Assessing the Impact of ChatGPT on Education: Sentiment Analysis and Topic Modeling of Reddit Discussions

Nawal Sarwar, Marcos Padilla, Nicholas Eversgerd, and Shebuti Rayana





Introduction

SUNY OLD WESTBURY

The rapid rise of Generative AI, such as ChatGPT, introduced by OpenAI in November 2022, has significantly reshaped the role of Artificial Intelligence (AI) in education. This study explores the impact of ChatGPT on teaching methods, learning outcomes, and student engagement by analyzing usergenerated content from Reddit.com. By examining diverse perspectives from educators, students, and professionals, the study provides valuable insights into the challenges and opportunities associated with integrating ChatGPT into settings. To understand the emotional landscape surrounding ChatGPT's integration, the study employs sentiment analysis, topic modeling, and data visualization techniques. This approach uncovers a spectrum of reactions, ranging from enthusiasm about its transformative potential to dilemmas misuse. The analysis is based on a dataset of 2,338 subreddit posts collected from seven educational subreddits, e.g., r/professors, r/teachers, r/education, r/college, r/academia, r/university, and r/getStuding between November 2022 and August 2024. A validation dataset of 200 posts was manually annotated. Sentiment classification was performed using models including VADER, BERT, roBERTa, and GPT, while topic modeling algorithm such as LDA was used to extract themes from posts with varying sentiments. Our qualitative analysis indicates that the sentiment labels generated by GPT-3.5 Turbo and GPT-4.o, when accompanied by reasoning, demonstrated higher quality than human annotations with reasoning. Furthermore, the content analysis suggests that while ChatGPT presents opportunities to enhance educational practices, it also raises concerns about students' over-reliance on AI, which may hinder critical thinking and authentic learning. A key finding is the growing concern over issues such as plagiarism, cheating, academic integrity, equitable access, and many more critical challenges that have emerged amid the widespread adoption of Generative AI. Educators have expressed challenges in accurately assessing student skills due to AI-generated responses. This study underscores the need for responsible AI integration in teaching and learning, and clear guidelines to ensure ethical adoption while maintaining educational integrity. Future research should explore the broader societal implications of AI in academia and beyond, developing best practices that balance innovation with accountability.

Method

Sentiment Analysis

Sentiment analysis is the process of using natural language processing and machine learning to identify and categorize emotions or opinions expressed in text as positive, negative, or neutral.

V.A.D.E.R:

A rule-based sentiment analysis tool that evaluates the sentiment of text, like social media posts, using a pre-built dictionary of words and their associated sentiment scores, making it easy to use without training data and suitable for quick sentiment analysis tasks.

BERT

The BERT sentiment analysis model uses bidirectional transformers to understand context from both directions in text, enabling it to capture subtle sentiment cues. Pretrained on large corpora and fine-tuned on sentiment tasks, BERT offers robust performance for analyzing complex language, including educational discourse.

RoBERT

The RoBERTa sentiment analysis model is a transformer-based classifier trained on large-scale datasets like SST-2 and Twitter data to detect sentiment with high accuracy. It improves on BERT by using more training data and optimized hyperparameters, making it well-suited for nuanced sentiment detection in social media and educational contexts.

GPT 3.5 Turbo & GPT 4.o:

Advanced transformer-based language models by OpenAI capable of understanding and generating human-like text. In sentiment modeling, these models go beyond traditional classifiers by deeply interpreting context, sarcasm, and nuanced emotion, allowing them to assign sentiment labels with high accuracy and provide reasoning—making them powerful tools for high-quality, human-aligned sentiment analysis.

Topic Modeling

Latent Dirichlet Allocation(LDA):

Latent Dirichlet Allocation (LDA) is a generative probabilistic model used in topic modeling to discover hidden topics in a set of documents based on word co-occurrence patterns.

Manual Labeling

A validation dataset of 200 posts was annotated by three undergraduate students following a structured labeling guideline refined through three iterative phases. Each phase improved the guideline to account for nuances such as sarcasm, passive aggression, and variations in sentiment intensity. In addition to sentiment labels, the annotators provided reasoning and assigned a confidence level (high, medium, or low) to their assessments.

Tables/Figures

	Maighted Average				
V.A.D.E.R	Weighted Average	0.62	0.47	0.43	0.475
	Positive	0.42	0.69	0.52	
	Neutral	0.47	0.68	0.55	
	Negative	0.91	0.13	0.23	
BERT	Weighted Average	0.70	0.57	0.52	0.5750
	Positive	1.00	0.15	0.27	
	Neutral	0.61	0.73	0.67	
	Negative	0.50	0.83	0.62	
roBERTa	Weighted Average	0.82	0.78	0.76	0.7750
	Positive	1.00	0.40	0.57	
	Neutral	0.65	0.87	0.74	
	Negative	0.88	0.93	0.90	

Table 1: Evaluation Results of Sentiment Analysis Models on manually labeled test data

