SOCIAL INNOVATIONS Journal
Vol. 30 (2025): The Future of AI in Education: Innovations and Insights

Scaffolding Creativity or Homogenizing Outputs? GPT-based Virtual TA's Impact on Student Ideation in Graduate Education

By: Nidhi Nasiar[1], Mamta Shah[1,2], Ryan S. Baker[1], Maciej Pankiewicz[1]

[1]Graduate School of Education, University of Pennsylvania
[2]Elsevier

Keywords: Generative AI, virtual TA, ChatGPT, creativity, ideation, graduate education

Abstract

The rapid integration of large language models in education presents both opportunities and challenges for creative ideation in graduate learning. This quasi-experimental study examines JeepyTA, a GPT-powered teaching assistant and discussion agent, and its impact on students' educational applications of a commercial video game. Comparing 2023 (without JeepyTA) and 2024 (with JeepyTA), findings show that JeepyTA significantly increased the production of fully formed ideas, suggesting its role in scaffolding ideation. However, its impact on idea quality depended on prior domain knowledge, with familiarity moderating effectiveness. While JeepyTA boosted idea quantity, it did not simply generate superficial responses. Notably, JeepyTA-driven ideation led to a higher volume of contributions but with creative homogeneity, whereas student-driven ideation produced more unique, context-rich ideas. These findings highlight the importance of prior knowledge in maximizing the benefits of generative AI and offer insights into integrating AI tools to support, rather than substitute, independent creative thinking in graduate education.

Introduction

The swift integration of large language models (LLMs) through tools like ChatGPT into educational settings has both unveiled promising opportunities and sparked significant concerns (Kasneci et al., 2023; Sok & Heng, 2024). Studies highlight generative AI's potential to transform learning, particularly in higher education, by scaffolding complex cognitive tasks and guiding learners (Lee et al., 2024).

Potential and Challenges of LLMs in Higher Education

One of the common recent uses of LLM technologies is AI discussion agents and chatbots. For instance, Imundo et al. (2024) state that domain experts can leverage AI chatbots as collaborative tools, enhancing their capacity to generate, refine, and iterate on complex ideas. By offloading lower-level cognitive tasks such as drafting and information retrieval, these technologies can potentially augment professional workflows and serve as training tools for novices developing expertise. However, non-expert users might struggle to effectively utilize these AI tools for higher-order cognitive tasks, primarily due to insufficient domain knowledge (Imundo et al., 2024). Fan et al. (2024) also note the risk of "metacognitive laziness," where learners can become overly reliant on AI-generated suggestions, potentially reducing their motivation and

capacity for deep, self-regulated learning. Research suggests that while generative AI tools might improve immediate task outcomes, they may not necessarily enhance intrinsic motivation or facilitate meaningful knowledge transfer (Fan et al., 2024), and that over-reliance on AI chatbots could inadvertently hinder expertise development by reducing opportunities for deliberate practice and disrupting the organic exchange of ideas within knowledge communities (Imundo et al., 2024). There has also been considerable recent interest in integrating LLMs into course discussion forums as simulated teaching assistants (TAs) (Laney & Dewan, 2024; Liu et al., in press), producing more rapid responses to student questions and initial feedback on assignments. Findings from Liu and M'hiri (2024) highlight the potential of virtual TAs to provide more detailed responses than humans; however, they emphasize the need for human supervision, as responses from virtual TAs can sometimes overwhelm beginners.

AI Discussion Agents and Creativity

ChatGPT—an LLM-based chatbot tool—has shown to emulate and support human creativity in controlled evaluations, suggesting a novel paradigm for human-AI co-creativity (Rafner et al., 2023). However, others have raised concerns about the risks of homogenized content and diminished thematic diversity in ideas when using ChatGPT (Liu et al., 2024). There is not yet sufficient evidence on which of the trends is most salient in the context of GPT-assisted creative ideation in graduate education. One of the key definitions of creativity, which we adopt in this study, is Henriksen et al.'s (2015) definition, which defines creativity, and by extension, artifacts resulting from creative ideation, as those which include elements of novelty, effectiveness and wholeness (positive aesthetic qualities for a specific context) (Henriksen, Mishra & Mehta, 2015). Novel idea generation represents a critical component of creativity, a fundamental goal of graduate education aimed at cultivating independent, innovative thinkers. As learners increasingly engage with AI-assisted technologies in their advanced studies, understanding the technology's impact on their creative output becomes paramount. In particular, there is a need for exploration of how AI influences both convergent (focused, problem-solving) and divergent (exploratory, open-ended) creativity.

