User Research Report

for

Exploring Pedagogical Applications of Large Language Models in English Literature Education

by Natasha Ravinand

Background

Domain Background

Large language models – also known as LLMs – have revolutionized the world in a relatively short period. LLMs are composed of transformer networks that learn "context" and "meaning" during their analysis of sequential relationships in input data (4). Popular LLMs, such as GPT-3, have been integrated into consumer-facing web applications that have generated public engagement with AI.

While these consumer applications have broadly been used in personal and professional settings, they may be additionally suitable in an educational context. Advancements in technology have long had impacts on pedagogy. For example, the release of Khan Academy began a chapter of globalized self-learning powered by the Internet. LLM-integrated software applications can contribute to even more personalized learning experiences, where a student can interface with a virtual instructor using a method best suited to that student's pedagogical profile.

Project Background

This senior project is interested in how LLMs can be fine-tuned with retrieval-augmented generation (RAG) and integrated into software applications to enhance English literature education at the undergraduate level. To view a more detailed project plan, visit the proposal document here.

Ethics Statement on AI Pedagogy Approaches

While LLMs offer the potential for a more productive pedagogical experience, it may not be wise to fully replace the human student-instructor dynamic. AI assistants may democratize study for those who cannot access high-quality academic spaces. Additionally, AI assistants can be used as a supplemental tool to in-person learning. But dialogue between human students and human teachers should not be something to replace entirely. Rather, AI should focus on enhancing, rather than replacing, a person's relationship with their learning.

Selected Overview of the Literature on LLMs and Pedagogy

Background on the Literature

The relationship between LLMs and pedagogy is quite nascent. Still, there are a few texts in the computer science and pedagogical studies literature that may provide a useful background for further discussion. The current literature places a special emphasis on English language and literature pedagogy, presumably due to the strong textual analysis abilities of existing LLMs.

How LLMs Can Create Pedagogical Opportunities across Grade Levels

Ahuja et al.'s "ChatGPT for Good" was published in April of 2023 in Volume 103 in the journal *Learning and Individual Differences*. Through a Human-Computer Interaction framework, they focused on the introduction of LLMs into pedagogical settings. After careful study, Ahuja et al. segment these opportunities by grade level.

For students enrolled at the elementary level (grades 1 through 5), LLMs may be useful in developing prompts to promote foundational reading and writing skills. They may be additionally used in summarization so that students can learn to parse text more effectively and strategically (1).

For students enrolled at the middle and high school level (grades 6 through 12), the domain of LLMs can be expanded beyond foundational reading and writing to supplement the learning of other subjects. LLMs may be used as a personal assessment generation tool, where students use practice tests to improve their active recall of the material (1).

For undergraduate students enrolled at the university level, LLMs can be used in efficient information retrieval of large corpora of text, such as academic papers, to enhance their research process. Additionally, LLMs may be useful in providing relevant

information that may be missing or ambiguous in a student's research draft (1). Opportunities at the undergraduate level may be more focused on developing analytical skills.

While the potential opportunities may differ across grade levels, Ahuja et al. maintain that the strength of LLM-integrated software applications is in their ability to facilitate a personalized pedagogical approach (1).

How LLMs are Currently Used by Instructors in the Classroom

"Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT" published in the journal Education and Information Technologies focuses on the current usage of LLMs in classrooms. Researchers conducted a small study with an "exploratory qualitative approach" in South Korea to uncover how instructors use ChatGPT to enhance their students' pedagogical experiences (2). In the classrooms included in the study, ChatGPT was often used as a tool for English language learning.

Researchers found 4 main roles of ChatGPT in the classroom.

1. With the interlocutor role, instructors often used ChatGPT as a role-player in conversations to enhance language learning. Instructors used the prompt "act as a [blank]" to engineer an appropriate conversation framework. Students then engaged with the chatbot in a dialogue (2).
2. With the content provider role, instructors used ChatGPT to generate textual materials, including "dialogue scripts, short stories, and sample words or sentences." A teacher could also personalize these contents for different students by prompting ChatGPT to modify the initial versions (2).
3. With the teaching assistant role, some instructors let students interact with ChatGPT to uncover grammatical or semantic errors in their writing.

4. With the evaluator role, instructors report using chatbots to provide initial feedback to students on their writing (2).

Researchers determined that, in classrooms with direct chatbot usage, students often became active investigators in their learning with more direct engagement.

Critical Features in Pedagogical LLM Design

Researchers Yu-Ju Lan and Nian-Shing Chen offer useful guidance on pedagogical LLM engineering in “Teachers’ agency in the era of LLM and generative AI: Designing pedagogical AI agents.”

They offer 5 desired features in an AI agent (3).

1. The theories/principles-guided learning feature requires that the AI agent must enhance the depth of understanding in a student.
2. The adaptive and personalized learning paths feature requires that the AI agent be adaptive and personalize the experience to best fit the student’s learning style.
3. The interactive learning feature requires that the AI agent reduce a student’s cognitive load by interactively educating on one concept at any given time.
4. The real-time assessment and feedback feature requires that the AI agent give instant feedback that is constructive and valuable.
5. The enforcement and quality control feature requires that the AI agent must ensure students learn and retain the most important information they learn.