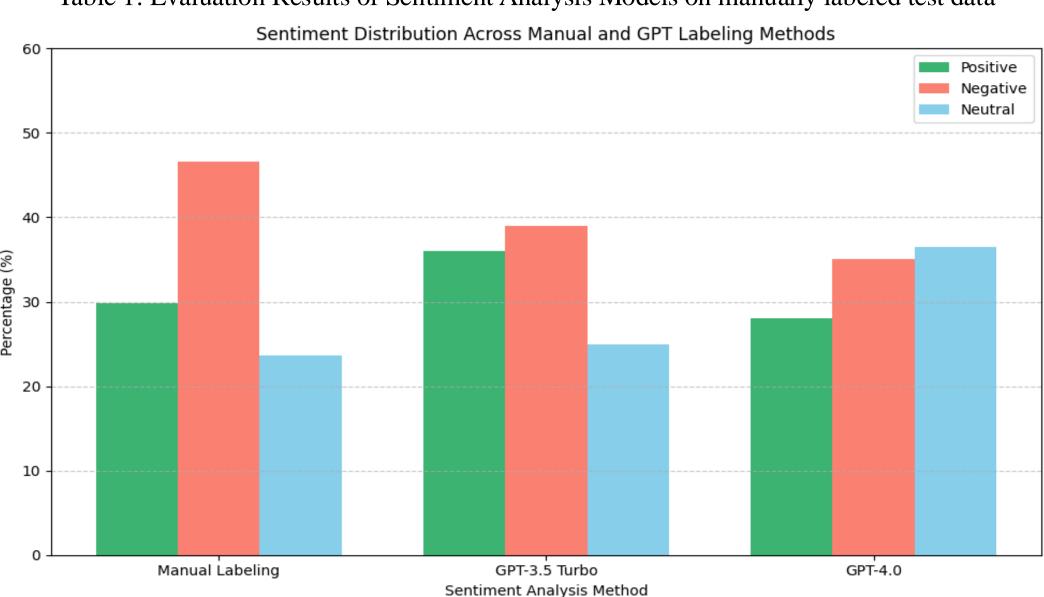


Figure 1: Sentiment distribution across different methods



Figure 2: LDA topic word cloud

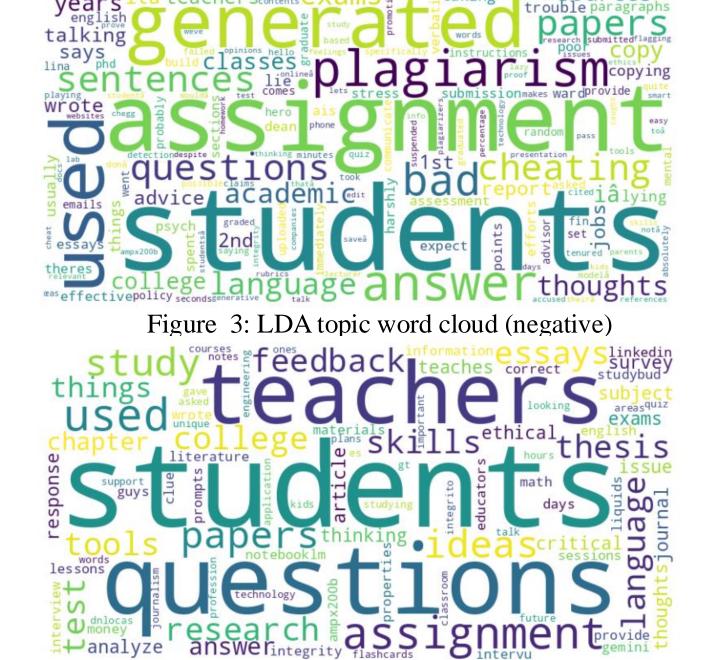


	Figure 4: LDA topic word cloud (positive)	
Overall Topics	issue, essay, plagiarism, used, teacher, detection, cheating, worried, lying, generate, writing, research, assignment, lazy, integrity, suspicious,copy, suspended, citation. harshly, caught. programing, report, policy, fail, verbatim, helpful, learned, flagging, penalty, ethical	
Positive Sentiment Topics	flashcards, answer, questions, notes, information, study, advise, growth, used, intrigued literature, optimize, perfect, notebooklm, studybud, support, essay, assignment, questions	
Negative Sentiment Topics	Cited, cheating, bad, copy, lying, used, verbatim, harshly, dean, plagiarism, generated, solution suspended, failed, denying, proof, accused, exams, issues, flagging, lazy, papers, caught	
Neutral Sentiment masters, research, problem, used, interesting, easier, homework, thesis, papers, de interesting, integrating, thoughts, opinions,		

Table 2: LDA Topic Modeling: Sentiment-based Insights

Results

Results from Topic Modeling:

Using LDA for topic modeling, we extracted dominant themes across the full dataset, identifying recurring concerns and discussion points around AI and education. Some of the most prominent keywords included: Issue, essay, plagiarism, used, teacher, detection, cheating, worried, lying, generate, writing, research, assignment, lazy, integrity, suspicious, copy, suspended, citation, harshly, caught, programming, report, policy, fail, verbatim, helpful, learned, flagging, penalty, ethical.

