

Rethinking AI from Classroom Adversary to Classroom Ally

Erik Brynjolfsson*

Matthew Merrigan†

August 2025

Introduction

We introduced RealAvatar, an AI-enabled learning companion, into Stanford’s course, *The AI Awakening*, taught by Professor Erik Brynjolfsson in the Spring Quarter of 2025. The course ran for 10 weeks, mostly comprised of MBA students from the Graduate School of Business and Computer Science / Engineering students. There were 138 students enrolled in the class.

As part of a micro-study, we utilized RealAvatar in order to encourage student-usage of AI-enabled technology. RealAvatar is a learning partner for students. Trained using course materials and readings, as well as Professor Brynjolfsson’s voice and mannerisms, the “Erik Avatar” provides students with real-time feedback, information, and technical mentorship. All of the students were asked to engage with RealAvatar over the duration of the course.

The teaching team was able to tailor the degree of combativeness and acquiescence of the Erik Avatar in response to student feedback. In general, we received feedback that the baseline version of the Erik Avatar was too acquiescent; students were concerned that RealAvatar was not providing sufficient constructive criticism. Instead, students noted that the original version of the Erik Avatar validated students’ opinions too frequently, potentially to the point of encouraging false information. The teaching team course corrected during Week 3 of the quarter, scaling down the acquiescent nature of the Erik Avatar and scaling up the combative nature. Following this, students reported increased satisfaction with RealAvatar.

We did not perform any experimentation on students enrolled in the course; all students had equal access to RealAvatar. Students could interact with the Erik Avatar in any way they wanted, and we did not limit the time spent with RealAvatar for any student at any point. Furthermore, RealAvatar was trained on the same data and given the same prompts to ask all students for each assignment. Thus, for this report, no experiment was possible.

While we elicited regular feedback in person from students, we did not collect weekly feedback surveys. We did, however, ask students to submit feedback via a quarter-end survey. 69 students completed the feedback survey.

Results

In this section, we present the results of the survey. Students, in general, engaged with the Erik Avatar across multiple weeks, and generally report high satisfaction scores. Recall that students were able to interact with the Erik Avatar via voice and / or chat. The avatar is trained on Professor Brynjolfsson’s voice and mannerisms, and both the voice and text features focus on engaging in natural conversation about a wide

*Stanford University. Email: erikb@stanford.edu

†Stanford University. Email: merrigan@stanford.edu

variety of topics.

Voice and Text Quality

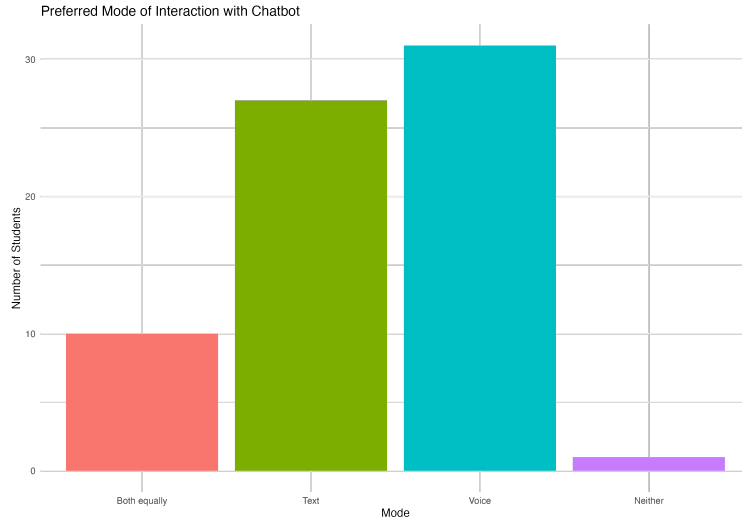


Figure 1: Preferred Interaction with RealAvatar

Figure 1 shows the preferred interaction of students with RealAvatar. Students were able to choose between text, voice, both, or neither. As the figure illustrates, the plurality of students prefer to interact with the Erik Avatar using the voice feature. Students reported that using the voice feature made their experience with RealAvatar feel more human. Engaging with voice features makes interactions more reminiscent of traditional conversations with professors in office hours or study groups. Many students describe how immersive it is to see a familiar face of their professor, and the novelty of talking to a chatbot that looks, sounds, and feels like Professor Brynjolfsson.

