WIP: From Tweets to Trends: Tracing the Public's Perception of AI in Education Post-ChatGPT

1st Fan Zhang
University of Florida
USA
f.zhang1@ufl.edu

2nd Rui Guo
University of Florida
USA
rui.guo@ufl.edu

3rd Wanli Xing
University of Florida
USA
wanli.xing@coe.ufl.edu

4th Wangda Zhu
University of Florida
USA
wangdazhu@ufl.edu

5th Taehyun Kim
University of Florida
USA
kim.taehyun@ufl.edu

6th Zifeng Liu
University of Florida
USA
liuzifeng@ufl.edu

***Abstract*—This study examines public sentiment towards AI in education, focusing on the impact of ChatGPT's launch by OpenAI on November 30, 2022. Analyzing around 80,000 Twitter posts from before and after the launch, we conducted a comprehensive sentiment analysis using a fine-tuned BERT, outperforming traditional methods such as VADER and SVM. We applied an RDD to assess the causal impacts of ChatGPT's introduction on public sentiment track sentiment shifts, highlighting how the introduction of AI technologies like ChatGPT has influenced educational discourse. Our findings reveal significant public sentiment changes post-launch, contributing new insights into AI's role in education and public discourse.**

Keywords—Public Sentiment, ChatGPT Launch, Sentiment Analysis, AI in Education

I. Introduction

The role of technology in modern education is increasingly complex, driven by the growing reliance on digital tools and artificial intelligence (AI). AI, representing the pinnacle of advancements in computer and communication technologies, enables machines to mimic human functions (Chen et al., 2020). This has significantly transformed learning, making it more engaging and accessible, especially for younger students (Lachhwani, 2022). The integration of AI into education offers benefits such as personalized learning, streamlined operations, and enhanced feedback (Harry, 2023; Zaman, 2023). However, these advancements bring concerns related to privacy, security, and biases within AI systems (Dignum, 2021; Lampou, 2023).

ChatGPT, developed by OpenAI, exemplifies AI's impact in education. Since its launch in November 2022, it quickly became a widely used tool, prompting educators to assess its utility. For example, it has been employed to create lesson plans and educational materials, although its rapid adoption has also raised issues such as AI-assisted cheating and potential risks to academic integrity (Anders, 2023; Sok & Heng, 2023). Ongoing research is focused on optimizing AI's role in education while addressing these challenges. Despite extensive studies on AI's benefits and drawbacks, understanding the perspectives of those affected by these technologies remains crucial to ensuring their effective and equitable integration (Mitchell, 2016; Relucio & Palaoag, 2018).

Social media serves as a vital platform for educational stakeholders to voice their opinions and visions regarding AI in education. It provides a rich, time-stamped repository of community sentiment (Wang & Fikis, 2019), serving as a unique resource for gauging how educational communities perceive AI and the impact of these perceptions. Increased engagement on platforms like Twitter by teachers (Carpenter & Krutka, 2014), school principals, and superintendents (Cho, 2016; Cox & McLeod, 2014), along with educational institutions, facilitates the exchange of ideas and collective sentiments about AI technologies. Nevertheless, research into these sentiments and their evolution on social media is relatively limited. This study aims to analyze how the sentiment of educational communities towards AI has shifted, particularly focusing on the period before and after ChatGPT's launch, which has significantly influenced the rapid and widespread adoption of AI in educational settings (Whalen & Mouza, 2023).

II. Related Work

Sentiment analysis of Twitter data has emerged as a crucial area of research due to the platform's widespread use for expressing opinions on various topics, including education and AI. Agarwal et al. (2011) demonstrated that sentiment analysis on Twitter requires specialized techniques due to the informal nature of the language and the brevity of the posts. They proposed using a combination of unigram models, feature-based models, and tree kernels to classify tweets into positive, negative, and neutral sentiments. Their work highlighted the importance of Twitter-specific features, such as emoticons and hashtags, in improving classification accuracy.

Building on these approaches, Kumar and Sebastian (2012) explored a hybrid sentiment analysis method that integrates both corpus-based and dictionary-based techniques. This approach is particularly effective in handling the varied linguistic expressions found in Twitter data. They emphasized the potential of sentiment analysis in applications ranging from business intelligence to political monitoring, demonstrating its versatility and relevance across different domains. Their work also noted the challenges posed by noisy and biased data, common in social media platforms like Twitter.

