's moral failure worse than the lawyer's hubris? How might the fictional justice systems reflect real-world inequalities? Could a character's moral choices be reevaluated using contemporary ethical theories? Students interacted with AI-generated sample arguments and rebuttals as critical interlocutors—testing the AI's reasoning, spotting logical gaps, and sometimes challenging its ethical assumptions. This dialogic exchange between human learners and the AI fostered metacognitive awareness about argument quality, bias, and persuasiveness.

The AI's role went beyond modeling argumentation; it provided adaptive scaffolding by adjusting the complexity of counterarguments based on student input. For quieter or less confident students who are often marginalized in fast-paced oral debates, this created a low-pressure way to engage in rhetorical contestation. The teacher used the platform's analytics dashboard to monitor not just accuracy and argument structure but also engagement indicators like voluntary rebuttals and depth of evidence.

In this way, Debate Guru's thematic integration at School A served multiple functions: reinforcing close reading and literary interpretation, cultivating transferable skills in evidence-based reasoning, and introducing ethical deliberation that was both personal and public, mediated by AI yet anchored in shared texts. In an under-resourced educa-

tional environment, the platform operated as both a literary companion and a catalyst for ethical inquiry, providing students with structured opportunities to articulate and refine their moral reasoning within a digitally mediated public sphere.

School B: Complete AI-Integrated Curriculum

School B's implementation of the Debate Guru platform within an 8th-grade Social Studies context exemplifies a comprehensive, scaffolded approach to debate instruction that integrates AI as a central pedagogical tool. Over the course of ten consecutive lessons spanning two weeks, the curriculum systematically introduced students to foundational debate skills and progressively advanced toward complex rhetorical and metacognitive competencies. The instructional design foregrounded essential debate concepts, including the structure of arguments via the claim-evidence-impact (CEI) model, rebuttal and refutation techniques, and interactive debate strategies such as Points of Information (POIs) and respectful interruption. This progression culminated in public speaking drills, team debate practice, and reflective activities, all tightly aligned to build students' confidence and competence in argumentation and oral communication. Such a sequenced, incremental approach ensured that students engaged with the material at multiple levels (cognitive, affective, and performative) allowing for sustained skill development over time.

Central to the innovative curriculum was the integration of Debate Guru's AI capabilities, which served multiple instructional functions. AI-generated arguments provided students with examples of claim construction and logical organization, while its ability to simulate opponents offered a dynamic, low-stakes environment for students to practice rebuttals and develop real-time response skills. The AI's adaptive scaffolding adjusted the complexity of counterarguments to meet students' varying proficiency levels, thereby facilitating differentiated learning opportunities within a heterogeneous classroom. Moreover, the platform's real-time and post-activity feedback mechanisms enabled an iterative learning cycle wherein students could revise arguments based on AI critiques, fostering metacognitive awareness of reasoning quality and rhetorical effectiveness. Importantly, this interaction with AI went beyond passive reception; students critically interrogated AI-generated content, identifying logical fallacies and ethical assumptions, thus engaging in reciprocal critique that deepened their analytical rigor.

The program's emphasis on public speaking was similarly enhanced by AI integration. Through voice recording features and automated analysis, students received targeted feedback on tone, pacing, and body language, supporting the development of communication skills essential for effective oral advocacy. The curriculum's culminating activities enabled students to synthesize the skills acquired, collaborate in argument construction, and present in front of peers, while AI provided evaluative support through simulated judging and scoring. This multifaceted approach underscored the platform's capacity not only to teach procedural debate skills but to cultivate transferable competencies in critical thinking, collaboration, and digital literacy.

Debate Guru was used not just as a tool for argument generation but also as an interactive assistant, coach, and assessment platform. Teachers leveraged the customizable elements, such as editable prompts and rubric-aligned scoring, to tailor instruction to their specific curricular goals. The contrast between School A's literature-focused application and School B's curriculum-centered approach provided a rich basis for comparative analysis.