Study Purpose and Research Questions

We examine the implications of a generative AI discussion forum agent on students' creative ideation around the educational application of a video game in a graduate-level course. This study examines an existing dataset on JeepyTA, a GPT-powered virtual teaching assistant and discussion agent, to explore the balance between its benefits and risks. While Shah et al. (2024) identified unique interaction patterns among graduate students using JeepyTA for play journal assignments, they did not assess its impact on idea quality, quantity, or creative engagement. By comparing student work over two years in a quasi-experimental design, this study investigates how JeepyTA influences students' ability to use AI as a tool for creative ideation. To that end, this paper addresses the following research questions using data from a 2023 cohort acting as the control group (without JeepyTA) and a 2024 cohort acting as the treatment group (with JeepyTA):

1. RQ1: What are the differences in ideas generated between the two cohorts (2023 and 2024), and do these effects persist when accounting for prior experience with Minecraft?
2. RQ2: What are the differences in the use of literature, anecdotes, and external sources in the ideas proposed by the two cohorts?
3. RQ3: What are the emerging themes in students' interaction with JeepyTA (2024) for ideation?

JeepyTA

The Penn Center for Learning Analytics at the University of Pennsylvania developed and launched JeepyTA in the Fall of 2023 (Liu et al., in press). JeepyTA, a combination of "GPT" (from the OpenAI language model it is based on) and "TA" (its role as a teaching assistant), leverages the dialogue feature of the pre-trained language model GPT, allowing it to engage in conversations, respond to questions, and provide assignment feedback. Currently deployed across 16 sessions of 14 courses at three institutions as of Spring 2025, JeepyTA customizes its responses using course-specific materials provided by instructors, including syllabi, textbooks, reading materials, past feedback examples, and previous instructor forum responses. The system employs embeddings and semantic search to ensure accurate, contextually appropriate responses. For selected forum topics, instructors can opt to review and approve JeepyTA's responses before they become visible to students. Liu et al. (in press) show promise of JeepyTA as a virtual TA in disseminating course information and facilitating communication. This study further evaluates the efficacy of JeepyTA specifically as a virtual discussion forum agent capable of thought-partnering and as a consultant for brainstorming and generating new ideas.

Study Context

The data for this study was collected from a graduate-level course on "Video Games and Virtual Worlds as Sites for Learning and Engagement". The course explores the multifaceted field of game studies through an interdisciplinary lens, examining key concepts including theories of play and learning, game evolution, design principles, and the educational potential of both digital and analog games. The curriculum covers pedagogical frameworks, educational outcomes, and game-based assessment and analytics. Students engage with various technologies, from crafting custom controllers with MakeyMakey to exploring augmented and virtual reality applications. Assessment combines individual work (including play journals, game design documentation, and a term paper) with collaborative projects focused on knowledge building through discussion and annotation. Students also develop practical game design skills by creating playable prototypes using platforms such as Scratch and OctoStudio. For a comprehensive course overview, refer to Kafai and Shah (2021). The course has been taught by the second author during Spring semesters since 2020, with JeepyTA integrated into the curriculum in 2024.

Application of JeepyTA in a graduate-level course

JeepyTA, powered by OpenAI's GPT-3.5 model (gpt-3.5-turbo-16k-0613) within this course, was integrated into the course's discussion forum to support three key pedagogical functions: answering course-related queries, providing discussion summaries for the instructor, and

engaging with students on two play journal assignments. A play journal is a structured reflective assignment in which students document and analyze their gameplay experiences to critically engage with a video game by examining their design, narrative, and educational potential. These journals provide a way for students to develop knowledge of a game through both direct and vicarious experiences, establishing at least some baseline equivalence of knowledge. This foundation enables students to propose creative and meaningful ways to repurpose games for learning in classroom settings.

The primary modification in the course from 2023 to 2024 was giving participants access to JeepyTA for consulting innovative educational applications of Minecraft, a popular commercial entertainment video game that has often been adopted for educational purposes (Nkadimeng & Ankiewicz, 2022). Students were urged to "thought-partner" with JeepyTA, though without specific requirements for how to utilize its recommendations. Students used prompts to offer JeepyTA context about their gameplay experiences and insights gained from both direct play and secondary sources. Specific prompts that students used to interact with JeepyTA for the assignment are given in detail in Shah et al. (2024). This article focuses on the same play journal assignment, which focused on proposed educational applications of Minecraft.

Methods

Data Source

42 play journals were collected from participants in the 2023 (without JeepyTA) cohort (N=20) and the 2024 (with JeepyTA) cohort (N=22). Data was extracted from the Canvas repositories for the course in Summer 2024 at a university in northeastern America.