However, students also report that the voice feature has areas for improvement. Students mention the quality of the voice as a major concern, with many students explicitly describing the voice as “robotic” or “unnatural.” Many students often cite the voice feature as having a substantial lag time between when a student poses or answers a question and when the Erik Avatar responds. Students note that this takes them out of the immersive experience.

Students also enjoyed the text feature. Students report that the context window for text conversations was sufficiently long. Moreover, students note that the text feature was sufficiently informative and curious, asking guiding questions of many students. Some students noted that the text feature of the chatbot, however, would often engage in “circular” reasoning or conversations. This would often lead to answers that were “too long” or conversations that “deviate” from the main topic.

Figure 2 illustrates the student rating of the quality of the voice and text components of RealAvatar, respectively. Students were asked to rate the quality of RealAvatar voice and text on a scale of 1 to 5. We see that the distribution of text quality is slightly shifted to the left as compared to voice quality. Text has an overall mean of 3.22, while voice has a mean of 3.25. The two means are not significantly different.

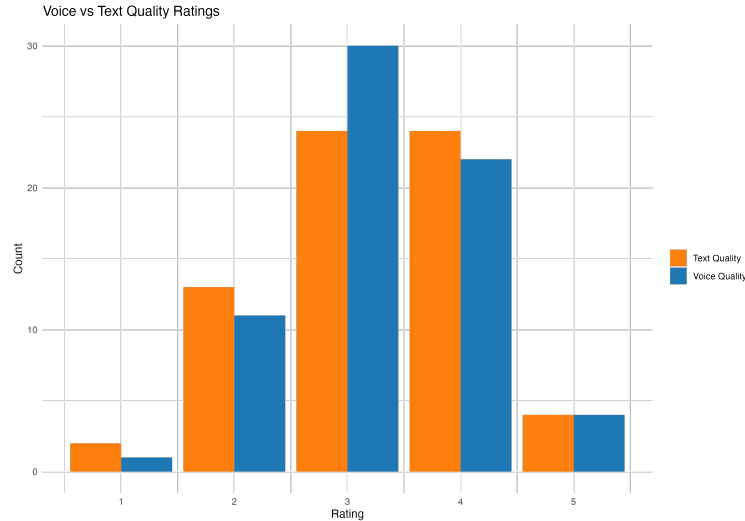


Figure 2: Student Rating of Voice and Text Quality

Comparing Assignments with and without RealAvatar

Not all assignments utilized RealAvatar to elicit student engagement. Some assignments did not include a RealAvatar component, instead simply asking students to complete and submit an assignment on their own. In the final survey, we also asked students to compare their personal experience across assignments that *did* have a RealAvatar component and assignments that *did not* have a RealAvatar component.

We are interested in whether or not students engage with course material (readings, lectures, assignments) differently when they have the additional opportunity to interact with RealAvatar. As such, we ask students to reflect on their own degree of engagement with course materials. We ask this across a variety of pedagogical areas.

Figure 3 displays a boxplot across the four areas of pedagogical interest: how carefully students thought about assignments, how much effort students put into assignments, how much original thought went into assignments, and how well students retained course material (in clockwise order starting from the top left). In each sub-figure, the red boxplot pertains to assignments in which students did not use RealAvatar, while the blue plot is for assignments in which students did use RealAvatar.

We see that, in general, the averages for all four pedagogical concerns are nearly equivalent across groups. Albeit we see marginal differences in the amount of effort that students exerted. Having said that, none of the two means are statistically different across the four groups. We note that we are asking students to make an ex-post evaluation of their own pedagogical approaches to assignments, which is a difficult task.

Several considerations may help interpret these findings:

- **Potential ceiling effects:** Median scores are already high across all conditions, leaving little room for observable gains from RealAvatar.
- **Novelty versus sustained impact:** Any initial engagement boost from using RealAvatar may fade over time, making it less salient in retrospective self-assessments.
- **Measurement constraints:** Self-reported, retrospective measures are vulnerable to recall bias and social desirability bias, potentially masking small differences.
- **Task-content interaction:** RealAvatar may have more effect on certain types of assignments (e.g., open-ended, discussion-based) than others; aggregating across tasks could obscure these effects.

