Embracing AI in Education: Understanding the Surge in Large Language Model Use by Secondary Students

Anonymous ACL submission

Abstract

The impressive essay writing and problem-solving capabilities of large language models (LLMs) like OpenAI's ChatGPT have opened up new avenues in education. Our goal is to gain insights into the widespread use of LLMs among secondary students to inform their future development. Despite school restrictions, our survey of over 300 middle and high school students revealed that a remarkable 70% of students have utilized LLMs, higher than the usage percentage among young adults, and this percentage remains consistent across 7th to 12th grade. Students also reported using LLMs for multiple subjects, including language arts, history, and math assignments, but expressed mixed thoughts on their effectiveness due to occasional hallucinations in historical contexts and incorrect answers for lack of rigorous reasoning. The survey feedback called for LLMs better adapted for students, and also raised questions to developers and educators on how to help students from underserved communities leverage LLMs' capabilities for equal access to advanced education resources. We propose a few ideas to address such issues, including subject-specific models, personalized learning, and AI classrooms.

1 Introduction

The remarkable generative capabilities of InstructGPT [16], introduced over two years ago, has led to widespread usage of large language models (LLMs). Since then, the capabilities of LLMs like OpenAI's ChatGPT [13], Google Gemini [18] and Meta Llama [20] have advanced significantly. According to an OpenAI report [15], GPT-4 achieves above the 90th percentile on exams like AP Biology, AP Calculus BC, and SAT math. Consequently, many middle and high school students have begun using LLMs for schoolwork assistance.

This growth in capabilities of artificial intelligence (AI) raises concerns about its impact on students' learning experiences [8]. We have observed schools implement policies restricting students from using LLMs due to concerns for cheating as well as inaccurate answers. Even OpenAI cautions secondary students against using ChatGPT and specifically prohibits usage by children under the age of 13 [12]. While OpenAI is developing ChatGPT Edu for educational purposes[14], its audience is currently only university students. Consequently, the secondary school age group lacks support and influence over the services and developments of LLMs [16].

Prohibited or not, access to ChatGPT is easy with no registration requirement. Such openness makes it challenging to restrict LLM usage by students. Instead of imposing restrictions, we believe the best approach is to understand how students utilize LLMs and develop curriculum and adapt LLMs to accommodate them. While some research [3, 1] has been conducted on personalized learning with LLMs, large-scale comprehensive surveys understanding the trend of student usage of LLMs are lacking. Our study aims to fill this gap, as shown in Table 1.

A key question we sought to address was the prevalence of LLM usage among middle and high school students. Our survey confirmed our suspicion of its widespread usage, with over 70% of respondents reporting usage, higher than the 43% usage among young adults as reported in a February poll [10]. Surprisingly, we found that the percentage of students who have used LLMs was quite consistent across grades 7 to 12, as shown in Figure 1. We had originally anticipated lower usage among students in lower grades.

In the subsequent sections of this paper, we will present the details of our survey methodology and the results obtained for a range of other questions, including:

- What subjects do students seek assistance from LLMs and how satisfied are they?

Table 1: Our study has more participants than past studies that investigated LLM Usage

Study	Demographic	Participants	Purpose
Li et. al [9]	Chinese secondary schoolers	76	ChatGPT usage motives
Belghith et. al [5]	Middle school students	24	AI interactions
Rahman et. al [17]	College programming students	N/A	ChatGPT for coding
This study	Middle and high schoolers	306	LLM trends & potential

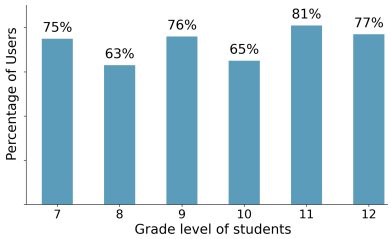

Figure 1: LLM usage remained consistent among 7th-12th graders.

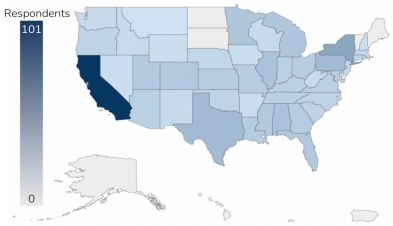

Figure 2: We received responses from students in 43 states.

- How stringent are school policies regarding LLM usage and how do students regard them?
- Are there demographic differences among students that affect LLM usage?

To address the survey feedback, we will present proposals on how to further improve LLMs for education and embrace LLMs in classrooms to enhance learning for students from different backgrounds.

In summary, our contributions are several fold. First, with over 300 respondents across the United States, we obtained a collective view of LLM usage from a large, diverse group of secondary students. Second, we drew insights on critical LLM usage questions that are informative for educators, policymakers, and developers to embrace AI in education. Third, we provided proposals to integrate LLMs into education more effectively and responsibly.