Coding the Data

Within this paper, we analyze students' journals proposing the use of Minecraft in an educational context. The second author, who also served as the instructor for the course, coded students' prior experience with Minecraft into five categories:

1. Has not played before
2. Has played before, time/extent undetermined
3. Has played for over 10 years
4. Cannot determine prior experience with the game

These categorizations provided a basis for understanding participants' familiarity with Minecraft as they engaged with the assignment. Next, we coded the journals themselves across multiple dimensions. The journals were anonymized before coding. For coding purposes, only the sections of the journals that explicitly described the educational application of Minecraft were analyzed. The coders developed a coding scheme from the journals using a thematic analysis approach to categorize and capture key themes and distinct ideas (Braun & Clark, 2006). The coding scheme was finalized after multiple iterations of assigning codes, reflecting, and refining the scheme. Table 1 lists the final codes with the categories, definitions, and examples.

Table 1. Coding Scheme

Code	Category	Definition	Examples
Total Ideas	Quantity	Count of distinct ideas.	3
Idea Theme	Quality	The main topics or themes of ideas presented.	Idea 1: Virtual Field Trips Idea 2: Programming and Coding Idea 3: Collaboration and Problem Solving
Types of Ideas: a) Trivial	Quality	Superficial ideas lacking depth and concrete examples. Ideas that offer little to no details about their application in real-life settings, with no relationship to game mechanics or domain knowledge.	"... I think the game has a very wide range of educational uses. It can be used as a teaching tool in a wide range of formal and informal settings. For example, in courses on design, the game can be used as a simulator for students to create and design their own works in the virtual world to develop their creativity and design-thinking skills."
Types of Ideas: b) Promising	Quality	Ideas showing potential, with more details than trivial, but requiring further development. Ideas that offer some detail about their application in real-life setting, game mechanics, and domain knowledge, but still lack concrete examples.	"[Minecraft] can also be used in higher education for city planning and architecture majors to explore innovative ideas. Students can learn about different materials and landscapes by engaging in the gameplay, practice their ability to construct three-dimensional structures, and improve their creativity in a safe space. However, due to the steep learning curve and time-consuming nature, I think it is more appropriate for students to explore Minecraft as assignments and projects outside of class than learning in class."
Types of Ideas: c) Fully Formed (FF)	Quality	Comprehensive and detailed ideas with concrete examples. Ideas that offer details about their application in real-life settings, game mechanics, domain knowledge, and specific examples.	"After having a rich conversation with JeepyTA, I found that many characteristics of Minecraft made this game a complement that can be used with formal learning materials in educational settings. One agreement that me and JeepyTA have in common on implementing Minecraft in Education is that Minecraft is a perfect platform for children to work collaboratively and develop their creativity and teamwork skill. The construction feature of Minecraft is suitable for supporting these developments. Teachers can assign students a specific theme, such as

			constructing a library. This project definitely cannot be done by a single student, so they need to work together to do this project. They then can take on different roles. Some students can be designers. They need to estimate the area of the library, sketch the inside layout, and think about what should be included. Some students can be resource managers. They need to think about what materials will be used and categorize different resources. Other students can be constructors. They will build different objects. During the process, they need to negotiate with each other, make improvements on the plan, and then implement the changes. Since teachers do not provide a set model for them to follow and replicate, students will creatively design the library on their own. Any difficulties encountered during the process will also become a solid experience for students to practice decision-making and problem-solving skills. Students who want to be the leaders of the group will also learn leadership skills."
Anecdotal	Source	Mention prior experience with the game or relevant domain knowledge, and draw from personal experience.	"By synthesizing my personal playing experience and the reading materials, I think there are several educational uses for this game…"
External sources	Source	Mention the use of external internet sources like YouTube or Google for coming up with ideas for the assignment.	"...YouTube has a large collection of videos of redstone creations, including a working platforming game made entirely out of redstone…"
Reference to literature	Source	Reference academic articles to support their ideas.	"...In conclusion, Minecraft exemplifies the potential of video games as versatile educational tools that can bridge historical knowledge with innovative technology. By engaging students in both constructing and experiencing virtual worlds, Minecraft facilitates a deeper understanding of curriculum content across disciplines. This approach not only aligns with Kafai's framework of connected gaming but also leverages the instructionist and constructionist

			paradigms to foster a comprehensive, interactive learning environment (Dezuanni, 2018)."
Student-driven ideation:	Direction of ideation	Ideas generated initially by the student. JeepyTA, if used, added details to the idea but didn't produce new ideas.	"I think Minecraft can be used in many educational settings. For example, Minecraft can be used to develop students' programming skills…."
JeepyTA-driven ideation	Direction of ideation	Ideas generated initially by JeepyTA as a result of the student's prompt. New ideas in this case came from JeepyTA and not the student.	"My conversation with JeepyTA helped me find an interesting educational use case of Minecraft, although it's probably far from novel, if not a predictable choice…."