Further advancements in sentiment analysis have focused on improving the accuracy and efficiency of these methods. Researchers have increasingly leveraged machine learning models such as Naive Bayes, Maximum Entropy, and Support Vector Machines (SVM) to enhance sentiment classification. Go et al. (2009) employed distant supervision by using emoticons as noisy labels to automatically annotate large datasets of tweets, finding that SVM models outperformed others in terms of accuracy. This method of leveraging weak supervision has proven to be a robust approach in sentiment analysis, particularly in the context of massive and unstructured Twitter data.

III. METHODS

Workflow

Figure 1 illustrates the workflow of our experiment. Initially, we gathered relevant data and enlisted research assistants with expertise in education to manually label 1,000 data points for sentiment, categorizing the posts as positive, negative, or neutral. In the second phase, these 1,000 manually labeled entries were used to train and evaluate several models, with VADER serving as a benchmark. A comprehensive assessment using metrics such as AUC, accuracy, and F1 score was conducted to identify the optimal model for sentiment analysis of all posts (a detailed description of the sentiment analysis methods is provided in the "Sentiment Analysis Methodology" section below). Subsequently, the processed posts were subjected to RDD for causal impact analysis (step 3). In future work, we plan to incorporate LDA time-series analyses (step 4) to further enhance our understanding of sentiment dynamics over time.

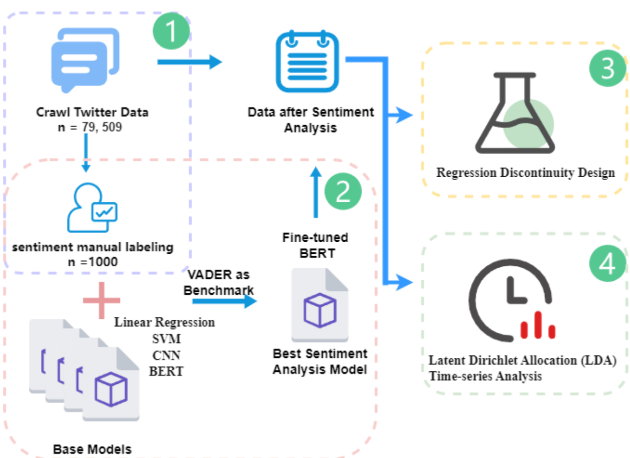

Fig 1 Workflow

Data Collection

The study aimed to assess public sentiment towards artificial intelligence (AI) in education, focusing on the significant impact of ChatGPT's launch by OpenAI on November 30, 2022. To achieve this, discussions and comments about AI in education from Twitter were collected over an extensive period, starting in January 2022 and ending in December 2023, amassing approximately 79,509 posts. The Apify Twitter Scraper API tool facilitated the efficient scraping of Twitter data related to "AI in education." To ensure relevance to our study goals, search terms related to both "AI" and "education" were intersected. From the collected dataset, a random sample of 1,000 posts was extracted for a preliminary check to confirm their pertinence to AI in education, setting the stage for subsequent sentiment analysis. This validation step affirmed the relevance of the posts to the AI-education domain, achieving a consistency rate of 96.7%. Through this dataset, we sought to explore deeply the evolution of sentiment regarding AI's role in education before and after ChatGPT's introduction to the public. This timeline allows for a nuanced analysis of how perceptions may have shifted due to the widespread discussion and implementation of ChatGPT in educational contexts. By leveraging advanced sentiment analysis techniques, the study intends to unveil trends, patterns, and significant shifts in public opinion on this transformative technological integration into education.

Sentiment Analysis

Sentiment analysis, also known as opinion mining, is a research area focused on evaluating people's sentiments, attitudes, and emotions regarding specific topics, products, or services expressed through written language (Medhat et al., 2014). The primary aim of sentiment analysis is to decipher the subjective aspects of textual data, enabling organizations, marketers, and researchers to assess public opinion, track brand and product sentiment, and understand customer needs and concerns. By leveraging natural language processing (NLP), machine learning, and computational linguistics, sentiment analysis classifies opinions in text as positive, negative, or neutral, among other detailed emotional states (Wankhade et al., 2022). This technique provides decision-makers with actionable insights from social media, reviews, forums, and other forms of user-generated content, thus supporting informed strategies in marketing, product development, customer service, and policymaking.