2 Data Collection

Our target group consisted of middle and high school students throughout the United States and encapsulated a diverse population.

The survey form took just a few minutes to complete, which encouraged respondents to take the survey without feeling lethargic and answering randomly. We used Centiment [6], a survey platform that allowed for connections with respondents throughout the United States, who were compensated over 80 cents for completing the short survey. We also created a Google Form survey, which we shared by visiting local tutoring programs and posting on social media.

The following information was provided to the survey takers: *"Results of this survey are completely confidential and will be used for data analysis to understand students' LLM usage."* All responses will be made available in a public GitHub repository with the respondents' personally identifiable information removed.

Grade	6th	7th	8th	9th	10th	11th	12th
Students	3	17	56	99	52	48	31

Table 2: Number of responses from students in grades 6-12

Responses included students ranging from sixth to twelfth grade (Table 2). California had the highest number of responses, as the in-person surveys were conducted in this state, and we received feedback from a total of 43 states (Figure 2). The respondents also came from various educational backgrounds, including private, public, charter, and home schooled. Lastly, the demographic encompassed students who consistently had access and students who seldom had access to technology for personal purposes.

3 Data Analysis

Figure 3 shows the usage frequency per week of LLMs, including ChatGPT, LLaMA, Gemini and Bing Copilot. Our survey shows that 71% of the correspondents have used LLMs at least once, while 9% used them daily. Since our survey shows 84% of the LLM users use ChatGPT, this translates to 60% of ChatGPT users among all the respon-

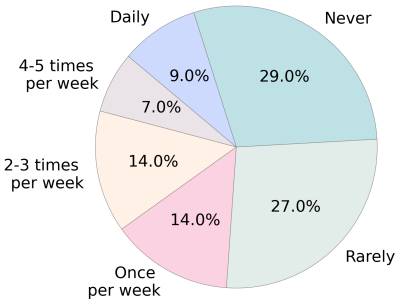

Figure 3: Over 70% of students have used LLMs.

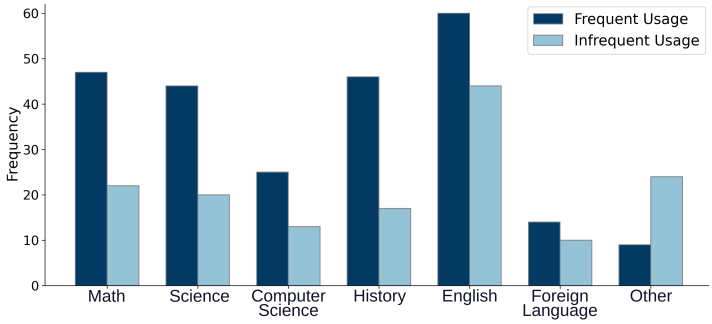

Figure 4: LLMs are used in all school subjects.

dents. In comparison, 43% of young adults of age 18-29 have used ChatGPT[10]. Our survey clearly shows popularity of LLMs among the secondary students and consistency across the grades, as reported in Figure 1.

LLMs are used in all school subjects. Previous research reveals that LLMs exhibit inferior mathematical abilities compared to addressing open-ended requests in English [7]. Despite this, our survey shows 28% students still used LLMs for the math subject. They also use LLMs in various other subjects, including foreign languages and history classes (Figure. 4). According to our survey, 57% of LLM users have reported finding them useful for writing text, while 43% have found them useful for math and STEM questions. We also noticed that among students who use LLMs at least twice a week, the majority of them have tried LLMs for all class subjects.

Students share various thoughts regarding LLM impact. Responses to an open-ended question regarding the impact of LLMs on students' academic performance were diverse, despite the common usage. The following are a few student responses with contrasting thoughts on LLMs' helpfulness:

- *"Yes, they have helped simplify complex topics."*
- *"No, increased reliance on such tools may lower the level of other fundamental skills that you develop."*
- *"Sometimes, as it may be inaccurate at certain points."*

Some students, like the last, believe LLMs have inaccuracies when answering factual questions, which is supported by previous research that discovered ChatGPT hallucinates the most when questions are inputted consecutively [2]. However, many other students fail to recognize the inaccuracies [21].

Most students desire LLM improvements for schoolwork. When asked about potential improvements to LLMs, 3% of individuals expressed no desire for any modifications. 51% desired for more coherent responses, accurate answers and a great ability to address complex questions, indicating somewhat dissatisfaction with the current models. Only 20% of respondents expressed interest in improving LLMs' understanding of emotions and better conversations, indicating students' preference for using LLMs to answer questions rather than using LLMs as a conversational tool.

LLM usage is prevalent despite school policies. Participants were asked to rate from 1 to 5 how strict their school policies are on getting help from LLMs, with a score of 5 meaning severe punishment if LLM usage was found in their schoolwork. The average rating was 3.48, signaling generally restrictive policies.