Both coders independently coded all 42 play journals using the predefined coding scheme. They resolved coding differences through social moderation. Inter-rater reliability was calculated for multiple variables, with Cohen's kappa coefficients ranging from moderate to very high agreement. Perfect agreement ($\kappa = 1.00$) was achieved for coding references to literature, while citations of external sources like YouTube or Google showed excellent reliability ($\kappa = 0.90$). The coding of JeepyTA-driven ideation also demonstrated strong reliability ($\kappa = 0.91$), as did student-driven ideation ($\kappa = 0.814$). Student-driven and JeepyTA-driven ideation codes were only applied to the 2024 cohort, as they were relevant only for the condition where students had access to JeepyTA. Identification of fully formed ideas (FF) ($\kappa = 0.802$) also showed strong agreement. Moderate to substantial agreement was achieved for counting distinct ideas ($\kappa = 0.77$), identifying trivial ideas ($\kappa = 0.74$), and recognizing promising ideas ($\kappa = 0.63$). Initially, the coding of anecdotal and domain knowledge showed lower reliability ($\kappa = 0.56$); however, after revising the coding scheme and clarifying the scheme by dropping the "domain knowledge" category due to its implicit nature, the researchers achieved full agreement through social moderation.

Data Analysis

We employed a mixed-methods approach, using quantitative statistical analyses as well as a preliminary thematic analysis to get a comprehensive understanding of JeepyTA's impact on students' idea generation. To answer RQ1, first, descriptive statistics (frequencies, means, and standard deviations) of idea types (trivial, promising, fully formed) were compared across the 2023 (control, N=20) and 2024 (JeepyTA, N=22) cohorts. Next, to evaluate JeepyTA's influence on the type of ideas generated, three ANCOVAs were conducted (one per idea type), with cohort as the independent variable and total ideas as a covariate. This controlled for individual productivity differences among students. A follow-up ANCOVA added prior Minecraft experience as a second covariate to assess whether JeepyTA's effects persisted after accounting for domain-specific expertise.

RQ2 was investigated by performing chi-square tests to examine the association between cohort year and the inclusion of different sources (e.g., anecdotal evidence, external sources, literature references) in the assignment. A chi-square test was used here since these categories were relatively rare, and analyzing them as present or absent was therefore more appropriate than examining their frequency.

For RQ3, researchers conducted a preliminary thematic analysis (Braun and Clark, 2006) to identify patterns in ways of engaging with JeepyTA. These emerging themes enabled a deeper understanding of student interaction in the 2024 cohort and the outcome of ideas with JeepyTA.

Results

The findings are organized to first compare the quantity and types of ideas generated between the 2023 and 2024 cohorts (RQ1), followed by an analysis of how students incorporated external sources, literature, and anecdotal evidence in their responses (RQ2). Finally, we explore emerging themes in student interactions with JeepyTA in 2024 to understand its role in fostering creative ideation (RQ3).

RQ1 Differences in ideas generated across cohorts

Table 2 presents descriptive statistics for the key codes for the two cohort years 2023 and 2024. The table gives a frequency count, mean, and SD of the total number of ideas and their breakdown into the three types of ideas. Table 3 summarizes the ANCOVA results from controlling for total ideas and controlling for total ideas along with previous Minecraft experience.

Table 2. Comparison of Descriptive Statistics for Codes Across Cohorts

Codes	**2023 (Control)** N=20	**2023 (Control)** Mean (SD)	**2024 (JeepyTA)** N=22	**2024 (JeepyTA)** Mean (SD)
Total Ideas	48	2.40 (1.43)	67	3.05 (2.3)
Fully formed (FF)	7	0.35 (0.49)	18	0.82 (0.96)
Promising	17	0.85 (0.93)	11	0.50 (1.06)
Trivial	24	1.20 (1.36)	38	1.73 (2.53)
Anecdotal evidence	15	0.75 (0.44)	8	0.36 (0.49)
External source	6	0.3 (0.47)	1	0.045 (0.21)
Reference to literature	7	0.35 (0.49)	4	0.182(0.39)