Conclusions from the Existing Literature

A brief survey of the current literature suggests the following trends:

- LLM-integrated software applications show great promise in providing personalized pedagogical experiences
- The opportunities provided by these applications may vary across grade levels. At the undergraduate level, there is more opportunity to interface with students in exploratory and analytical methods
- In the classroom, instructors may use LLMs to act as interlocutors, content providers, teaching assistants, and evaluators
- Pedagogical AI agents should provide a sound learning environment that is adaptive and comprehensive while focusing on providing constructive feedback

Study overview

Motivations Behind Conducting a UX Research Survey

We are interested in how LLMs can be fine-tuned and then integrated into pedagogical software applications in teaching undergraduate-level English literature. In addition to findings from existing research papers, it is necessary to understand how current students currently engage with AI both in and out of the classroom. With this understanding, we can ascertain which qualities of an LLM-integrated pedagogical application are most useful.

Materials and Methods

This senior project is interested in English literature pedagogy at the undergraduate level, so it was necessary to restrict participants in the survey to current or recently graduated college students. Responses were only solicited from those with Yale College affiliation. The survey was publicized through the CPSC 490 proposal presentation as well as the CPSC 439 Canvas page.

The survey was constructed with opportunities to obtain both qualitative and quantitative data from students on their current experiences with AI agents, as well as what they desire from a pedagogical experience. It included a list of approximately 20 questions, most of which were multiple-choice.

The survey included 5 separate sections.

1. The academic background section intended to obtain information on the academic background of participants, including their major(s) and familiarity with humanities courses.
2. The baseline assessment of the academic English literature background section is intended to assess the existing experience of a student in

English literature training. This section also obtained information on a student's extracurricular reading habits.

3. The current usage of AI assistants section intended to evaluate how students currently use AI assistants personally, professionally, and academically, across 4 categories: education & learning (ex. summarizing information, learning new information, understanding topics); productivity (ex. managing tasks, drafting emails, translation); professional (ex. market research, building presentations); creative (ex. finding writing prompts, getting personalized movie/TV/book suggestions)
4. The designing of an effective pedagogical user experience section intended to solicit what a desired user experience might be for a student interacting with an AI assistant to interface with English literature course materials.

The survey was conducted via Google Form and obtained 15 anonymized responses over approximately two weeks. It can still be accessed here, though it is no longer accepting responses.

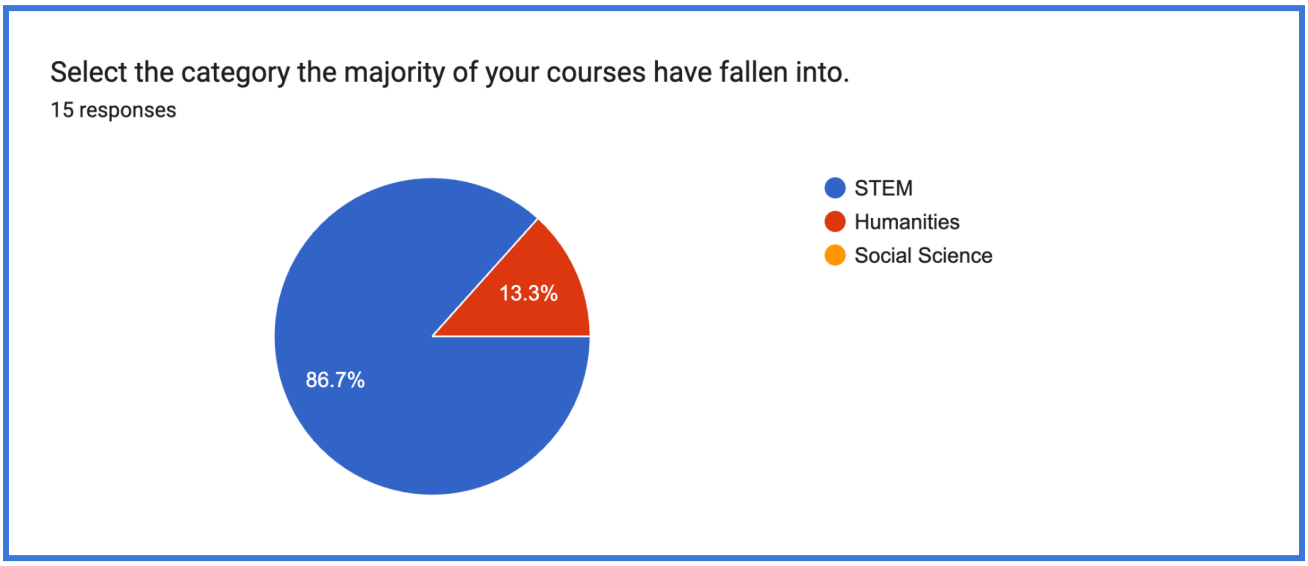
Findings

Key Observations from an Analysis of the Data

After conducting the survey, several insights were gathered. We can elaborate on them by section.