Methodology

This study incorporates both quantitative and qualitative data collection, to evaluate the impact of the Debate Guru platform on student debate skills, AI engagement, and communication growth. Qualitative information was collected through semi-structured interviews with one teacher from School A. This provides specific details into the use case in integrated classrooms. The central quantitative instrument was a pre- and post-program student survey administered to 50 participants across two schools, respectively 25 students at School A and School B. The survey consisted of four sections: Speaking Confidence & Structure, Building & Responding to Arguments, Broader Thinking & Application, and AI Tools & Feedback. Students rated their agreement with statements on a 1-7 Likert scale (1 representing "Strongly Disagree" and 7 representing "Strongly Agree"). Sample statements included "I feel confident speaking in front of others," "I can give a strong rebuttal to an argument I disagree with," and "AI helps me see how I can improve my debate skills."

Survey results were analyzed to measure change across three primary domains: student confidence in public speaking, understanding of argument structure, and overall engagement with debate. These categories were triangulated with platform usage logs (e.g., session length, feedback interactions) and classroom observation notes. To determine the statistical significance of observed changes, a one-way repeated-measures ANOVA was conducted. This analysis compared pre- and post-survey means within the same student sample across the two schools.

	Mean	Std. Deviation	N
Pre_Confidence	14.24	3.63	25
Post_Confidence	20.32	3.30	25
Pre_Arguments	18.52	2.62	25
Post_Arguments	24.48	2.82	25
Pre_Application	10.32	2.91	25
Post_Application	14.28	2.65	25
Pre_AITools	13.52	3.19	25
Post_AITools	18.64	3.37	25
Pre_Total	56.60	5.99	25
Post_Total	77.72	6.02	25

Table 1: Descriptive Statistics (School A)

Source	Measure	df	Mean Square	F	Sig.
time	Confidence	1	462.08	185.08	<.001
	Arguments	1	444.02	195.60	<.001
	Application	1	196.02	254.57	<.001
	AITools	1	327.68	236.02	<.001
	Total	1	5575.68	668.01	<.001
Error	Confidence	24	2.50		
	Arguments	24	2.27		
	Application	24	.77		
	AITools	24	1.39		
	Total	24	8.35		

Table 2: Tests of Within-Subjects Contrasts (School A)

Preliminary Findings

In both implementation contexts, pre-post survey results yield measurable increases in students' debate-related competencies, confidence, and AI engagement. In both classes, the observed changes yielded p-values less than 0.05, indicating statistical significance. Table 1-4 present the results for both schools. When the gains were disaggregated into four distinct categories, both schools restate the statistical significance ($p < 0.05$).

	Mean	Std. Deviation	N
Pre_Confidence	14.16	4.25	25
Post_Confidence	22.68	2.80	25
Pre_Arguments	18.44	2.92	25
Post_Arguments	28.28	4.25	25
Pre_Application	10.32	3.38	25
Post_Application	16.80	2.81	25
Pre_AITools	13.16	3.29	25
Post_AITools	22.12	3.68	25
Pre_Total	56.08	6.63	25
Post_Total	89.88	6.29	25

Table 3: Descriptive Statistics (School B)

However, the magnitude, consistency, and domain-specific distribution of these gains varied between the two approaches. Across all 16 survey items, School A recorded a mean improvement of +1.32 points on the 7-point Likert scale, while School B achieved a mean improvement of +2.11 points, corresponding to a relative lift of 60.04% from School A to School B. The median gains mirrored this trend, with School A at +1.16 and School B at +2.08. This indicates the merits of the full-curriculum approach distributed broadly across the dataset.