Table 3. ANCOVA Summary Table

Idea Type	Covariate	Factor	F-value	p-value
Fully formed (FF)	Total Ideas	**Cohort**	4.652	**0.037**
		Total Ideas	1.534	0.223
	Total Ideas and Minecraft Experience	**Cohort**	4.83	**0.034**
		Total Ideas	1.407	0.243
		Cannot determine prior experience with the game	0.034	0.855
		Has played before	0.069	0.795
		Has played before for over 10 years	0.641	0.428
Promising	Total Ideas	Cohort	2.504	0.122
		Total Ideas	5.721	**0.022**
	Total Ideas and Minecraft Experience	**Cohort**	4.5	**0.041**
		Total Ideas	8.2	**0.007**
		Cannot determine prior experience with the game	0.585	0.449
		Has played before	5.388	**0.026**
		Has played before for over 10 years	0.025	0.875
Trivial	Total Ideas	Cohort	0.019	0.891
		Total Ideas	92.613	**<0.001**
	Total Ideas and Minecraft Experience	Cohort	0.011	0.917
		Total Ideas	88.066	**<0.001**
		Cannot determine prior experience with the game	0.217	0.644
		Has played before	2.637	0.113
		Has played before for over 10 years	0.194	0.663

Fully Formed (FF) ideas: Descriptive results show that the 2023 cohort averaged 0.35 (SD = 0.49) FF codes per student, which is lower than the 0.82 (SD = 0.96) observed in the 2024 cohort. In the context of FF ideas, a significant cohort effect was observed when controlling for total ideas ($F(1, 39) = 4.652$, $p = 0.037$), but the effect of total ideas was not significant ($F(1, 39) = 1.534$, $p = 0.223$). This suggests that although access to JeepyTA (2024) had an effect on the number of FF ideas produced, the total ideas generated did not (see Table 3).

When Minecraft experience was included in the model, the cohort effect remained significant ($F(1, 39) = 4.83$, $p = 0.034$), but again, total ideas did not significantly affect outcomes ($F(1, 39) = 1.407$, $p = 0.243$). This shows that access to JeepyTA (2024) had an effect on the FF ideas produced, but the total ideas generated did not. The Minecraft experience variables "Cannot determine prior experience with the game" ($F(1, 39) = 0.034$, $p = 0.855$), "Has played before" ($F(1, 39) = 0.069$, $p = 0.795$), and "Has played before for over 10 years" ($F(1, 39) = 0.641$, $p = 0.428$) showed no significant impact compared to the baseline of no experience.

Promising ideas: The 2023 cohort (no JeepyTA) reported an average of 0.85 (SD = 0.93) Promising codes per student, compared to 0.50 (SD = 1.06) in the 2024 JeepyTA cohort. For Promising ideas, ANCOVA analysis without considering Minecraft experience (but only controlling for total ideas) indicated no significant cohort effects ($F(1, 39) = 2.504$, $p = 0.122$) but a significant influence of total ideas ($F(1, 39) = 5.721$, $p = 0.022$). This suggests that students who generated more ideas overall were more likely to produce ideas categorized as Promising but no effects of JeepyTA were found (see above in table 3).

However, when Minecraft experience was included in the model, the cohort effect became significant ($F(1, 39) = 4.5$, $p = 0.041$) as well as the effect of total ideas ($F(1, 39) = 8.2$, $p = 0.007$). Hence, when students' past Minecraft experiences are also taken into account, students who had access to JeepyTA (2024 cohort) generated a larger number of promising ideas than the group without JeepyTA (2023). Among the Minecraft experiences, "Has played before" significantly influenced promising ideas production ($F(1, 39) = 5.388$, $p = 0.026$) relative to those with no prior experience, whereas "Cannot determine prior experience with the game" ($F(1, 39) = 0.585$, $p = 0.449$) and "Has played before for over 10 years" ($F(1, 39) = 0.025$, $p = 0.875$) did not. Overall, having prior experience with Minecraft influenced the production of Promising ideas.

Trivial ideas: Descriptives from Table 2 show that the 2023 cohort (no JeepyTA) had an average of 1.20 (SD = 1.36) Trivial ideas per student, compared to 1.73 (SD = 2.53) in the 2024 JeepyTA cohort. However, the ANCOVA analysis for Trivial ideas revealed no significant differences between cohort years when only controlling for total ideas ($F(1, 39) = 0.019$, $p = 0.891$), but a highly significant effect was found for the total number of ideas ($F(1, 39) = 92.613$, $p < 0.0001$). This suggests that students who generated more ideas overall were more likely to produce ideas categorized as Trivial but no effects of JeepyTA were observed (see above in table 3). When also considering Minecraft experience, results again showed no significant differences between cohorts ($F(1, 39) = 0.011$, $p = 0.917$) with a continued significant effect of total ideas on trivial ideas produced ($F(1, 39) = 88.066$, $p < 0.0001$). Specific Minecraft experiences such as "Cannot determine prior experience with the game" ($F(1, 39) = 0.217$, $p = 0.644$), "Has played before" ($F(1, 39) = 2.637$, $p = 0.113$), and "Has played before for over 10 years" ($F(1, 39) = 0.194$, $p = 0.663$) did not significantly impact Trivial idea production compared to the baseline group of those who had not played before. Overall, no significant effect of JeepyTA was observed on students' generation of Trivial ideas for the assignment.