In this study, we employed various models for sentiment analysis, including Linear Regression, SVM, CNN, VADER, and BERT. We compared these models to identify the most optimal approach for our final analysis. Table 1 summarizes these methods, offering a brief introduction and citation for each.

TABLE 1 SENTIMENT ANALYSIS MODEL DESCRIPTION

Model	*Introduction*	*Citation*
Linear Regression	A statistical method used to predict sentiment scores based on text features, offering a straightforward approach but may not capture complex language nuances.	Cakra and Trisedya (2015)
SVM	Utilizes a hyperplane in an N-dimensional space to distinctly classify sentiment categories, effective for clear-margin datasets.	Ahmad et al. (2018)
CNN	Applies spatial hierarchy capturing capabilities of neural networks to identify text patterns for sentiment	

Model	*Introduction*	*Citation*
	classification, excelling in local context recognition.	Liao et al. (2017)
VADER	A lexicon and rule-based tool specifically designed for sentiment analysis in social media, offering quick analysis without extensive training data.	Elbagir and Yang (2019)
BERT	A transformer-based model that revolutionizes NLP by pre-training on bidirectional context, highly effective for nuanced sentiment analysis.	Xu et al. (2019)

RDD Framework

Upon analyzing sentiments extracted from posts using sentiment analysis tools, we applied the Regression Discontinuity Design (RDD) for a comprehensive evaluation. RDD is a quasi-experimental method that estimates the causal impact of an intervention by utilizing a pre-established cutoff point (Maciejewski & Basu, 2020). In our investigation of ChatGPT's launch on public sentiment towards AI in education, RDD employs the launch date as a natural cutoff, dividing the data into pre-launch and post-launch intervals. By comparing sentiment scores adjacent to this cutoff, RDD mimics a randomized controlled trial, providing a robust basis for causal inferences. Through precise determination of a bandwidth around the cutoff and the application of regression models to measure the sentiment shift at this point, RDD enables a sophisticated examination of the impact of significant technological introductions like ChatGPT on public sentiment. This approach is crucial in isolating the direct effects of ChatGPT's launch on sentiments regarding AI in education, highlighting changes in public attitudes with potentially far-reaching implications for the evolution of educational technology and policy development.

IV. RESULTS

A. *Sentiment Analysis Results*

In benchmarking sentiment analysis, as illustrated by the performance metrics in Table 2, the fine-tuned BERT model emerged as notably superior. This model not only demonstrated the highest accuracy (ACC) and micro-averaged area under the curve (MICRO AUC) but also excelled in precision, recall, and F1 score. These results underscore BERT's advanced capacity for context-sensitive sentiment analysis, setting a high standard for accuracy and nuance in detecting sentiment within the specific context of AI's impact on education. Further analysis of examples highlights the fine-tuned BERT model's unique ability to accurately parse and analyze sentiments related to the prospects and descriptions of AI in education within individual statements. Unlike more generic sentiment analysis, which might assess the overall sentiment of an entire statement, the fine-tuned BERT model adeptly focuses on specific segments relevant to AI and education (Due to space constraints, we did not include detailed examples or the comparative results of the model classifications and manual classifications). This targeted approach allows for a more nuanced sentiment analysis that aligns closely with the user's intent, capturing both the optimism and concerns expressed about the role of AI in transforming educational paradigms.

TABLE 2 BENCHMARKING SCORE OF EACH ANALYSIS METHOD

	Linear Regression	*SVM*	*CNN*	*Vader*	*Finetuned Bert Results*
ACC	0.63	0.61	0.62	0.54	**0.68**
MICRO AUC	0.76	0.75	0.74	0.54	**0.81**
Precision	0.59	0.49	0.41	0.48	**0.63**
Recall	0.50	0.46	0.44	0.46	**0.60**
F1 Score	0.51	0.47	0.43	0.43	**0.61**