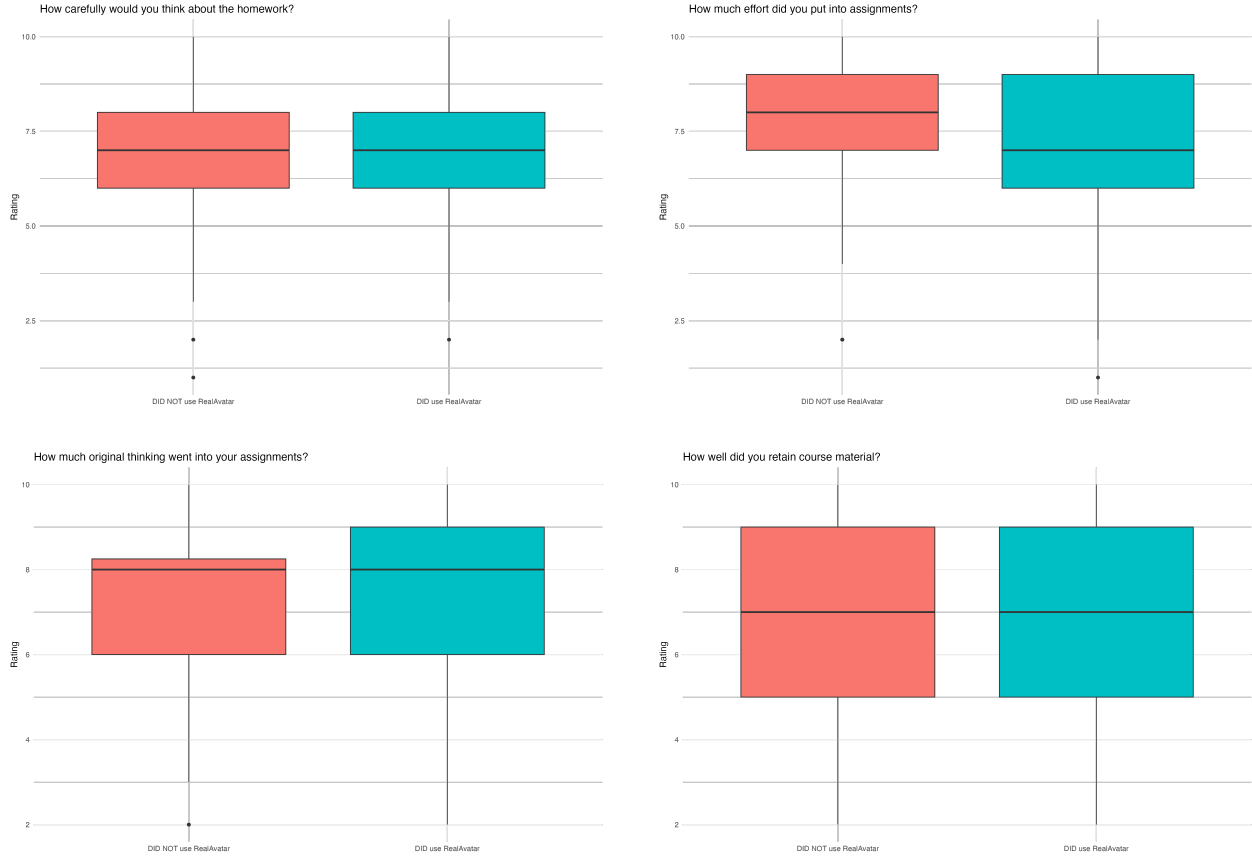


Figure 3: Student self-rating of pedagogical motivation across assignments with and without RealAvatar

- **Pedagogical interpretation:** Similar self-reports do not preclude differences in objective performance outcomes; future work could compare subjective ratings to graded performance or originality scores.

Having said that, we do not see any difference between how students reflect on their approaches to course materials when they use RealAvatar vs. when they do not. Taken together, students appear to conceptualize their understanding and approach to course material as nearly identical whether or not they use RealAvatar. If differences exist, they may be subtle, context-dependent, or detectable only through objective performance metrics rather than retrospective self-ratings.

The Personality of RealAvatar

One common point of feedback that we heard from the students pertains to the nature of RealAvatar. Many students felt that Erik Avatar was too acquiescent. Students wanted the chatbot to “agree less,” noting that RealAvatar “did not critique or challenge ideas effectively.” With that in mind, at several times throughout the quarter, the teaching staff modified the degree of acquiescence and combativeness of the Erik Avatar.