RQ2 Differences in the use of sources (literature, anecdotes, and external sources) in the ideas generated across cohorts

Results from the chi-square test show a significant association between the cohort year and the reference to anecdotal evidence in the student journal ($\chi^2 = 4.85$, $df = 1$, $p = 0.03$). More students used anecdotes to support their ideas for suggesting educational uses of Minecraft in the year without JeepyTA (2023) than when they had access to the virtual TA. The chi-square test results were not significant, i.e., showing no differences between cohorts, for references to external sources like YouTube videos ($\chi^2 = 3.23$, $df = 1$, $p = 0.07$), or to prior research literature ($\chi^2 = 0.79$, $df = 1$, $p = 0.38$).

RQ3 Emerging themes from JeepyTA interaction in the 2024 cohort

An analysis of students' interactions with JeepyTA for the 2024 cohort highlighted two emerging themes. These themes provide insights into the mechanisms of idea generation with an LLM-based TA and the type of ideas produced when the interaction is either student-led or JeepyTA-led.

Differences between JeepyTA-driven and student-driven ideas

Table 4 highlights the differences in the total quantity and distribution by the quality of ideas (trivial, promising, FF) when the approach for ideation was student-driven vs. JeepyTA-driven.

Table 4. JeepyTA-Driven and Student-Driven Ideation in 2024

Category	Total Ideas	N of students	Avg ideas per student	Trivial	Promising	FF
JeepyTA-driven ideation	50	18	2.78	36	8	6
Student-driven ideation	17	10	1.7	2	3	12

Table 4 shows that a total of 50 ideas were JeepyTA-led and 17 were student-led. Some students overlapped in both categories because they contributed ideas classified in each group. Specifically, JeepyTA-led ideas came from 18 students (averaging 2.78 ideas per student) and included 36 Trivial, 8 Promising, and 6 Fully Formed (FF). Meanwhile, student-led ideas were generated by 10 students (averaging 1.7 ideas per student) and included 12 FF, three promising, and two trivial. Overall, JeepyTA-driven ideation outnumbered student-led ideation by producing 2.78 ideas per student compared to 1.70. However, 72% of JeepyTA-led ideas (36 out of 50) were labeled Trivial, whereas 70.6% of student-led ideas (12 out of 17) were categorized as FF, demonstrating greater depth and specificity. Although JeepyTA-driven ideation resulted in the production of more ideas overall, student-driven ideas appeared to have greater depth and quality.

Because of the small sample size, no statistical tests were conducted, so these observations warrant further inquiry in future cohorts. Additionally, as noted earlier, since JeepyTA helps generate a quantity of ideas, students could be supported in drawing on their knowledge of a game, domain, and insights from resources to develop Fully Formed (FF) ideas, especially from those that were initially JeepyTA-driven.

Homogeneity of ideas

A thematic overlap in ideas was observed for JeepyTA-driven ideas. Among 18 students who proposed JeepyTA-driven ideas (see table 4), the virtual TA's suggestions included repetitive themes. Upon analyzing the frequency of each unique theme suggested by JeepyTA across all students, the three most frequent themes of ideas identified were as follows: "Teamwork and

collaboration" (included in responses of 10 students), "Creativity with storytelling" (included in responses of 8 students), and "Virtual field trips" (included in responses of 6 students). In contrast, student-led ideas exhibited greater diversity, spanning unique applications, such as "Ethics of animal treatment," "Ecological concepts," and "Hyperbolic geometry in world design." The most frequent themes of student-driven ideas included architecture (included in the response of 3 students) and math (included in the response of 2 students). Upon analyzing the full set of responses for the frequent themes for student-driven ideation, it was observed that even the students with the same themes in their ideas included nuances and details that differentiated these ideas. This is relevant especially since a majority of student-driven ideation included FF ideas (12 FF; 3 promising; 5 trivial; see table 4), which were labeled as such due to their detailed explanations and concrete examples. Contrastingly, for JeepyTA-driven ideation, the majority of ideas were trivial (6 FF; 8 promising; 36 trivial; see table 4), signaling that most of these ideas did not include sufficient details and were the same across participants. Once again, these observations warrant further investigation in future cohorts. As a takeaway, students could be encouraged to “add their voice” to the ideas they generate in collaboration with JeepyTA instead of solely relying on the virtual TA’s recommendations.