Source	Measure	df	Mean Square	F	Sig.
time	Confidence	1	907.38	155.42	<.001
	Arguments	1	1210.32	456.15	<.001
	Application	1	524.88	256.46	<.001
	AITools	1	1003.52	404.92	<.001
	Total	1	14280.50	1422.12	<.001
Error	Confidence	24	5.84		
	Arguments	24	2.65		
	Application	24	2.05		
	AITools	24	2.48		
	Total	24	10.04		

Table 4: Tests of Within-Subjects Contrasts (School B)

Item-level analysis revealed particularly strong School B advantages in three domains: (1) perspective-taking ("I understand why it's important to consider other points of view") improved by +2.24 in School B compared to +0.80 in School A (difference = +1.44); (2) evidence-gathering ("I can find good evidence to support my ideas") increased by +2.12 in School B versus +0.88 in School A (difference = +1.24); and (3) trust in AI feedback ("I trust the feedback that AI gives me about my debate skills") rose by +2.20 in School B compared to +0.92 in School A (difference = +1.28). Several additional items showed School B advantages exceeding one point, including understanding speech organization (+1.04) and curiosity about AI's role in skill development (+1.12).

In contrast, three measures showed minimal or reversed differences between the two contexts: respectful disagreement ("I can respond respectfully when someone disagrees with me") saw a slightly higher gain in School A (+2.64) than in School B (+2.48), difference = –0.16; explanation of personal opinions showed a similar pattern (School A: +1.84; School B: +1.64; difference = –0.20); and real-world application ("I can connect what I learn in debate to real-world issues") showed only a marginal difference (difference = –0.04), suggesting both contexts were similarly effective in promoting transfer of debate skills beyond the classroom. This convergence suggests that integrating debate into thematic and ethical contexts, as done in School

A, effectively supports interpersonal discourse skills and civic-minded reflection.

Clearer patterns emerged across thematic clusters. For core debate mechanics including argument clarity, evidence use, rebuttal delivery, and speech organization, School B averaged +1.91 improvement compared to School A's +1.03, an 85% greater gain. In the domain of AI engagement covering trust in feedback, understanding AI use, curiosity, and AI-supported speech planning, School B averaged +2.23 compared to School A's +1.31, a 70% advantage. By contrast, the difference narrowed to 18% in civic and interpersonal discourse skills, such as respectful disagreement, perspective-taking, and real-world application. This suggests that thematic, literature-based integration may be equally capable of cultivating these interpersonal and civic competencies.

Interview with the lead teacher in School A offers describes Debate Guru's application in subject-related contexts. The teacher used the platform alongside cultural and literary texts, such as *Stand and Deliver* and Chekhov's *The Bet*. Under the prompt about socioeconomic status and access to academic excellence, the teacher then instructed the students to formulate claims, receive immediate counterarguments, and iteratively refine their reasoning.

Engagement is the recurring theme in his narrative. The teacher noticed that students who tended to be less active began responding to the "pushbacks" provided by the AI. Debate Guru also piqued students' innate curiosity: one student "*buried herself within Debate Guru – and she went two rounds or three… with the AI*." The teacher concluded that Debate Guru "*did the most important thing to [him] as a teacher: engage intellectual curiosity*." This dynamic reframed the classroom from teacher-centered to a participatory environment where students practiced constructing evidence-based arguments. In an integrated setting, Debate Guru supported inquiry-based learning by stimulating critical discourse within a familiar digital medium.

The above findings resonate with previous studies. Subject-specific integration of AI chatbots yielded proven significance in supporting student learning. Students saw improvements in both their literary competencies and AI preparedness. The use of AI models in both cases harnessed the retention of debate skills, inspiring practical applications to real-world scenarios.

Past research corroborates the usage of large language models like ChatGPT in debate-based classrooms (de la Puente et al., 2024; Darmawansah et al., 2024; Grubaugh & Levitt, 2025). The most direct advantages of AI-supported debate training are attributed to real-time feedback that refines argumentation strategies and speech style (Grubaugh & Levitt, 2025). AI agents could open up new perspectives in building effective arguments and grasping complex subjects (de la Puente et al., 2024). The facilitation of contextually grounded learning through AI also augments the practicality of the acquired skills. Students are thus more prepared to exercise these interpersonal skills in everyday life (de la Puente et al., 2024). Furthermore, AI integration in a collaborative classroom is linked to greater growth in students' argumentative speaking skills, critical thinking, and technical fluency (Darmawansah et al., 2024). Students' co-learning with peers and AI agents deepened the quality of their arguments as well. Interactivity retained student engagement with the lessons and built up the essential interpersonal skills that didactic tutoring may not provide (Wang et al., 2025).