We asked students to rate their ethical concerns on using LLMs, and got an average rating of 2.83 on a scale of 1-5, where a higher number meant a larger concern. The value is slightly less than what they perceived as severity of punishment rating, signaling usage outweighs ethical concerns, as shown by figure 5. Interestingly, we found that 2% of the students who rated themselves as a 5 on ethical concern reported using LLMs at least once a week, indicating that resisting using LLMs can be difficult.

More technologically advanced regions have more users. According to the 2022 State Technology and Science Index, which ranks the states based on their technological economies, the five highest ranking states are Massachusetts, California, Colorado, Maryland, and Utah [11]. We dis-

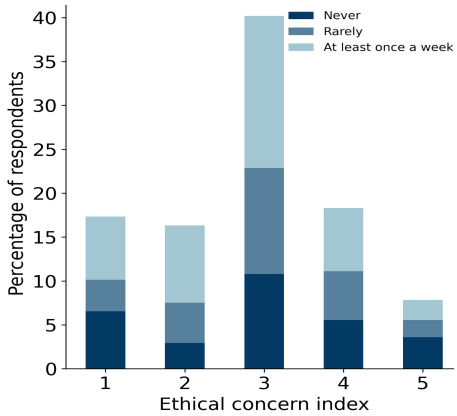

Figure 5: LLM usage persists despite ethical concerns.

covered that 80.2% of students in these five states and 64.3% of students in the rest of the states had used LLMs. This difference of over 15% shows a concerning gap between students with different tech savviness levels. However, it also points to a trend of increasing LLM usage among students since LLMs' continuing improvement and decreasing access barrier will attract more users from states with lower tech benchmarks.

We also discovered a small usage gap between private and public school students. Our survey shows 76.7% of private school students reported LLM usage versus 71.3% of public school students. Since students attending private schools tend to be from wealthier families, this 5% difference is likely caused by family income differences.

Users who can pay for models have an advantage. Out of the 17 participants who possess a paid version of the models, the outcomes regarding its usefulness are predominantly in favor of a positive response. Of these 17, all except one reported using LLMs at least 2-3 times a week, significantly more frequent than students using free models. As companies release higher-quality paid versions, students who have paid models will gain a greater advantage. This further proves the existence of an LLM usage gap between families with different income levels, making it urgent to ensure students from various economic backgrounds can all benefit from LLMs equally.

4 Discussion and Conclusion

The survey shows the use of LLMs in secondary education is widespread and is expected to grow further with the advent of more capable models and increased awareness. It also shows that resource-rich students are more likely to take advantage of LLMs with their studies, potentially causing underserved students with limited learning resources to fall further behind. We therefore propose a few ideas to address the feedback and to make LLMs more useful to students of all backgrounds.

AI models should be fine-tuned for student usage. Rather than restricting the use of LLMs in education, educators and AI developers should collaborate to build fine-tuned models trained on textbooks and teacher inputs [4]. Such models will hallucinate less and be more relevant to topics at hand. Firstly, the LLMs must integrate a diverse range of educational resources to allow the model to generate well-rounded and informed responses. Secondly, they should be trained with teacher inputs that provide additional context on topics to encourage critical thinking by prompting students to explore topics more deeply. Lastly, safeguards must be implemented to prevent the generation of inappropriate or harmful content and protect privacy.

AI tutors can personalize learning. An AI tutor will make students' educational experiences more accessible, comprehensive and engaging due to its easy access, vast knowledge and context-aware conversation capabilities [19]. This is particularly beneficial to students in remote areas, those with disabilities, or learners who need additional help outside of traditional classrooms. AI tutors can also provide content in multiple languages and formats, catering to diverse learning needs and making education more inclusive. They can be made more useful by creating customized learning paths that adapt to a student's learning habits, providing immediate feedback and tailored quizzes, and aligning with the curriculum to address specific learning objects of various subjects and grades.

AI classrooms promote learning equality. LLMs provide an opportunity for AI developers and teachers to work together to level the learning field for students from different demographics. We propose the idea of "AI classrooms:" the "AI teachers" in the virtual classrooms are trained on teaching videos and notes of experienced teachers. Using the multi-modal capabilities of LLMs, the "AI teachers" can "teach" and interact with the students in real-time. They will also proactively ask students questions and quiz them to provoke critical thinking skills. These multi-modal capabilities can also be personalized for differing learning paces. With government and industry support, the "AI classrooms" can be made free for those under-resourced districts as a supplement to students' education in school.

5 Limitations

Our survey was online-based, requiring respondents to have access access to the internet to take the forms. Therefore, it might have introduced accessibility bias, which could lead to a higher LLM usage percentage than the national truth. The study provides educators, policymakers, and developers with information regarding LLMs in education but may not encompass all viewpoints due to the focus on students.