B. *Identify the Headings*

Figure 2 illustrates the changes in the volume of discussions about AI in education and their sentiment over time, with a specific focus on the period surrounding the launch of ChatGPT. The data reveals a significant increase in discussion volume about AI in education, rising from an average of approximately 1,000 posts per month to an average of 4,000 to 5,000 posts per month following the deployment of ChatGPT. However, the proportion of positive sentiment posts about AI in education remained consistently around 70% before the launch of ChatGPT, except for September 2022. In the month of ChatGPT's launch, this figure increased to nearly 80% but subsequently returned to the pre-launch level of 70%. To further explore these observations, we employed Regression Discontinuity Design (RDD). This methodology allowed us to examine the causal impact of ChatGPT's launch on sentiment and thematic evolution within AI education discussions, providing a structured and nuanced understanding of how this pivotal event influenced public discourse over time.

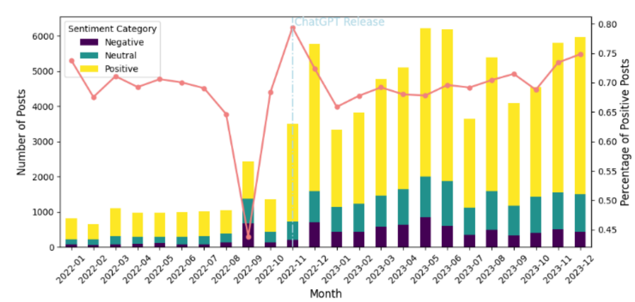

Fig 2 Monthly Distribution of Sentiment Categories

C. ***RDD analysis***

In our analysis, we conducted an Ordinary Least Squares (OLS) regression to investigate the impact of a specific variable on our outcome of interest. The fit of this model is visually represented in Figure 3, which illustrates how well our regression line captures the underlying data points. Additionally, the results of the Regression Discontinuity Design (RDD) analysis are detailed in Table 3.

TABLE 3 OLS REGRESSION RESULTS

	coefficient	*standard error*	*t value*	*P>|t|*
Intercept	0.5232	0.017	31.380	0.000
days_from_launch	-0.0006	0.001	-0.606	0.547
post_launch	-0.1083	0.032	-3.406	0.001

The OLS regression results indicate that the deployment of ChatGPT had a significant impact on sentiment towards AI in education. The model, with an R-squared of 0.533, explains approximately 53.3% of the variability in sentiment scores, suggesting a moderate fit. The statistically significant negative coefficient for post_launch (-0.1083, $p < 0.001$) implies that the deployment of ChatGPT is associated with a decrease in sentiment scores by approximately 0.108 units. This result is robust, with a 95% confidence interval ranging from -0.172 to -0.044, indicating a reliable negative effect. The days_from_launch variable, however, does not show a significant trend over time ($p = 0.547$), suggesting that the observed effect is specifically associated with the event of ChatGPT's launch rather than a general time trend. The model's diagnostics, including the Omnibus and Jarque-Bera tests, indicate some deviation from normality, which should be considered when interpreting the results. Overall, the analysis suggests that the introduction of ChatGPT had a measurable negative impact on public sentiment regarding AI in education, as captured through the sentiment scores in the period surrounding its launch.

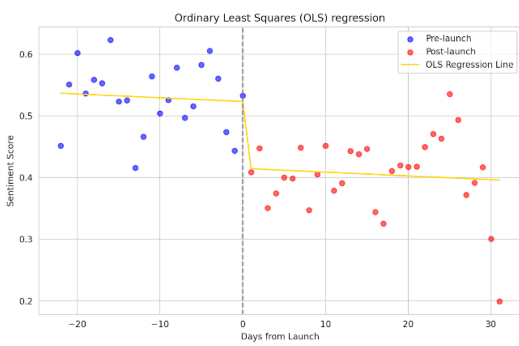

Fig 3 Ordinary Least Squares (OLS) Regression

V. DISCUSSION

In this paper, we explore the impact of ChatGPT's release on public sentiments regarding the educational use of AI. Initially, we conducted a benchmark analysis to identify the most appropriate sentiment analysis model for AI in education. Our findings revealed that the fine-tuned BERT model provided the most accurate and contextually relevant results. This model's refined analysis capability is particularly evident in its consistent performance across discussions—whether addressing the slow pace of change in formal education systems, the potential of e-learning and technological advances, or the need for educational systems to better prepare students for an AI-driven future. BERT's fine-tuned version reliably identifies and evaluates sentiments specific to the integration and impact of AI in educational contexts, demonstrating significant advancement over traditional models and automated tools such as VADER, which may exhibit variability in sentiment detection accuracy and struggle to capture nuanced perspectives on AI's role in education.