Academic Background

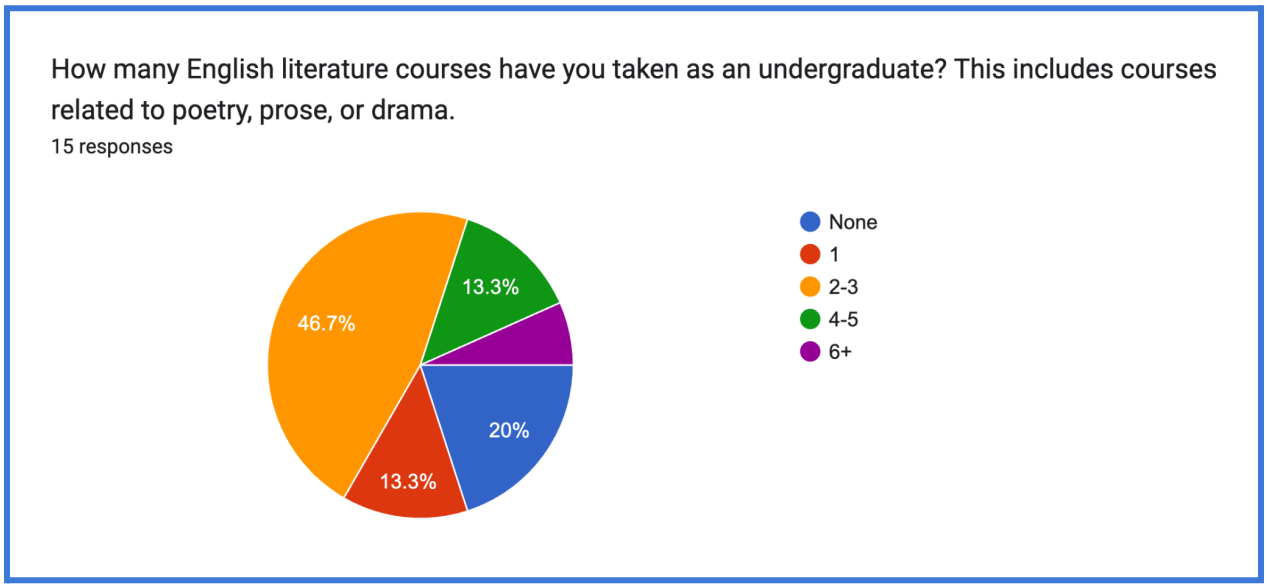
More than 80% of respondents had a primary major related to the STEM field, and 86.7% of the majority of their courses at the undergraduate level have been in STEM departments.

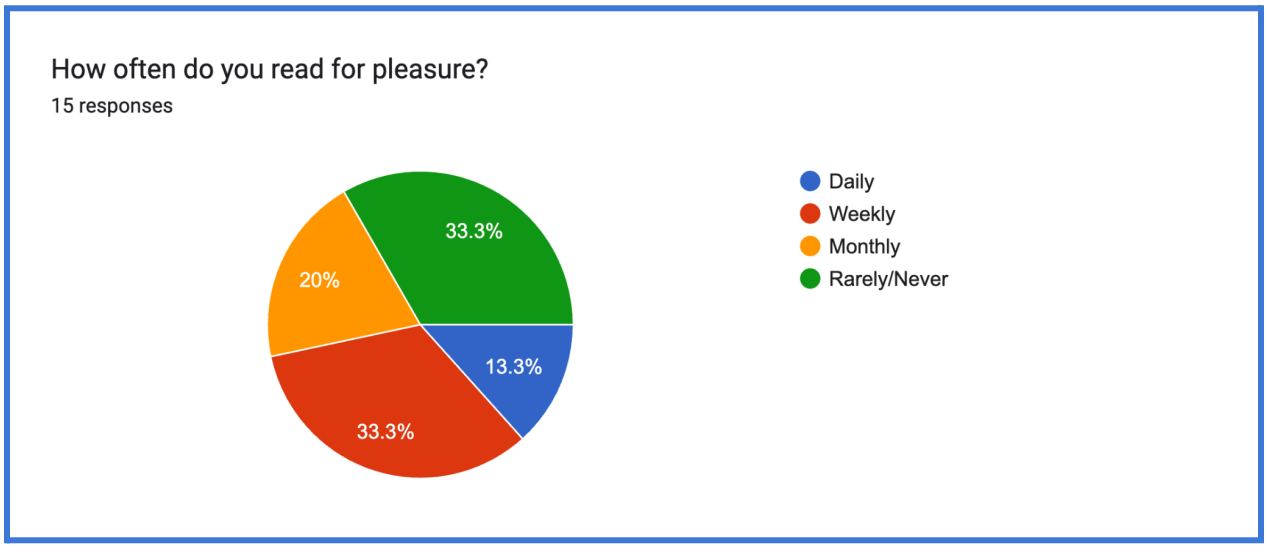


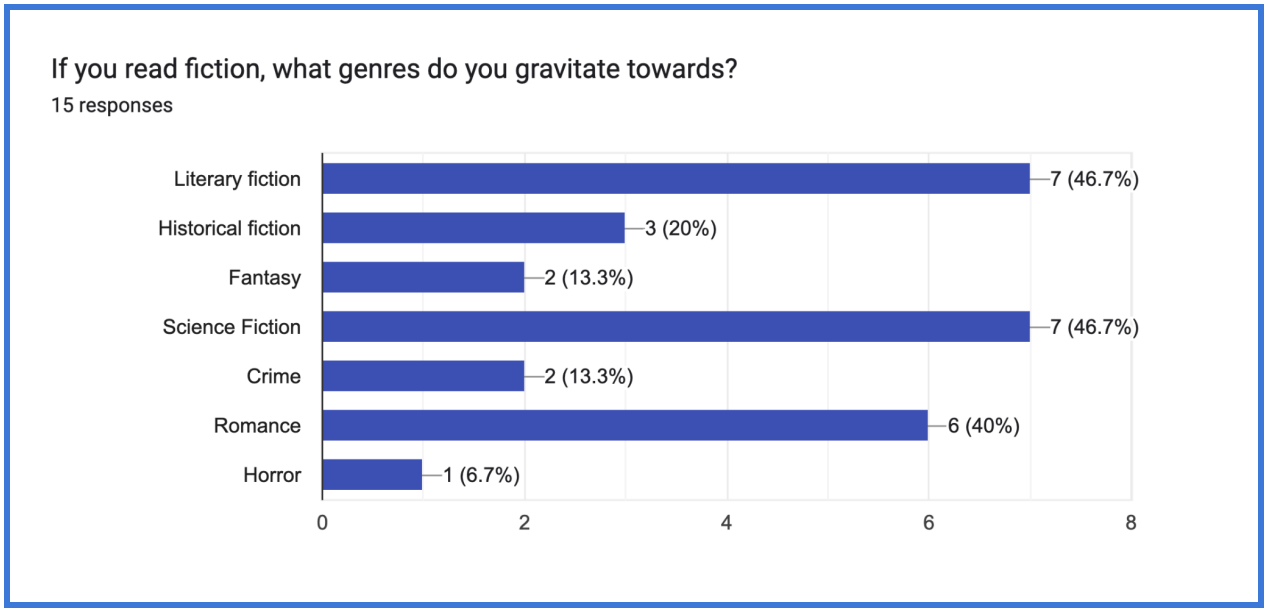
Baseline assessment of academic English literature background

