Exploring Public Opinions Toward the Use of Generative Artificial Intelligence Chatbot in Higher Education: An Insight from Topic Modelling and Sentiment Analysis

Sarsam, S., Alias, A. A., Su Mon, C., Al-Samarraie, H. & Al-Hatem, A. I

Original citation & hyperlink:
Sarsam, S, Alias, AA, Su Mon, C, Al-Samarraie, H & Al-Hatem , AI 2023, Exploring Public Opinions Toward the Use of Generative Artificial Intelligence Chatbot in Higher Education: An Insight from Topic Modelling and Sentiment Analysis. in 2023 International Conference on Big Data, Knowledge and Control Systems Engineering (BdKCSE). IEEE, pp. 1-6, 2023 International Conference on Big Data, Knowledge and Control Systems Engineering , Sofia, Bulgaria, 2/11/23.
https://doi.org/10.1109/BdKCSE59280.2023.10339760

DOI 10.1109/BdKCSE59280.2023.10339760
ISBN 9798350313253
ESBN 9798350313246

Publisher: [Publisher]

Exploring Public Opinions Toward the Use of Generative Artificial Intelligence Chatbot in Higher Education: An Insight from Topic Modelling and Sentiment Analysis

Samer Muthana Sarsam
School of Strategy and Leadership Coventry University, Coventry, UK
samer.sarsam@coventry.ac.uk (ORCID: 0000-0002-7404-1150)

Abd Aziz Alias
Department of Business Analytics, Sunway University, Business School Sunway University, Petaling Jaya Selangor, Malaysia
aliasa@sunway.edu.my

Chit Su Mon
School of Mathematical and Computer Sciences, Heriot-Watt University Malaysia, Putrajaya, Malaysia
c.mon@hw.ac.uk

Hosam Al-Samarraie
School of Design, University of Leeds, Leeds, UK
h.alsamarraie@leeds.ac.uk

Ahmed Isam Al-Hatem
School of Computer Engineering, Al-Nukhba University College, Baghdad, Iraq
a.essam@alnukhba.edu.iq

Corresponding Author: Samer Muthana Sarsam
samer.sarsam@coventry.ac.uk

***Abstract*—The Generative Artificial Intelligence chatbots (GAI chatbots) have emerged as promising tools in various domains, including higher education, so this study aims to investigate the role of Bard, a newly developed GAI chatbot, in higher education. English tweets were collected from Twitter's free streaming Application Programming Interface (API). The Latent Dirichlet Allocation (LDA) algorithm was applied to extract latent topics from the tweets. User sentiments were extracted using the NRC Affect Intensity Lexicon and SentiStrength tools. This study explored the benefits, challenges, and future implications of integrating GAI chatbots in higher education. The findings shed light on the potential power of such tools, exemplified by Bard, in enhancing the learning process and providing support to students throughout their educational journey.**

Keywords— Generative Artificial Intelligence chatbots, Bard, higher education, topic modelling, sentiment analysis

I. INTRODUCTION

Generative Artificial Intelligence (GAI) chatbots are advanced chatbot systems that use generative models to respond to user inputs in a human-like manner. GAI chatbots have the potential to generate interactive dialogues by establishing chatbots that simulate conversations with native speakers. They use different techniques to generate text—either generative pre-trained transformer (GPT) chatbot or large language models (LLMs).

The release of Open AI's conversational AI text-writing tool ChatGPT, an example of generative pre-trained transformer (GPT) chatbot, in November 2022 sparked a tempest in technology and business circles, revealing hitherto unseen prospects, worries, and obstacles (Murugesan & Cherukuri, 2023). Its achievement underlines the significance of multiple elements, including data volume, GPU count, and powerful pre-trained models. These elements have been critical in the transition from traditional learning frameworks to the new ChatGPT-like learning framework (Wang, Miao, Li, Wang, & Lin, 2023). Generative Pre-trained Transformer 3 (GPT-3) and other generative AI (artificial intelligence) chatbots receive user inputs, interpret the context, and generate responses word by word. To ensure that the generated responses are coherent and relevant to the discourse, these models employ sophisticated algorithms and neural architectures to learn the statistical patterns and language structures included in the training data during the training phase, allowing them to generate human-like responses.