As shown in Figure 2, prior to the launch of ChatGPT, the number of posts per month about AI in education typically hovered around 1,000. However, following the launch, this number increased dramatically, reaching at least 3,000 posts per month, with peaks as high as 6,000. This substantial increase indicates that ChatGPT's introduction has significantly amplified discussions about AI in education on platforms such as Twitter (now X). Before ChatGPT, AI was already utilized in educational settings, including adaptive learning systems like DreamBox Learning, AI-based language learning applications like Duolingo, and automated essay scoring systems using NLP technology like ETS's e-rater. This indicates that, although AI was not entirely new, the launch of ChatGPT has notably intensified public discussions about its role in education. Furthermore, the sustained high volume of posts over the year following the launch suggests that ChatGPT has effectively engaged educational stakeholders directly. Capturing the interest of both the public and educational stakeholders is crucial for any technology to be widely applied in education. In this regard, ChatGPT has played a significant role in raising awareness and fostering discussions about AI in education.

We also analyzed the sentiments of these posts to understand how educational stakeholders perceive AI. By using ordinary least squares (OLS) regression analysis to compare the 20 days before and after the launch of ChatGPT (see Figure 4), we discovered that posts following the launch tended to contain more negative sentiments about AI. However, when considering Figure 2, we observed that the proportion of positive posts surged to nearly 80% in November 2022, the month of ChatGPT's launch. It then declined for two consecutive months, followed by a gradual increase in positive sentiments. This pattern closely resembles the Gartner Hype Cycle (Linden & Fenn, 2003). Initially, as the technology was introduced (November 2022), positive sentiments peaked due to high expectations and interest. As limitations and problems became apparent, disappointment set in (December 2022–January 2023). Eventually, as issues were addressed and the technology's value and usefulness began to be recognized, it moved toward broader acceptance (from January 2023 onward).

For future work, we plan to introduce large language models(LLMs) such as Llama to further enhance the accuracy of sentiment analysis. By leveraging Llama's capabilities, we aim to capture even more nuanced insights into public sentiments. Additionally, we intend to employ time-series Latent Dirichlet Allocation (LDA) to extract and analyze themes from the posts over time. This approach will allow us to identify and separate distinct topics, providing a deeper understanding of the evolving discourse around AI in education.

REFERENCES

[1] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. J. (2011, June). Sentiment analysis of twitter data. In *Proceedings of the workshop on language in social media (LSM 2011)* (pp. 30-38).

[2] Ahmad, M., Aftab, S., Bashir, M. S., & Hameed, N. (2018). Sentiment analysis using SVM: A systematic literature review. International Journal of Advanced Computer Science and Applications, 9(2).

[3] Anders, B. A. (2023). Is using ChatGPT cheating, plagiarism, both, neither, or forward thinking?. Patterns, 4(3).

[4] Cakra, Y. E., & Trisedya, B. D. (2015, October). Stock price prediction using linear regression based on sentiment analysis. In 2015 international conference on advanced computer science and information systems (ICACSIS) (pp. 147-154). IEEE.

[5] Carpenter, J. P., & Krutka, D. G. (2014). How and why educators use Twitter: A survey of the field. Journal of Research on Technology in Education, 46, 414-434.

[6] Chandrasekaran, S. (2014). Factors Affecting to Technology Adoption by Teachers. International Journal of Humanities and Social Science Invention. 3(9), 03-09.

[7] Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. Ieee Access, 8, 75264-75278.

[8] Chen, X., Xie, H., & Hwang, G. J. (2020). A multi-perspective study on artificial intelligence in education: grants, conferences, journals, software tools, institutions, and researchers. Computers and Education: Artificial Intelligence, 1, 100005. https://doi.org/10.1016/j.caeai.2020.100005

[9] Cho, V. (2016). Administrators' professional learning via Twitter: The dissonance between beliefs and actions. Journal of Educational Administration, 54, 340-356.