A plurality of respondents (46.7%) had taken 2-3 English literature courses at the undergraduate level, while 20% had taken none. This suggests that many respondents were familiar with the undergraduate English literature curriculum but were not at an advanced level.

When asked about whether they read for pleasure, responses were split roughly evenly: 13.3% of respondents reported reading daily, 33.3% reported reading weekly, 20% reported reading monthly, and 33.3% reported rarely or never reading. Of those who reported reading for pleasure, the majority (60%) reported reading predominantly fiction. Interestingly, literary and science fiction were the most popular genres.



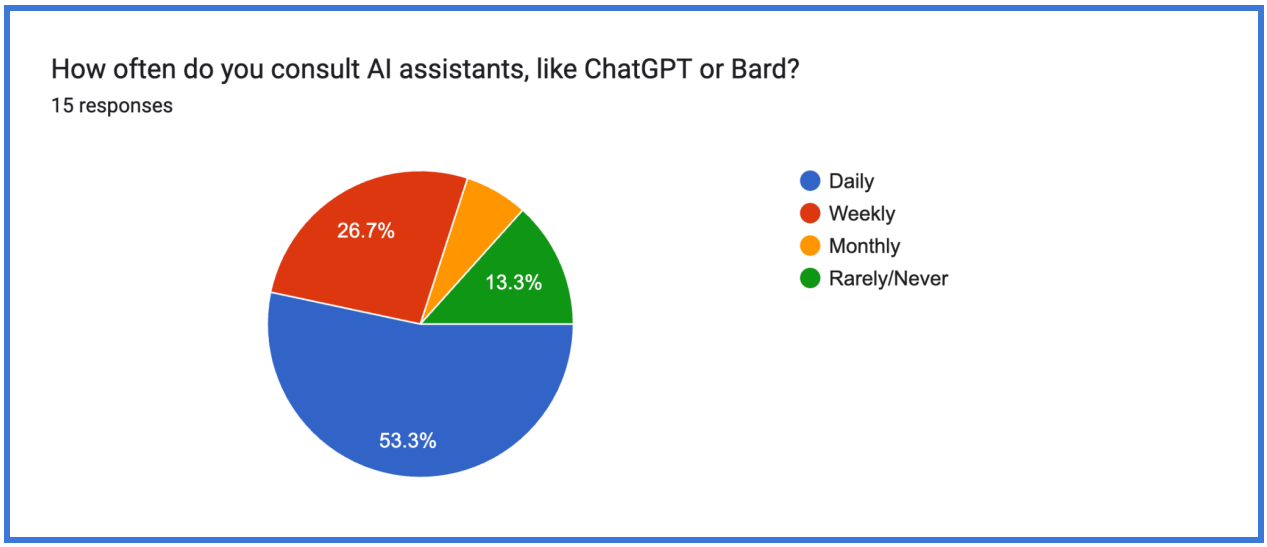


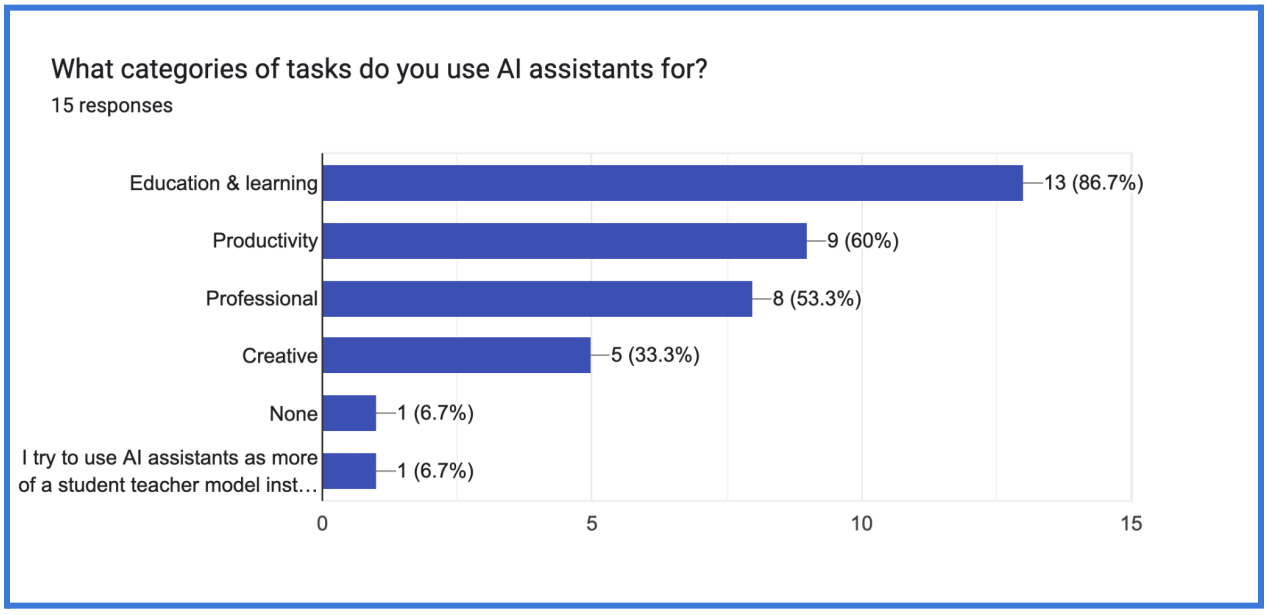