Recently, Google has introduced a new GAI chatbot that is called "Bard", an example of LLMs. Google Bard was first launched on February 6, 2023. It was made available to a limited number of users in the United States and the United Kingdom on March 21, 2023. On May 10, 2023, Google removed the waitlist and made Bard available in over 180 countries and territories. Bard generates product suggestions powered by state-of-the-art language models and sophisticated machine learning techniques. Such tools allow generate text, translate languages, write different kinds of creative content, and answer users' questions in an informative way. To do so, they rely on advanced techniques from Natural language processing (NLP), Machine learning, and Deep learning (e.g., Recurrent neural networks (RNNs), Convolutional neural networks (CNNs), and Transformers). Therefore, Bard can generate human-like answers in response to input prompts. It generates text using the grammar-based generation (unlike ChatGPT that uses attention-based generation) to produce text that is grammatically correct and relevant to the conversation. Bard understands a wide range of themes and conversational circumstances, enabling more meaningful and contextually appropriate conversations (Rahaman et al., 2023). It has potential to access to a vast knowledge collection covering a wide range of topics such as general knowledge, current affairs, entertainment, science, and more. It provides the user with up to date and reliable information on a wide range of topics (Plevris, Papazafeiropoulos, & Rios, 2023).

Social media provide a rich source for information that allows understanding individuals' activities and determine their topics and opinions from their experience (Di Minin, Tenkanen, & Toivonen, 2015). Precisely, social media refers to web-based services that permit its users to connect, interact, and establishing an online community by enabling them to create, modify, and engage with the resulting textual content in easy and flexible way. Twitter is the most popular social media platform that allows its users to publish their thoughts and viewpoints (Giachanou & Crestani, 2016). Twitter messages (tweets) contain individuals' opinions towards products, services, and events. This motivates several high school students to share their learning experience on Twitter and establish discussion about certain learning method or academic-related topics (Tang, Hew, & education, 2017).

In order to extract the embedded topics from the huge amounts of textual data, topic modelling algorithms are applied to large numbers of documents and showed high potential in finding interesting patterns even from social media data. Latent Dirichlet Allocation (LDA) is an example topic modelling algorithm that analyzes words from social media content and find themes in the corpus (Ostrowski, 2015). It produces the words with the highest trend and the topic with the highest term frequency. For instance, a study by Zhou, Kan, Huang, and Silbernagel (2023) effectively used LDA method to investigate temporal latent topics from tweets during a disaster event, the 2020 Hurricane Laura. Jelodar et al. (2019) proposed a natural language processing (NLP) model with LDA to extract detailed information Twitter data, such as affected individuals, donations and support, caution and advice about Hurricane Irma. Also, Kim and Shim (2014) proposed a recommendation system for Twitter using probabilistic modelling based on LDA that recommended top-K users to follow as well as top-K tweets to read for a user.

Understanding users' opinions from their own social media posts plays a significant role in understanding their viewpoint towards particular topics. Sentiment analysis gives useful information for organizations, researchers, governments, and individuals to make wise decisions (Waheeb, Khan, & Shang, 2022). To obtain such knowledge, sentiment analysis or opinion mining is technique to be used since it helps in detecting opinions of people on social media platforms like Twitter. For instance, a study by Paul, Li, Teja, Yu, and Frost (2017) showed the possibility of using sentiment analysis to understand individuals' opinions towards political matters like elections. Also, Liao, Wang, Yu, Sato, and Cheng (2017) proposed an approach to understand real-world events using sentiment analysis of Twitter data via deep learning techniques. The researchers were able to predict user satisfaction of a product, happiness with some particular environment or destroy situation after disasters.