As a part of the feedback survey, we asked students to rate how acquiescent and combative they found RealAvatar over the course of the quarter. Note that students are evaluating the nature of RealAvatar over the entirety of the quarter, so these ex-post evaluations should be interpreted as an average. Figure 4 shows the distribution of student ratings of the RealAvatar’s personality. The sub-figure on the left is student ratings of Erik Avatar’s acquiescence, and the right sub-figure is that of combativeness. We see that there is larger variation in how combative students found the RealAvatar; most students found the Erik Avatar to be highly acquiescent.

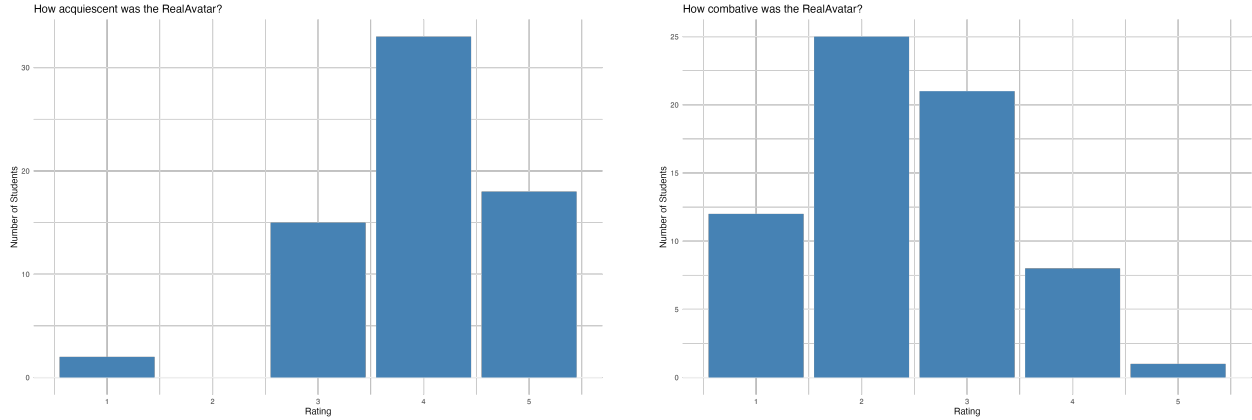


Figure 4: Student self-rating of RealAvatar personality

We wanted to additionally ask students about how RealAvatar aided their learning process across four main areas:

- Usefulness of the chatbot
- Thinking deeply about assignments
- Achieving learning goals
- How informative were the RealAvatar’s responses

We split students into low- and high-acquiescence and low- and high-combative groups. That is, we divide students into groups whether they perceived the chatbot to be more/less acquiescent or combative than the median, respectively. Of the 68 students that responded, 51 found the chatbot to be more acquiescent than the median and 55 found the chatbot to be more combative than the median.

Figure 5 illustrates how students responded to each prompt conditional on their perceived level of acquiescence. In general, the distribution of students who found RealAvatar to be more acquiescent is shifted to the right of those that had lower perceived levels of acquiescence. However, a chi-squared test reveals that there is no significant difference between the two distributions across any of the four outcomes. In general, students found the RealAvatar to be useful. Most students felt that they thought more deeply about course material, and many, particularly those in the high-acquiescence group, felt that RealAvatar helped them achieve their learning goals and had informative responses.

In contrast, Figure 6 illustrates these same four outcomes varied by perceived combativeness. Here we see the same trend as in Figure 5: those who had a higher level of perceived combativeness, in general, have more favorable views of RealAvatar. Those who found the chatbot to be more combative found RealAvatar to be more useful, thought more deeply about the course materials, achieved learning goals, and thought the chatbot had more informative responses. Notably, chi-squared tests tell us that the differences between the distributions of the two groups across all four outcomes are significantly different.

Students who find the chatbot to be more combative report meaningfully higher levels of satisfaction with RealAvatar across these outcomes. This echoes many of the open-ended feedback submitted by students describing a desire for the Avatar to push back on ideas more. These findings reflect an earnest call from students for AI-Avatars that are capable of providing critical feedback.