Participants reported different methods of reading and analyzing literary texts in a classroom setting. A few popular ones were reading and rereading text, annotating on digital or hard copies, and taking notes in a separate notebook.

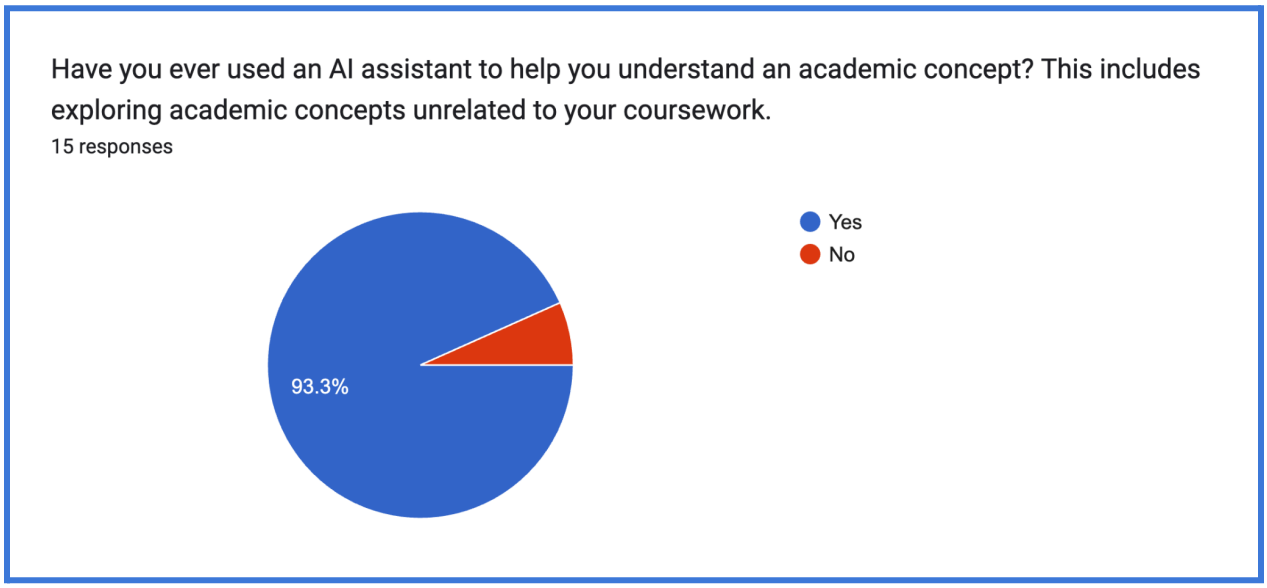
Current Usage of AI assistants

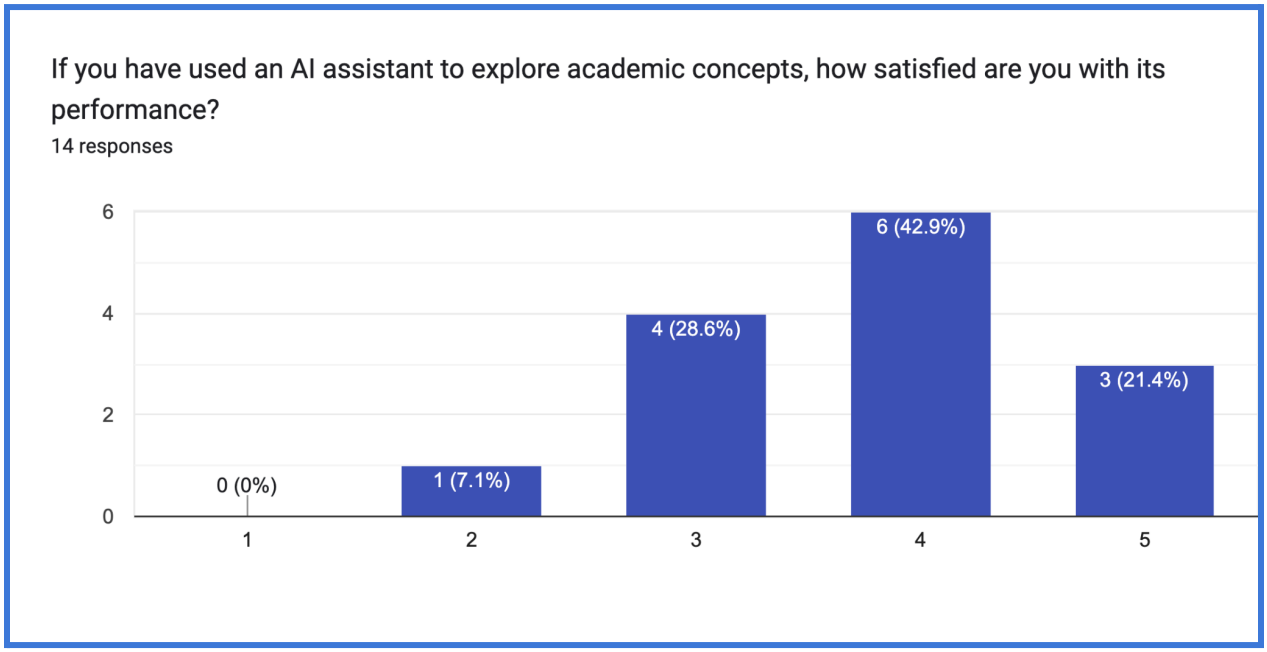
A majority of respondents report using AI assistants like ChatGPT or Bard daily at 53.3%, and 26.7% report using them weekly. This suggests that AI-integrated software applications that focus on dialogue between the user and model are exceedingly popular. Only 13.3% of participants report rarely or never using these applications. Respondents generally use AI assistants for education/learning (86.7%) and productivity (60%).