Descriptive findings suggest that the full-curriculum model yielded more consistent, domain-general skill gains within this pilot; however, these non-experimental results should be viewed as exploratory rather than causal. This model's advantages are particularly pronounced in skills that require procedural fluency (e.g., organizing a CEI speech, generating effective rebuttals) and in fostering sustained engagement with AI as a feedback mechanism. By contrast, the literature-adapted model produced strong but more targeted gains, with improvements concentrated in areas where debate practice was closely tied to thematic or ethical questions embedded in the literary text. The partial integration model also appeared to support slightly higher growth in interpersonal and reflective dimensions, such as respectful disagreement and articulating personal viewpoints, potentially due to the deeper contextual grounding and moral reasoning stimulated by literary content. However, without the structured, daily reinforcement of the full curriculum, gains in procedural debate skills and AI familiarity were less pronounced. Furthermore, Wang et al. (2025) emphasize that dialogic interactions facilitated by AI enhance both cognitive skills and intrinsic motivation, underscoring the value of Debate Guru's interactive features in fostering holistic debate competencies.

In sum, quantitative evidence from the pilot suggests that Debate Guru helps improve student outcomes across diverse instructional contexts, but that the scope and uniformity of gains depend on the integration model. School B's descriptive averages were approximately 0.79 points higher per item than School A, a difference that may reflect contextual factors (e.g., teacher experience or subject focus) rather than curriculum design. School A's model, while producing smaller overall gains, demonstrated that Debate Guru can still be effective when strategically embedded within disciplinary learning, especially for fostering nuanced discussion and reflective discourse. These exploratory findings suggest that different integration modes

may serve distinct instructional priorities; future work should test these relationships under controlled conditions.

Conclusions and Implications

The pilot of Debate Guru across two distinct instructional contexts demonstrates both the adaptability and the potential impact of AI-supported debate instruction in secondary education. We observed initial quantitative gains for both models in confidence, procedural debate skills, and AI engagement. These findings align with emerging research on AI-mediated learning environments, particularly those emphasizing adaptive feedback and student-centered interactivity (Grubaugh & Levitt, 2025; Darmawansah et al., 2025). Using the AI-supported collaborative argumentation models (de la Puente et al., 2024), Debate Guru's integration of automated rebuttal generation and iterative feedback loops can accelerate skill acquisition while promoting metacognitive reflection on argument quality.

We acknowledge the limitations of the pilot that warrant attention and refinement. First, its short duration of 2 weeks limits our ability to assess the long-term retention of debate skills and whether observed gains persist or compound over time. Prior work (Wang et al., 2025) suggests that sustained dialogic interaction with AI tutors yields stronger, more stable learning outcomes. A longer implementation window is necessary. Second, while both contexts showed improvement, the literature-integrated approach at School A revealed uneven gains in procedural skills, suggesting that without daily structured reinforcement, some students may plateau. Third, the current platform is primarily optimized for English-speaking learners and monolingual debate formats; as Hamza (2024) notes in the context of AI and literary creativity, cross-linguistic adaptability and culturally responsive scaffolding remain crucial for equitable adoption. Finally, while teacher feedback was generally positive, implementation still relied on some pedagogical familiarity. Additional professional development is required to prepare educators for adopting AI tools in pedagogically underpinned ways tailored to subject specificity.