RealAvatar vs. Other Aids

A final question we decided to ask students pertains to whether or not RealAvatar is more helpful than other forms of academic aid. Students in many classroom settings have access to a wide variety of online

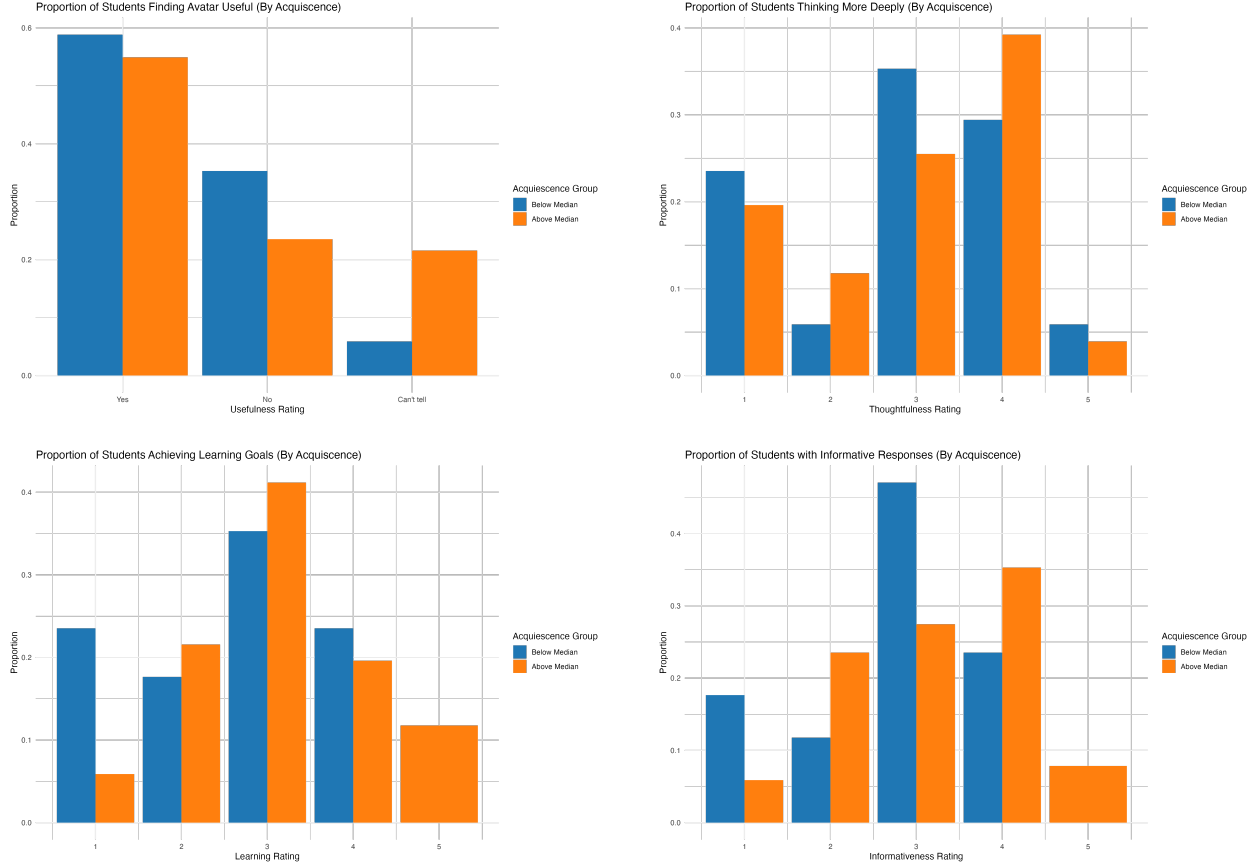


Figure 5: Perception of RealAvatar varied by acquiescence

and offline resources. These may include traditional office hours, course readings, extra materials. But in the age of LLMs, these may also include other AI-powered tools such as Chat-GPT, Claude, or Perplexity (to name a few).

To this end, we asked students how they would rate RealAvatar as compared to (1) other AI-powered tools and (2) other non-AI learning tools. Figure 7 presents these results. We see that in general, students find RealAvatar to be generally less helpful than other AI tools, but relatively helpful compared to non-AI tools. This again echoes feedback we heard from students. Students report that while RealAvatar is helpful and innovative, the shortcomings made it difficult to engage with fully as a one-stop shop for help with course materials. Students report they continued to use other LLMs to supplement RealAvatar. However, students report that it is more helpful than traditional non-AI tools to get immediate help on any questions or feedback on assignments.

To emphasize this point, Figure 8 illustrates that students, by and large, prefer interacting with other AI-based tools as compared to RealAvatar. When asked about which AI tool they most prefer to use, nearly all students report that they would rather use an application like ChatGPT or Claude as compared with RealAvatar.