How often do you consult AI assistants, like ChatGPT or Bard?
15 responses
26.7%
13.3%
53.3%
Daily
Weekly
Monthly
Rarely/Never

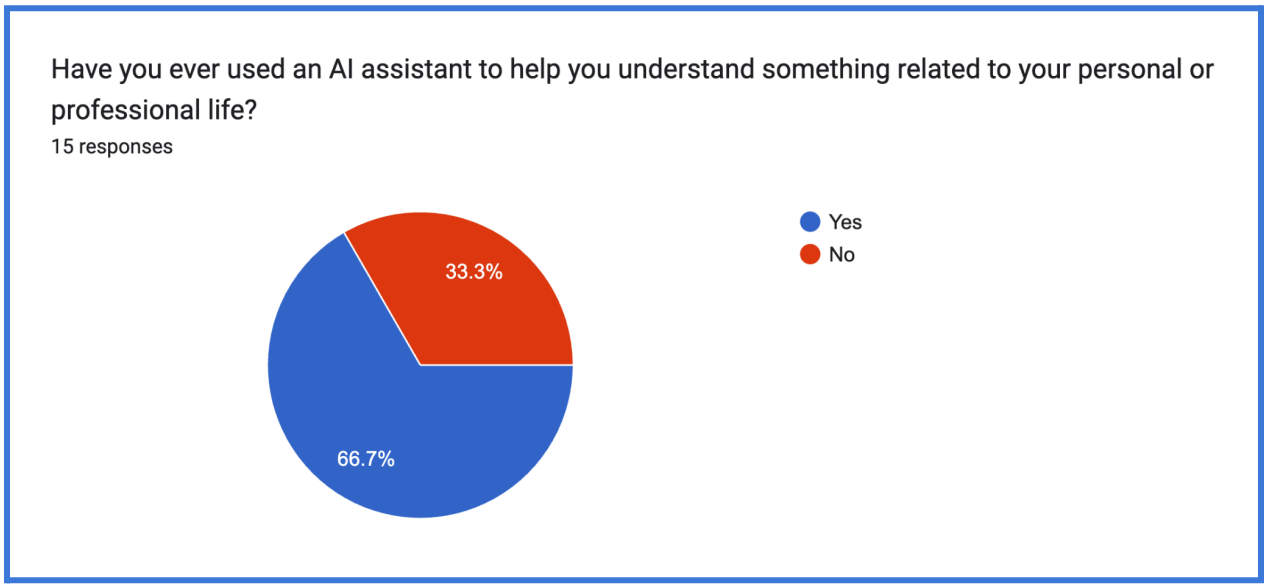

What categories of tasks do you use AI assistants for?
15 responses
Education & learning
13 (86.7%)
Productivity
9 (60%)
Professional
8 (53.3%)
Creative
5 (33.3%)
None
1 (6.7%)
I try to use AI assistants as more of a student teacher model inst…
1 (6.7%)
0
5
10
15