These limitations suggest several avenues for future research and development. Longitudinal studies over an academic year can measure both retention and transfer of skills into other domains such as writing and civic engagement. Controlled comparative trials could evaluate Debate Guru against non-AI-supported debate programs, helping isolate the specific contribution of AI feedback mechanisms to student growth. Technical enhancements might include real-time moderation tools for in-class debates, multilingual support, and integration with writing analytics platforms to strengthen cross-modal literacy skills (Ouyang et al., 2023). The AI tutoring system can also benefit from the further engineering of fairness and ethical principles (Verheij, 2016; Holstein et al., 2019). Meticulous curation of training data, content monitoring, and periodical adjustments (Holstein et al., 2019) to the tutoring system are needed to cater to moral discussions in the program. We also plan to explore Debate Guru's performance in under-resourced educational environments, building on research showing that AI-supported civic literacy programs can help close opportunity gaps (Grubaugh & Levitt, 2025).

Debate Guru positions itself as a viable tool for fostering critical thinking, civil discourse, and persuasive communication skills emphasized in teaching with AI in K-12 settings. As AI technologies in education continue to mature, Debate Guru can benefit from sustained empirical inquiry, coupled with iterative platform refinement. The continuous research in Debate Guru is expected to show that AI tutoring systems are more than supplementary tools, but an integral component of skill-building, academic success, professional preparedness, and civic engagement in the 21st century.

Acknowledgments

We would like to extend our sincere gratitude to MadeManifest and the Mount Vernon Public School District for their support and participation in this pilot project using the Debate Guru AI platform.

References

Abbas, S.; and Sawamura, H. 2012. Argument mining based on a structured database and its usage in an intelligent tutoring environment. *Knowledge and Information Systems* 30(1): 213–246.

Chen, R.; Lee, V. R.; and Lee, M. G. 2025. A cross-sectional look at teacher reactions, worries, and professional development needs related to generative AI in an urban school district. *Education and Information Technologies* 30: 16045–16082. doi.org/10.1007/s10639-025-13350-w.

Chin, D. B.; Dohmen, I. M.; Cheng, B. H.; Oppezzo, M. A.; Chase, C. C.; and Schwartz, D. L. 2010. Preparing students for future learning with Teachable Agents. *Educational Technology Research and Development* 58(6): 649–669. doi.org/10.1007/s11423-010-9154-5.

Darmawansah, D.; Rachman, D.; Febiyani, F.; and Hwang, G.-J. 2025. ChatGPT-supported collaborative argumentation: Integrating collaboration script and argument mapping to enhance EFL students' argumentation skills. *Education and Information Technologies* 30(3): 3803–3827. doi.org/10.1007/s10639-024-12986-4.

de la Puente, M.; Torres, J.; Troncoso, A. L. B.; Meza, Y. Y. H.; and Carrascal, J. X. M. 2024. Investigating the use of ChatGPT as a tool for enhancing critical thinking and argumentation skills in international relations debates among undergraduate students. *Smart Learning Environ-*

ments 11(1): Article 55. doi.org/10.1186/s40561-024-00347-0.

Grubaugh, S.; and Levitt, G. 2025. Debate centered civics education: Using AI to teach future-ready skills and better prepare students for college, careers, and civic life. *Technium Social Sciences Journal* 70(1): 165–170. doi.org/10.47577/tssj.v70i1.12635.

Gunawan, W.; and Suhartono, D. 2023. Classifying argument components using deep learning on English dataset. *Informatica (Ljubljana)* 47(5): 63–68. doi.org/10.31449/inf.v47i5.4617.

Gurevych, I.; Wachsmuth, H.; Al Khatib, K.; and Potthast, M. 2019. Argumentation mining in user-generated web discourse. *Foundations and Trends in Information Retrieval* 13(3): 130–243.

Hamza, M. M. A. 2024. Artificial intelligence and literary creativity: Examining new frontiers in English literature. In *Proceedings of the 2024 International Conference on IoT, Communication and Automation Technology (ICICAT)*, 773–778. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE).

Holstein, K.; Wortman Vaughan, J.; Daumé III, H.; Dudik, M.; and Wallach, H. 2020. Improving fairness in machine learning systems: What do industry practitioners need? In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* New York: Association for Computing Machinery.