Discussion and Conclusion

The introduction of RealAvatar into the classroom setting demonstrates both the promise and the challenges of integrating AI-enabled learning companions into higher education. Across multiple dimensions, students

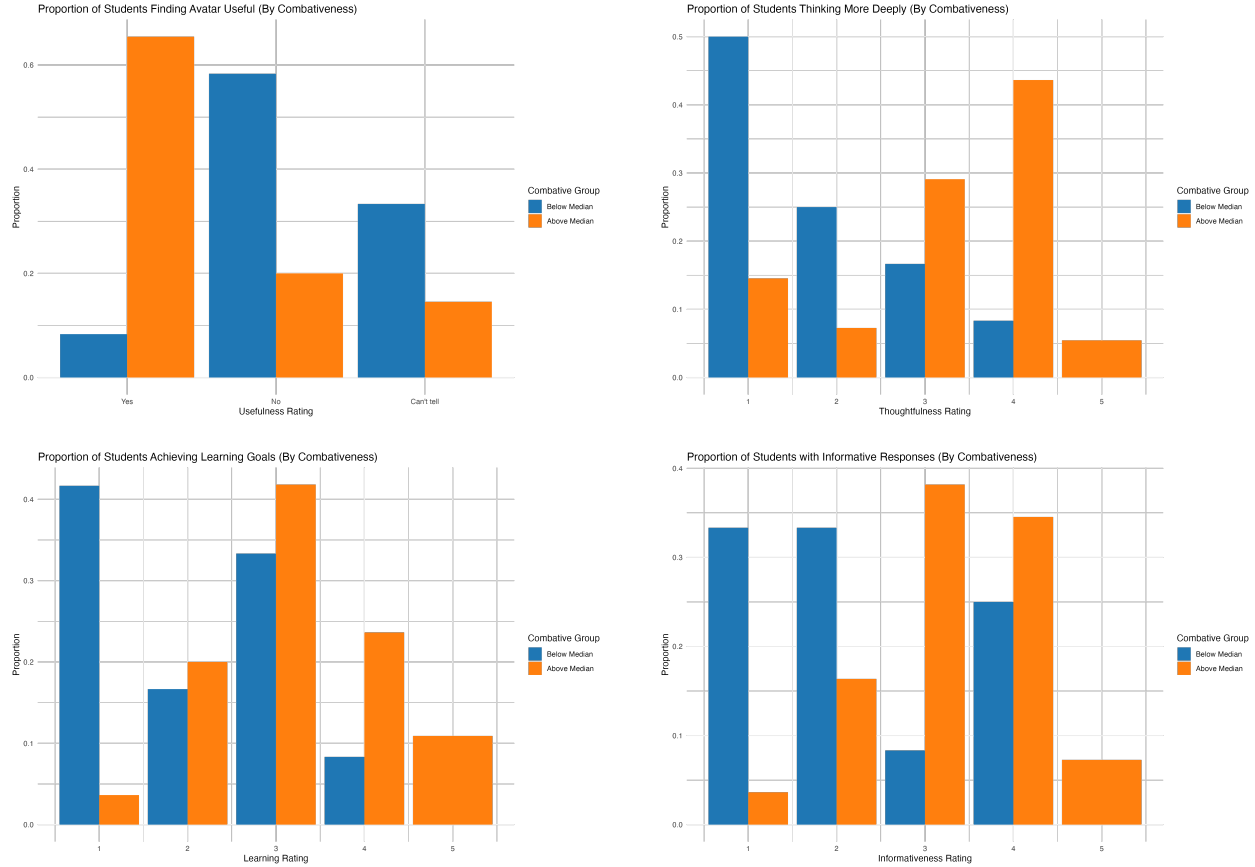


Figure 6: Perception of RealAvatar varied by combativeness

reported generally positive experiences, with many noting the novelty and human-like qualities of interacting with an avatar modeled on their professor. The voice feature, in particular, was cited as contributing to a more immersive and personal learning environment, evoking the feel of in-person office hours. At the same time, students highlighted technical limitations—such as perceived “robotic” voice quality and latency—that detracted from the experience and interrupted conversational flow.

The findings from our comparison of assignments with and without RealAvatar suggest that the tool did not substantially alter students’ self-reported engagement, effort, originality, or retention. While these null results may reflect a true absence of impact, several alternative explanations are plausible. First, high baseline ratings across all conditions indicate potential ceiling effects, leaving limited room for measurable improvement. Second, the retrospective nature of the survey is subject to recall and social desirability biases, which may obscure subtle differences. Third, RealAvatar’s influence may be context-dependent, enhancing learning in certain types of assignments—particularly those that are open-ended or discussion-based—while having less relevance for more procedural tasks.