In order to understand peoples' opinions towards the application GAI chatbots in higher education, this study aims to answer the following research questions:

- What are the main discussed topics on Twitter about the use of GAI chatbots in higher education?
- What are the sentiments of Twitter users towards GAI chatbots technology?

To answer these questions, English tweets were collected from Twitter's free streaming Application Programming Interface (API). The Latent Dirichlet Allocation (LDA) algorithm was applied to extract latent topics from the collected tweets. User sentiments, including disgust, surprise, sadness, anger, fear, joy, anticipation, and trust, as well as positive and negative sentiments, were extracted using the NRC Affect Intensity Lexicon and SentiStrength tools.

II. METHOD

A. Data collection

We retrieved 4637 English-language tweets from February 10, 2023, to May 1, 2023, using the Twitter free streaming Application Programming Interface (API). To ensure their relevance to our research objectives, we obtained the tweets using predetermined keywords such as 'Bard', 'higher education', 'google', and 'chatbot'.

B. Data pre-processing

To perform a reliable analysis, we applied several data pre-processing approaches. In this context, all repeated tweets (i.e., (Retweets)) were removed from the data. The "Tokenization" technique was implanted via term frequency–inverse document frequency (TF-IDF) to extract all relevant words from the collected tweets and subsequently generated a dictionary to serve as the foundation of our main corpus. The mention symbol (@), URLs, and hashtags were deleted to keep necessary post-related content. Also, all the tweets-related words were converted to the lowercase form to standardize their format across the entire corpus. Finally, the non-essential words were deleted from the corpus using the Stopwords list technique.

C. Topic modelling

To extract the hidden topics from the collected tweets, the Latent Dirichlet Allocation (LDA) algorithm was applied based on the recommendation of Ostrowski (2015). LDA is an unsupervised method that models each document as a combination of topics to produce summaries of topics with a discrete probability distribution over words per topic and estimate per-document discrete distributions over topics.

In this study, we configured LDA by setting its alpha and beta parameters at 0.5 and 0.01, respectively. Then, the perplexity measure was used to determine the best LDA result. After extracting the topics, three social science experts were invited to provide an individual decision about each topic by reading

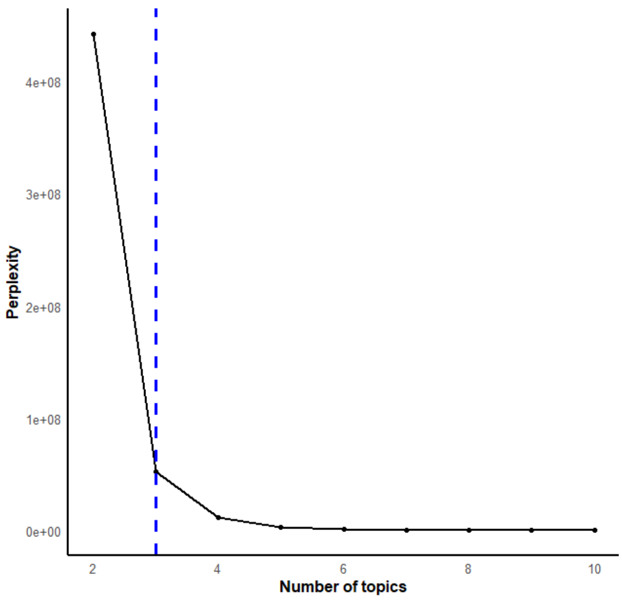

Fig. 1: Perplexity measure across 10 topic models (2-10). The continuous line represents the smoothed exclusivity curve. The dotted line represents the selected cutoff (three topics).

related tweets. Then, appropriate themes were provided to the tweets that share similar topics. Finally, to validate experts' decisions, the kappa statistic technique was used as a popular agreement approach.

D. Sentiment analysis

At this stage, users' sentiments were extracted from their tweets via two main approaches: "NRC Affect Intensity Lexicon" (Mohammad, 2017) and "SentiStrength" (Thelwall, 2017) using Waikato Environment for Knowledge Analysis (WEKA) software.