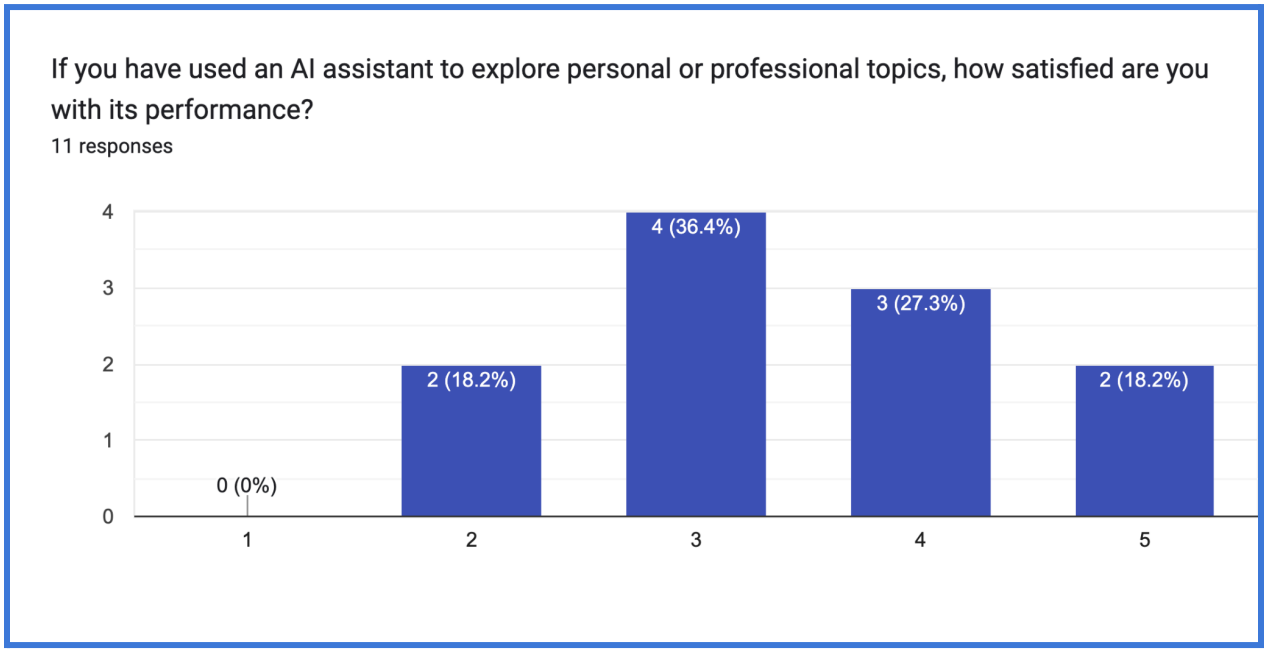
The overwhelming majority of respondents (93.3%) have used an AI assistant to understand an academic concept at some point. Generally, they are satisfied with the responses of the assistants they use; using a standard scale going from 1 (not helpful) to 5 (very helpful), 42.9% of respondents rated their assistants with a 4 while 21.4% of participants rated with a 5.





While exploring academics seems to be a popular use case for respondents, participants still reported using AI assistants in personal or professional capacities (at 66.7%). They are generally less satisfied with their performance in this context. Of those who use AI assistants outside of academics, 36.4% rate their assistants with a 3, 27.3% rate with a 4, and only 18.2% rate with a 5.