[10] Cox, D., & McLeod, S. (2014). Social media marketing and communications strategies for school superintendents. Journal of Educational Administration, 52, 850-868.

[11] Dignum, V. (2021). The role and challenges of education for responsible AI. London Review of Education, 19(1), 1-11.

[12] Elbagir, S., & Yang, J. (2019, March). Twitter sentiment analysis using natural language toolkit and VADER sentiment. In Proceedings of the international multiconference of engineers and computer scientists (Vol. 122, No. 16). sn.

[13] Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N project report, Stanford*, *1*(12), 2009.

[14] Harry, A. (2023). Role of AI in Education. Interdiciplinary Journal and Hummanity (INJURITY), 2(3), 260-268.

[15] Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artifcial intelligence in education. Computers and Education: Artifcial Intelligence, 1, 100001. https://doi.org/10.1016/j.caeai.2020.100001

[16] Justus, M. (2017). The role of pedagogical beliefs in emerging technology integration: An exploratory case study of faculty perspectives. The Qualitative Report, 22(2), 499-526.

[17] Kumar, A., & Sebastian, T. M. (2012). Sentiment analysis on twitter. *International Journal of Computer Science Issues (IJCSI)*, *9*(4), 372.

[18] Lachhwani, V. (2022). Role of Information Technology in Education Sector: A Review. Management Journal for Advanced Research, 2(6), 12-15.

[19] Lampou, R. (2023). The integration of artificial intelligence in education: opportunities and challenges. Review of Artificial Intelligence in Education, 4(00), e015-e015.

[20] Liao, S., Wang, J., Yu, R., Sato, K., & Cheng, Z. (2017). CNN for situations understanding based on sentiment analysis of twitter data. Procedia computer science, 111, 376-381.

[21] Linden, A., & Fenn, J. (2003). Understanding Gartner's hype cycles. Strategic Analysis Report Nº R-20-1971. Gartner, Inc, 88, 1423.

[22] Mitchell, G. W., Wohleb, E. C., & Skinner, L. B. (2016). Perceptions of public educators regarding accessibility to technology and the importance of integrating technology across the curriculum. Journal of research in business education, 57(2), 14-25.

[23] Relucio, F. S., & Palaoag, T. D. (2018, January). Sentiment analysis on educational posts from social media. In Proceedings of the 9th international conference on E-education, E-business, E-management and E-learning (pp. 99-102).

[24] Singer, N. (2023, February 6). At this school, computer science class now includes critiquing chatbots. The New York Times. https://www.nytimes.com/2023/02/06/technology/chatgpt-schoolsteachers-ai-ethics.html

[25] Sok, S., & Heng, K. (2023). ChatGPT for education and research: A review of benefits and risks. Available at SSRN 4378735.

[26] Wang, Y., & Fikis, D. J. (2019). Common core state standards on Twitter: Public sentiment and opinion leaders. Educational Policy, 33(4), 650-683.

[27] Whalen, J., & Mouza, C. (2023). ChatGPT: Challenges, opportunities, and implications for teacher education. Contemporary Issues in Technology and Teacher Education, 23(1), 1-23.

[28] Xu, H., Liu, B., Shu, L., & Yu, P. S. (2019). BERT post-training for review reading comprehension and aspect-based sentiment analysis. arXiv preprint arXiv:1904.02232.

[29] Xu, W., & Ouyang, F. (2022). The application of AI technologies in STEM education: a systematic review from 2011 to 2021. International Journal of STEM Education, 9(1), 59.

[30] Zaman, B. U. (2023). Transforming Education Through AI, Benefits, Risks, and Ethical Considerations. Authorea Preprints.

[31] Zampirolli, F. A., BorovinaJosko, J. M., Venero, M. L. F., Kobayashi, G., Fraga, F. J., Goya, D., & Savegnago, H. R. (2021). An experience of automated assessment in a large-scale introduction programming course. Computer Applications in Engineering Education, 29(5), 1284–1299. https://doi.org/10.1002/cae.22385

[32] Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. International Journal of Educational Technology in Higher Education, 16(1), 1-27.

[33] Zhai, X. (2022). ChatGPT user experience: Implications for education. Available at SSRN 4312418.