The analysis of RealAvatar’s perceived personality offers important design insights. Students who perceived the avatar as more combative—willing to challenge ideas and push back—reported significantly higher satisfaction and stronger perceptions of usefulness, deeper thinking, goal achievement, and informativeness. In contrast, perceived acquiescence did not meaningfully affect reported outcomes. This suggests that constructive challenge, rather than constant affirmation, may be a key ingredient in effective AI-assisted learning, echoing student calls for avatars that foster critical engagement rather than passive agreement.

When compared to other learning resources, students rated RealAvatar as less helpful than other AI-powered

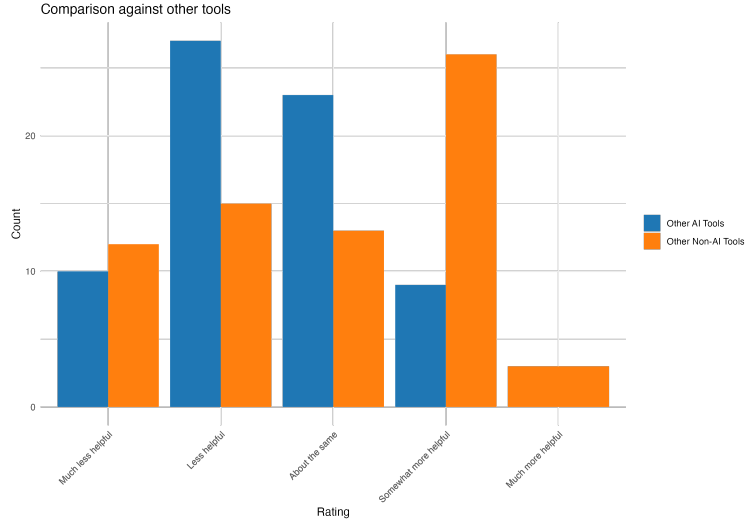


Figure 7: Comparison of RealAvatar with other learning tools

tools but more helpful than non-AI aids. This positioning highlights a tension between novelty and performance: while RealAvatar’s personalization and contextual awareness make it a unique resource, students continued to rely on other LLM-based tools for broader, faster, or more accurate responses. This implies that RealAvatar may currently serve best as a complementary, rather than primary, academic support tool.

Overall, our results underscore the importance of iterative design and evaluation in AI-assisted education. Technical refinements to improve responsiveness and voice quality, coupled with intentional calibration of conversational style toward constructive challenge, may increase both engagement and perceived value. Future research should extend beyond self-reports to include objective performance measures, longitudinal tracking of learning outcomes, and controlled experiments to isolate causal effects. As AI-enabled avatars become more prevalent, aligning their design with pedagogical best practices—particularly around fostering critical thinking—will be central to maximizing their educational impact.

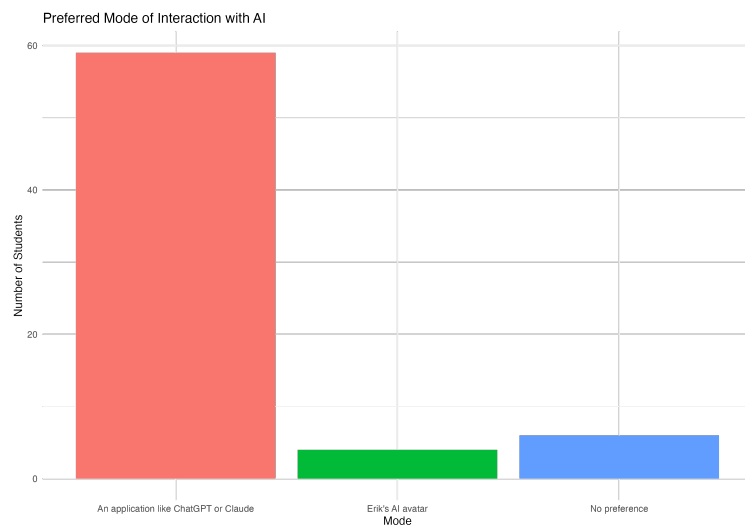


Figure 8: Comparison of RealAvatar with other AI tools