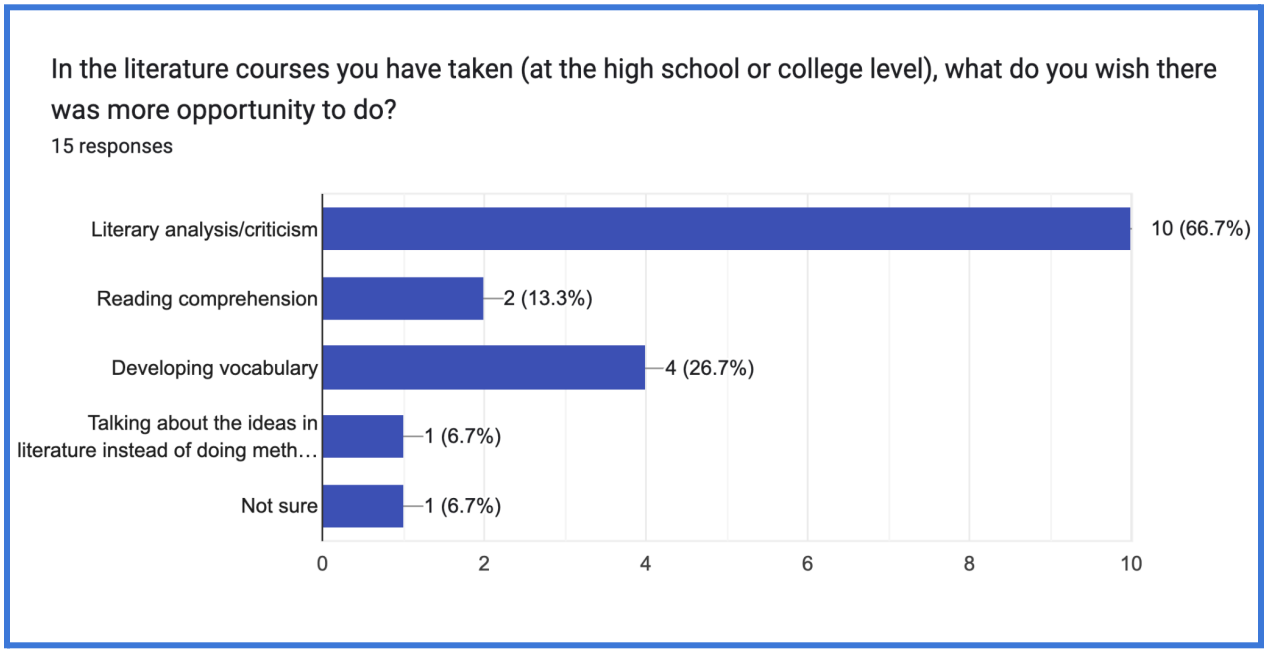


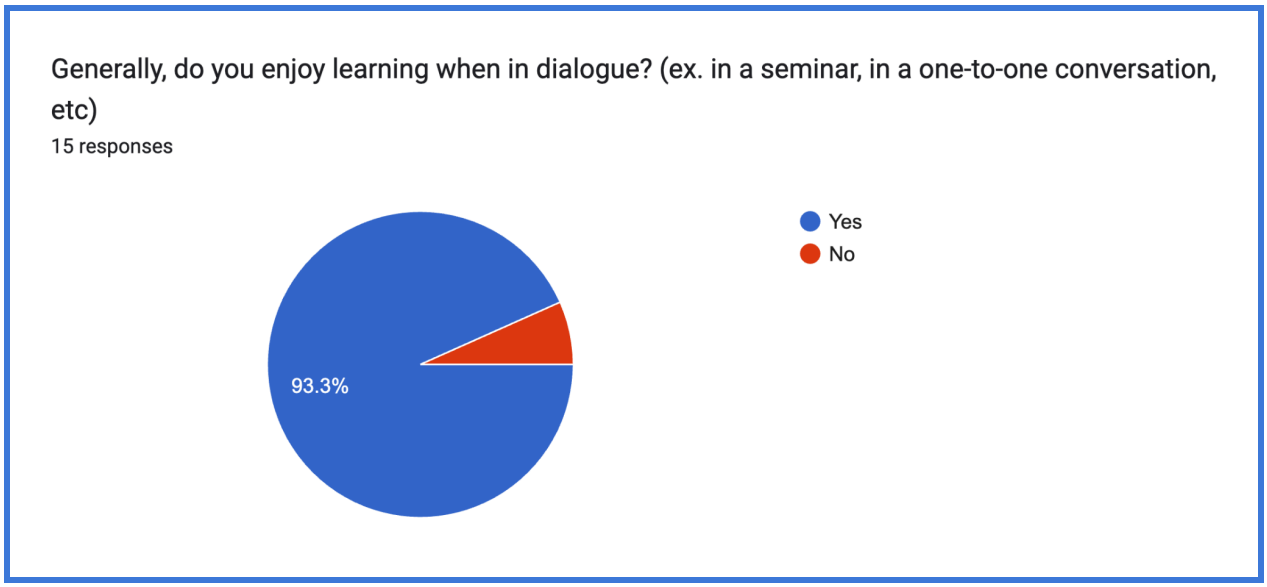


Designing an effective pedagogical user experience

The majority of respondents (66.7%) wish there were more opportunity to engage in literary analysis and criticism at the high school and undergraduate levels. 26.7% also wished to expand their vocabulary in English literature courses. Surprisingly, the desire to engage with reading comprehension was surprisingly less popular, with 13.3% expressing interest.

Small group discussions, and learning in dialogue, seem to be quite popular among respondents. 73.3% of participants reported engaging in small group discussions in their previous English courses. The overwhelming majority of respondents (93.3%) report enjoying dialogue-based learning.





When asked what they would look for in an English literature-focused AI assistant, responses widely varied, but a few responses were popular. Most respondents wanted their assistant to be personable and friendly yet “trustworthy”, “biased”, and “critical”. Being “inquisitive” and “curious” were also desired. Many respondents were interested in an assistant that offers the ability to explore ideas with the user, rather than one who took a more objective approach. They were also interested in engaging in dialogue with the assistant, where the assistant would ask probing questions and instigate conversation.

Limitations of the Survey

Given the avenues in which the survey was publicized (through the CPSC 439 and CPSC 490 network), the majority of respondents came from computer science backgrounds. This may have introduced a bias in how respondents thought about interacting with an AI assistant, and how they thought about English literature writ large.

In addition, if the survey elicited more participation and produced a larger sample (ex. Over 100 respondents), we might have obtained a better diversity of respondents and more information.

Next steps

This research has produced material insights that will prove invaluable during the next two phases of the project.

Implications for Fine-Tuning with RAG on an LLM

We learned from this study that users value certain characteristics in a pedagogical AI agent. This includes, but is not limited to:

- Having a personable and friendly profile yet also “trustworthy”, “biased”, and “critical”
- Having an inquisitive nature that is inviting to exploration
- Having a wealth of reliable domain knowledge
- Having the ability to detect when a user is struggling with a concept and offering to revise and reexplore

These insights may inform the kinds of training data we use in the RAG. For example, it will be crucial to use training data that comes from a pedagogical perspective in the English literature domain. However, this data should not influence the model into taking an objective stance on its interactions with a user; rather, it should invite experimentation and opportunities for the user to learn.

Implications for an LLM-integrated Pedagogical Web Application

This survey uncovered the desire to engage in dialogue-based learning, even with an AI assistant. Respondents report already using applications like ChatGPT at high frequencies and are generally satisfied with its responses. Additionally, they value learning in conversation and are generally interested in the idea of learning via an AI tool.

The web application will integrate these findings by prioritizing a user experience that allows seamless chat-based conversations with the AI assistant.

References

1. Ahuja, A.S., et al. “ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education.” *Learning and Individual Differences*, 9 Mar. 2023, www.sciencedirect.com/science/article/abs/pii/S1041608023000195.
2. Jeon, J., Lee, S. Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. Educ Inf Technol 28, 15873–15892 (2023). https://doi.org/10.1007/s10639-023-11834-1
3. Lan, Yu-Ju, and Nian-Shing Chen. “Teachers’ Agency in the Era of LLM and Generative AI: Designing Pedagogical AI Agents.” Educational Technology & Society, vol. 27, no. 1, 2024, p. I–XVIII. JSTOR, https://www.jstor.org/stable/48754837. Accessed 25 Feb. 2024.
4. “What Are Large Language Models?” NVIDIA, www.nvidia.com/en-us/glossary/large-language-models/. Accessed 29 Jan. 2024.