We obtained several types of emotions (disgust, surprise, sadness, anger, fear, joy, anticipation, and trust) using NRC Affect Intensity Lexicon that includes a list of English words and their associations with the embedded emotions. Also, tweets' polarity (positive and negative) was extracted via "SentiStrength" which is a lexicon- and rule-based sentiment analysis method that uses a group of rules to integrate the sentiment of words into the sentiment of the overall text.

III. RESULTS

A. Results of topic modelling

To determine the number of topics to be examined, we evaluated the performance of LDA algorithm using the perplexity measure. Perplexity is a valuable measure used in topic modelling, particularly with LDA, to assess the quality and coherence of the generated topics. To do so, the corpus is divided into a training set and a test set. The training set is used to estimate the LDA model, while the test set is used to evaluate the model's performance. After training LDA on the training set, we computed the heldout likelihood using the trained LDA model and the test set. The heldout likelihood represents how well the model predicts the unseen documents in the test set which shows how well the LDA model generalizes to unseen data. Perplexity is derived from the heldout likelihood. It is commonly computed as the exponential of the negative average log-likelihood per word in the test set. The perplexity result is summarized in Fig. 1.

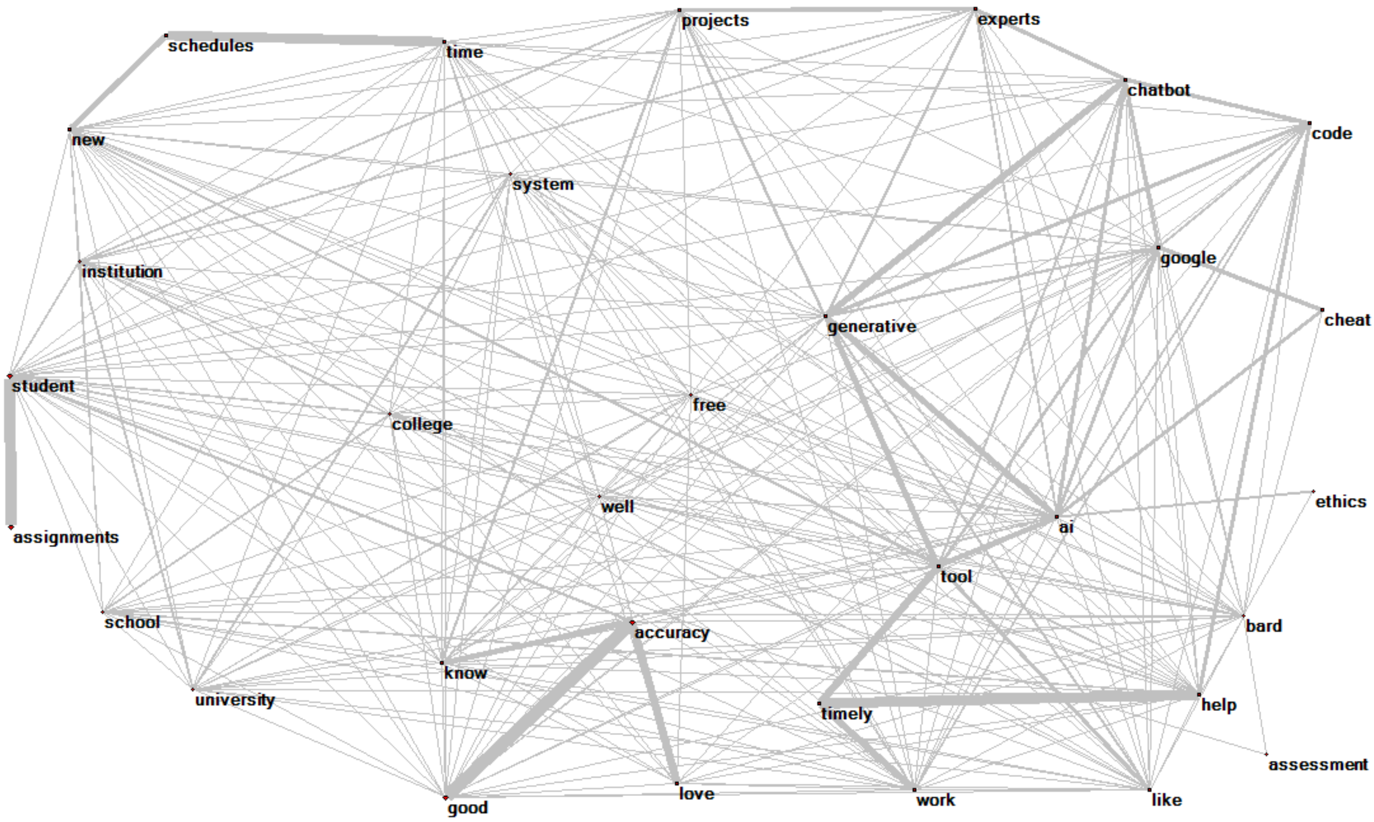

Fig. 2: Network model of relations between words (bigrams) from the whole corpus of tweets.

From this figure it can be observed that the perplexity values start to level off show diminishing improvement; hence, three topics were considered to be analysed. Finally, the authors validated the topics and assessed their interpretability. As a result, we found three main latent topics which are: ‘Applications of Bard to students’, ‘Bard challenges in higher education’, and ‘The future of Bard in higher education’.

First topics mainly focused on the potential applications of using Bard among the students. The tweets highlighted some of Bard’s support in learning, including personalized tutoring help in providing customized responses in association with the submitted query, provide a timely answers to their questions with high-level of accurately, and helping university students to stay on track with their studies by creating study schedules that are tailored to a student's individual needs, which helps students to stay organized and to make the most of their time.