6 Ethical Considerations

The data annotation in this paper is classified as exempt by Human Subject Committee under IRB protocols category 2. This paper includes the instructions given to participants before taking the survey and a GitHub link to the survey questions. There was fair compensation for participants who were crowdsourced, and no personally identifying information was collected.

References

[1] Ibrahim Adeshola and Adeola Praise Adepoju. 2023. The opportunities and challenges of chatgpt in education. *Interactive Learning Environments*, pages 1–14.

[2] Zakia Ahmad, Wahid Kaiser, and Sifatur Rahim. 2023. Hallucinations in chatgpt: An unreliable tool for learning. *Rupkatha Journal on Interdisciplinary Studies in Humanities*, 15.

[3] David Baidoo-anu and Leticia Owusu Ansah. 2023. Education in the era of generative artificial intelligence (ai): Understanding the potential benefits of chatgpt in promoting teaching and learning. *Journal of AI*, 7:52–62.

[4] Patrick Bassner, Eduard Frankford, and Stephan Krusche. 2024. Iris: An ai-driven virtual tutor for computer science education. *Innovation and Technology in Computer Science Education*, 1.

[5] Yasmine Belghith, Atefeh Mahdavi Goloujeh, Brian Magerko, Duri Long, Tom Mcklin, and Jessica Roberts. 2024. Testing, socializing, exploring: Characterizing middle schoolers' approaches to and conceptions of chatgpt. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA. Association for Computing Machinery.

[6] Centiment. Survey with centiment. https://www.centiment.co/.

[7] Simon Frieder, Luca Pinchetti, Ryan-Rhys Griffiths, Tommaso Salvatori, Thomas Lukasiewicz, Philipp Petersen, and Julius Berner. 2023. Mathematical capabilities of chatgpt. In *Advances in Neural Information Processing Systems*, volume 36, pages 27699–27744. Curran Associates, Inc.

[8] Enkelejda Kasneci, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, Stephan Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jürgen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, Jochen Weller, Jochen Kuhn, and Gjergji Kasneci. 2023. Chatgpt for good? on opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103:102274.

[9] Yalin Li, Jinyuan Chen, Haowen Zhou, Haoyang Yuan, and Ruolin Yang. 2023. Research on motivation and behavior of chatgpt use in middle school students. *International Journal of New Developments in Education*, 5.

[10] Colleen McClain. 2024. Americans' use of chatgpt is ticking up, but few trust its election information. https://pewrsr.ch/3VyXbJR.

[11] Milken Institute. 2022. State technology and science index. https://statetechandscience.org.

[12] OpenAI. 2023. Terms of use. https://openai.com/policies/terms-of-use/.

[13] OpenAI. 2024. Chatgpt. https://chat.openai.com.

[14] OpenAI. 2024. Openai for education. https://openai.com/index/introducing-chatgpt-edu/.

[15] OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, and Gabriel Bernadett-Shapiro et. al. 2024. Gpt-4 technical report. https://doi.org/10.48550/arXiv.2303.08774. *Preprint*, arXiv:2303.08774.

[16] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. https://doi.org/10.48550/arXiv.2203.02155. *Preprint*, arXiv:2203.02155.

[17] Md. Mostafizer Rahman and Yutaka Watanobe. 2023. Chatgpt for education and research: Opportunities, threats, and strategies. *Applied Sciences*, 13(9).

[18] Gemini Team, Machel Reid, Nikolay Savinov, Denis Teplyashin, Dmitry, Lepikhin, Timothy Lillicrap, Jean baptiste Alayrac, Radu Soricut, Angeliki Lazaridou, Orhan Firat, Julian Schrittwieser, Ioannis Antonoglou, Rohan Anil, Sebastian Borgeaud, Andrew Dai, Katie Millican, Ethan Dyer, Mia Glaese, Thibault Sottiaux, and ... 2024. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. https://doi.org/10.48550/arXiv.2403.05530. *Preprint*, arXiv:2403.05530.

[19] Danielle R Thomas, Jionghao Lin, Erin Gatz, Ashish Gurung, Shivang Gupta, Kole Norberg, Stephen E Fancsali, Vincent Aleven, Lee Branstetter, Emma Brunskill, and Kenneth R Koedinger. 2024. Improving student learning with hybrid human-ai tutoring: A three-study quasi-experimental investigation. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*, LAK '24. ACM.

[20] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models. https://doi.org/10.48550/arXiv.2302.13971. *Preprint*, arXiv:2302.13971.

[21] Shen Wang, Tianlong Xu, Hang Li, Chaoli Zhang, Joleen Liang, Jiliang Tang, Philip S. Yu, and Qingsong Wen. 2024. Large language models for education: A survey and outlook. https://doi.org/10.48550/arXiv.2403.18105. *Preprint*, arXiv:2403.18105.