Discussion and Conclusion

This quasi-experimental study provides evidence of the impact of JeepyTA, a GPT-powered virtual teaching assistant, on students' creative ideation in a graduate-level course. The findings indicate a significant effect of JeepyTA on the production of fully formed (FF) ideas, with students in the 2024 cohort producing a higher number of FF ideas compared to the control group (2023 cohort). This suggests that JeepyTA facilitates the development of well-articulated, detailed ideas, aligning with prior research on AI-supported scaffolding in education (Reiser, 2004; Roll & Ruth, 2016). For promising ideas, results were initially inconclusive, but when controlling for prior Minecraft experience, a significant difference between cohorts emerged. This finding suggests that JeepyTA's influence on promising ideas is contingent upon prior domain experience, indicating a nuanced interaction between AI assistance and students' pre-existing knowledge. These results underscore the role of prior knowledge in idea generation, suggesting that JeepyTA may be most effective when students possess foundational expertise in the subject matter, allowing them to refine and build upon AI-generated suggestions more effectively, a direction pointed out by Imundo et al (2024). In contrast, there was no observed effect for trivial ideas, with total idea production being the strongest predictor of trivial outputs. This pattern highlights that while JeepyTA aids in structuring and refining higher-quality ideas, it does not contribute to a proliferation of superficial student responses.

The preliminary thematic analysis further reveals that JeepyTA-driven ideation resulted in a greater quantity of ideas, but these were often categorized as trivial, lacking depth, and specificity. In contrast, student-driven ideation led to more fully formed ideas with detailed descriptions and concrete applications. This dichotomy highlights a key consideration in AI-supported learning: while generative AI can enhance idea fluency, it may simultaneously contribute to creative homogeneity by promoting common themes across multiple students (Anderson et al., 2024; Wenger & Kenett, 2025). However, prior knowledge plays a critical role in mitigating this effect, as students with deeper expertise were able to generate more unique and

nuanced ideas, even when leveraging AI assistance. Moreover, the analysis of supporting sources revealed that students in the 2023 cohort were more likely to incorporate anecdotal evidence into their assignments than those in the 2024 cohort. This suggests that reliance on JeepyTA may reduce students' use of personal experiences as a basis for idea generation. However, references to external sources and academic literature remained consistent across both cohorts, indicating that JeepyTA did not diminish engagement with external research.

A potential limitation of this study is self-selection bias, as students who engaged more with JeepyTA may have already been predisposed to AI-assisted learning. Additionally, the tool's presence may have influenced behaviors beyond ideation, such as increasing time spent on tasks or reliance on AI-generated suggestions. Future research should explore these behavioral shifts to understand how the presence of generative AI alters student engagement and creative processes.

Implications for practice

The findings from this study offer key takeaways for integrating AI-powered teaching assistants into educational settings. To maximize JeepyTA's benefits for creative assignments, educators should encourage students to use the tool as a brainstorming aid rather than as a definitive source of ideas. AI-generated suggestions should be treated as starting points, prompting students to refine and expand upon them with their own insights and knowledge. This approach can help mitigate the risk of creative homogeneity while fostering deeper engagement with course material. Drawing from self-reflections and prior experiences remains an important element of creative ideation, and the reduced use of anecdotal evidence in the 2024 cohort suggests a need to reinforce this practice. Instructors can incorporate assignment prompts that explicitly require students to integrate personal experiences alongside AI-generated recommendations. By doing so, they can help students balance computational efficiency with authentic self-expression.

The study also highlights the importance of domain knowledge in shaping AI-assisted creativity. Students with prior familiarity with Minecraft were better positioned to generate meaningful and well-developed ideas, demonstrating the interplay between prior knowledge and AI support. To enhance students' ability to engage productively with AI tools, curriculum design should include opportunities for foundational knowledge-building before introducing AI-assisted ideation tasks. Furthermore, the repetition of common themes in JeepyTA-generated ideas suggests a need for strategies to counteract homogenization. Instructors might diversify discussion prompts, encourage students to critique and elaborate on AI-generated ideas, and introduce peer review processes to stimulate unique perspectives. By promoting critical engagement, students can avoid creative convergence, move beyond surface-level suggestions, and purposefully develop ideas that are both innovative and contextually rich.

This study shows an initial promise for JeepyTA supporting student creativity with thoughtful pedagogical integration. Future work should focus on generalizing these findings with a larger data set, across diverse contexts, and with broader measures of creativity and human-AI collaboration. Lastly, a virtual TA like JeepyTA- customizable to a course- has the potential to serve students as a thought-partner for encouraging deeper reflection and helping them build and

refine their own ideas and interests in creative ways. By combining generative AI's efficiency with human experiences and knowledge, educators can create learning environments that empower students to engage deeply with their subject matter and support creative ideation. Overall, tools like JeepyTA have the potential to benefit learners in a variety of subject domains and for a variety of skills and competencies; we are currently working in partnership with over twenty instructors at several institutions to identify specific opportunities to adapt the tool for their specific courses and student population and test the results.