Kasih, E. N. E. W.; Indriyanto, K.; and Putra, A. V. L. 2024. Artificial intelligence for literature class: Trends and attitude. In *Proceedings of the 2024 10th International Conference on Education and Technology (ICET)*, 142–148. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE).

Khosravi, H.; Shibani, A.; Jovanovic, J.; Pardos, Z. A.; and Yan, L. 2025. Generative AI and learning analytics: Pushing boundaries, preserving principles. *Journal of Learning Analytics* 12(1): 1–11. doi.org/10.18608/jla.2025.8961.

Labadze, L.; Grigolia, M.; and Machaidze, L. 2023. Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education* 20(1): 56.

Mochales, R.; and Moens, M.-F. 2011. Argumentation mining. *Artificial Intelligence and Law* 19(1): 1–22. doi.org/10.1007/s10506-010-9104-x.

Ni, A.; and Cheung, A. 2023. Understanding secondary students' continuance intention to adopt AI-powered intelligent tutoring systems for English learning. *Education and Information Technologies* 28(3): 3191–3216. doi.org/10.1007/s10639-022-11305-z.

Ouyang, F.; Wu, M.; Zheng, L.; Zhang, L.; and Jiao, P. 2023. Integration of artificial intelligence performance prediction and learning analytics to improve student learning in an online engineering course. *International Journal of Educational Technology in Higher Education* 20(1): Article 4. doi.org/10.1186/s41239-022-00372-4.

Pargman, T. C.; McGrath, C.; and Milrad, M. 2025. Towards responsible AI in education: Challenges and implications for research and practice. *Computers and Education: Artificial Intelligence* 100345. doi.org/10.1016/j.caeai.2024.100345.

Singh, N.; Gunjan, V. K.; Mishra, A. K.; Mishra, R. K.; and Nawaz, N. 2022. SeisTutor: A custom-tailored intelligent tutoring system and sustainable education. *Sustainability* 14(7): 4167. doi.org/10.3390/su14074167.

Van Brummelen, J.; Heng, T.; and Tabunshchyk, V. 2021. Teaching tech to talk: K–12 conversational artificial intelligence literacy curriculum and development tools. In *Proceedings of the AAAI Conference on Artificial Intelligence* 35(17): 15655–15663. Palo Alto, CA: AAAI Press.

VanLehn, K. 2011. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist* 46(4): 197–221.

Verheij, B. 2016. Formalizing value-guided argumentation for ethical systems design. *Artificial Intelligence and Law* 24(4): 387–407. doi.org/10.1007/s10506-016-9189-y.

Wang, D.; Shan, D.; Ju, R.; Kao, B.; Zhang, C.; and Chen, G. 2025. Investigating dialogic interaction in K–12 online one-on-one mathematics tutoring using AI and sequence mining techniques. *Education and Information Technologies* 30(7): 9215–9240. doi.org/10.1007/s10639-024-13195-9.

Wijekumar, K. K.; Meyer, B. J. F.; and Lei, P. 2012. Large-scale randomized controlled trial with fourth graders using intelligent tutoring of the structure strategy to improve nonfiction reading comprehension. *Educational Technology Research and Development* 60(6): 987–1013. doi.org/10.1007/s11423-012-9263-4.

Wijekumar, K. K. 2006. Implementing web-based intelligent tutoring systems in K–12 settings: A case study on approach and challenges. *Journal of Educational Technology Systems* 35(2): 193–208. doi.org/10.2190/W3V7-4167-3510-3215.

Xing, W.; Lee, H.-S.; and Shibani, A. 2020. Identifying patterns in students' scientific argumentation: Content analysis through text mining using latent Dirichlet allocation. *Educational Technology Research and Development* 68(5): 2185–2214. doi.org/10.1007/s11423-020-09761-w.

Zhang, Y. 2024. Exploring the dynamical system modeling and evolutionary mechanisms for the integration of AI and civic education management in colleges and universities. *Applied Mathematics and Nonlinear Sciences* 9(1). doi.org/10.2478/amns-2024-0886.