These words were grouped into broader domains that reflect underlying societal concerns:

- Academic Misconduct: Topics like plagiarism, cheating, lying, suspicious, citation, and verbatim highlight ethical fears about how students may misuse AI tools.
- Educational Challenges: Terms such as assignment, fail, lazy, teacher, and report suggest issues around learning and adapting to new tech in classrooms.

 Consequences and Detection: Words like flagging, suspended, caught, and
- penalty reflect institutional responses and enforcement of academic integrity.
 Mixed Perception: Topics including helpful, learned, used, and generate indicate that some users find educational value and support in AI use. When
- filtered by sentiment, LDA topics further revealed differences:

 Negative sentiment topics often centered around cheating, suspension, and
- Positive sentiment responses featured words like helpful, learned, and used,
- suggesting a recognition of AI's educational benefits.
 Neutral sentiment frequently reflected procedural or descriptive mentions—such as writing, assignment, or generate—without strong emotional weight.

Overall, the combination of sentiment analysis and topic modeling provides a well-rounded understanding of public perception, with strong concerns about ethics and misuse balanced by optimism about educational support and productivity.

Discussion

Findings:

This study examined Reddit users' perceptions of ChatGPT in education through LDA topic modeling and sentiment analysis. Prominent concerns included plagiarism, cheating, academic dishonesty, and ethical issues, while some users also acknowledged the tool's potential to support learning. Sentiment trends varied across models: manual labeling showed the highest rate of negative sentiment (46.6%), GPT-3.5 Turbo presented a more balanced distribution, and GPT-4.0 assigned the highest proportion of neutral sentiment (36.5%), suggesting a more cautious interpretation of ambiguous posts. These findings reflect a mix of skepticism and optimism, offering insight into how users engage with AI in educational settings and how sentiment models differ in their interpretations.

Limitations:

The scope of this study was limited to English-language Reddit posts from educational subreddits, which may not represent the broader global population. Data was collected exclusively from publicly available content, excluding perspectives from private discussions. Additionally, our analysis was constrained to using publicly available, pre-trained models, which may not be fine-tuned for the nuances of education-related dialogue. Despite these limitations, the study remains valid in providing a snapshot of real user concerns and adds meaningful contributions to the broader conversation around AI and education.

Future plans:

In future work, we plan to identify the primary concerns of Reddit users and quantify the frequency of each opinion to deepen our understanding of common sentiments. Rather than focusing solely on ChatGPT, we will broaden our scope to include other AI tools by incorporating additional keywords related to popular AI systems and technologies. We also aim to expand our analysis by scraping data from other domains such as business and journalism, allowing us to compare how concerns about AI differ across various contexts. To evaluate model performance, we will assess the accuracy of different sentiment analysis methods by comparing them to human-labeled data. Additionally, we plan to conduct surveys and interviews with individuals from diverse fields to gather direct insights into their experiences and concerns. This next phase will help us build a more comprehensive understanding of AI's broader societal impact.

Summary and Conclusions

Our research offers several key contributions to the understanding of how users perceive AI tools, particularly ChatGPT, in educational contexts.

Development of Annotation Guidelines:

By manually reviewing and labeling 200 Reddit posts, we developed a clear set of annotation guidelines to consistently categorize sentiment labels. These guidelines are useful for future researchers working with similar social media data.

Manually Labeled Dataset

The manually labeled dataset was instrumental in shaping our analysis, serving both as a reference point for evaluating the accuracy of automated models and as a source of deep qualitative insight. It enabled us to ground our sentiment findings in human judgment and measure where machine models align or diverge.

Model Comparison and Performance:

Our study revealed that sentiment classifications varied significantly between models. Manual labeling showed the highest rate of negative sentiment (46.6%), GPT-3.5 Turbo yielded a more balanced output, while GPT-4.0 demonstrated a more neutral stance (36.5%). These differences point to the unique characteristics and limitations of each model when interpreting nuanced or ambiguous content.

Findings from Sentiment and Topic Modeling:

LDA topic modeling uncovered consistent themes of academic misconduct, plagiarism, and ethical concerns, alongside positive mentions of learning assistance and productivity. While positive sentiments reflect enthusiasm about ChatGPT's potential, negative and neutral posts emphasize real anxieties around academic integrity, overreliance on AI, and diminished critical thinking.

Real-World Relevance:

This work helps contextualize the social implications of AI adoption in education. Understanding public sentiment provides institutions, developers, and policymakers with valuable feedback that can inform responsible AI development, ethical guidelines, and educational strategies. As we move forward, our research will extend beyond ChatGPT to include other generative AI tools and explore user attitudes in fields such as business and journalism.

We encourage the research community to build upon our dataset, methods, and findings. Greater collaboration in this space can help address societal concerns, promote ethical innovation, and ensure that AI technology evolves in ways that serve and protect its users.

Acknowledgements:

We extend our heartfelt gratitude to Dr. Shebuti Rayana, and Nicholas Eversgerd who have contributed to the inception and execution of this project. We acknowledge their invaluable contribution in formulating the idea, curating the data, labeling data, and running the initial experiments. Their insights and analyses are instrumental in shaping the direction of this work.

Funding:

Funded by PRODiG-DEFT Micro-Grant 2025