References

Anderson, Barrett R., Jash Hemant Shah, and Max Kreminski. "Homogenization Effects of Large Language Models on Human Creative Ideation." In Proceedings of the 16th Conference on Creativity & Cognition, (2024): 413–25.

Fan, Yizhou, Luzhen Tang, Huixiao Le, Kejie Shen, Shufang Tan, Yueying Zhao, Yuan Shen, Xinyu Li, and Dragan Gašević. "Beware of Metacognitive Laziness: Effects of Generative Artificial Intelligence on Learning Motivation, Processes, and Performance." British Journal of Educational Technology (2024).

Henriksen, Danah, Punya Mishra, and Rohit Mehta. "Novel, Effective, Whole: Toward a NEW Framework for Evaluations of Creative Products." Journal of Technology and Teacher Education 23, no. 3 (2015): 455-478.

Imundo, Megan N., Micah Watanabe, Andrew H. Potter, Jiachen Gong, Tracy Arner, and Danielle S. McNamara. "Expert thinking with generative chatbots." Journal of Applied Research in Memory and Cognition (2024).

Kafai, Yasmin, and Mamta Shah. "Video Games as Sites for Learning and Engagement." In Teaching the Game: An Interdisciplinary Collection of Game Course Syllabi, vol. 2 (2021): 244–59.

Kasneci, Enkelejda, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser et al. "ChatGPT for good? On opportunities and challenges of large language models for education." Learning and individual differences 103 (2023): 102274.

Laney, Mason, and Prasun Dewan. "Human-AI Collaboration in a Student Discussion Forum." In Companion Proceedings of the 29th International Conference on Intelligent User Interfaces, pp. 74-77. 2024.

Lee, Hsin-Yu, Pei-Hua Chen, Wei-Sheng Wang, Yueh-Min Huang, and Ting-Ting Wu. "Empowering ChatGPT with guidance mechanism in blended learning: Effect of self-regulated learning, higher-order thinking skills, and knowledge construction." International Journal of Educational Technology in Higher Education 21, no. 1 (2024): 16.

Liu, Qinghan, Yiyong Zhou, Jihao Huang, and Guiquan Li. "When ChatGPT is gone: Creativity reverts and homogeneity persists." arXiv preprint arXiv:2401.06816 (2024).

Liu, Mengqi, and Faten M'hiri. "Beyond traditional teaching: Large language models as simulated teaching assistants in computer science." In Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1, pp. 743-749. 2024.

Liu, Xiner, Maciej Pankiewicz, Tanvi Gupta, Zhongtian Huang, and Ryan S. Baker. "A Step Towards Adaptive Online Learning: Exploring the Role of GPT as Virtual Teaching Assistants in Online Education." Manuscript under review (2024).

Nkadimeng, Mahlatse, and Piet Ankiewicz. "The Affordances of Minecraft Education as a game-based Learning Tool for Atomic Structure in Junior High School Science Education." Journal of Science Education and Technology 31, no. 5 (2022): 605-620.

Rafner, Janet, Roger E. Beaty, James C. Kaufman, Todd Lubart, and Jacob Sherson. "Creativity in the Age of Generative AI." Nature Human Behaviour 7, no. 11 (2023): 1836-1838.

Reiser, Brian J. "Scaffolding Complex Learning: The Mechanisms of Structuring and Problematizing Student Work." In Scaffolding, pp. 273-304. Psychology Press, 2018.

Roll, Ido, and Ruth Wylie. "Evolution and Revolution in Artificial Intelligence in Education." International Journal of Artificial Intelligence in Education 26 (2016): 582-599.

Shah, Mamta, Maciej Pankiewicz, Ryan S. Baker, Jiahui Chi, Yue Xin, Hetali Shah, and Dangela Fonseca. "Students' Use of an LLM-Powered Virtual Teaching Assistant for Recommending Educational Applications of Games." In Joint International Conference on Serious Games, pp. 19-24. Cham: Springer Nature Switzerland, 2024.

Sok, Sarin, and Kimkong Heng. "Opportunities, Challenges, and Strategies for Using ChatGPT in Higher Education: A Literature Review." Journal of Digital Educational Technology 4, no. 1 (2024): ep2401.

Wenger, Emily, and Yoed Kenett. "We're Different, We're the Same: Creative Homogeneity Across LLMs." arXiv preprint arXiv:2501.19361 (2025).