The second topic, focused on discussing number of ethical challenges that could arise in the use of Bard in higher education, including the potential use of generating misleading or bias repones either intentionally or unintentionally. Furthermore, Bard could be used to collect personal data about students, such as their academic performance, their interests, and their personal beliefs. In addition, some tweets stated that Bard could be used to cheat on tests or assignments by generating answers that are not the student's own work such as generate essays or code that a student could then submit as their own work.

The third topic focused on the future of Bard in higher education. Our analysis showed that many people expressed their excitement to see how Bard will be used in higher education in the future. They believe that Bard has the potential to revolutionize the way we teach and learn in different aspects, such as providing access to experts by allowing students to ask questions to experts in any field via direct connection. Also, a part of the future usage of Bard, it could be used to create new forms of assessment by allowing students to demonstrate their knowledge and skills in new and creative ways. For example, a student could be asked to write a poem or a song that demonstrates their understanding of a particular concept, and Bard could then be used to evaluate the student's work.

To further understand the content of the data we used the network model (see Fig. 2), which shows the relations between words (bigrams) from the whole corpus of tweets. The width of the connecting lines is proportional to the frequency of co-occurrences between words (i.e., a thicker line indicates a more frequent word co-occurrence).

B. Results of sentiment analysis

The overall sentiment analysis result is summarized in Fig. 3. From this figure, it can be observed that trust (**M** = 0.63, **SD** = 0.05), anticipation (**M** = 0.51, **SD** = 0.11), and joy (**M** = 0.39, **SD** = 0.08), respectively, are the dominant emotions. Also, fear (**M** = 0.28, **SD** = 0.02) and sadness (**M** = 0.27, **SD** = 0.06) were found to be higher than anger (**M** = 0.23, **SD** = 0.02), surprise (**M** = 0.20, **SD** = 0.09), and disgust (**M** = 0.14, **SD** = 0.07), respectively. Besides, our sentiment analysis results revealed the percentage of positivity (75.45 %) is higher than negativity (24.55 %) among the collected tweets. From previous results, on can observe that the positive polarity is dominant among the collected tweets. Also, these tweets have high levels of trust, anticipation, and joyful emotions. This means that Twitter users have a high level of satisfaction when using the Bard tool.

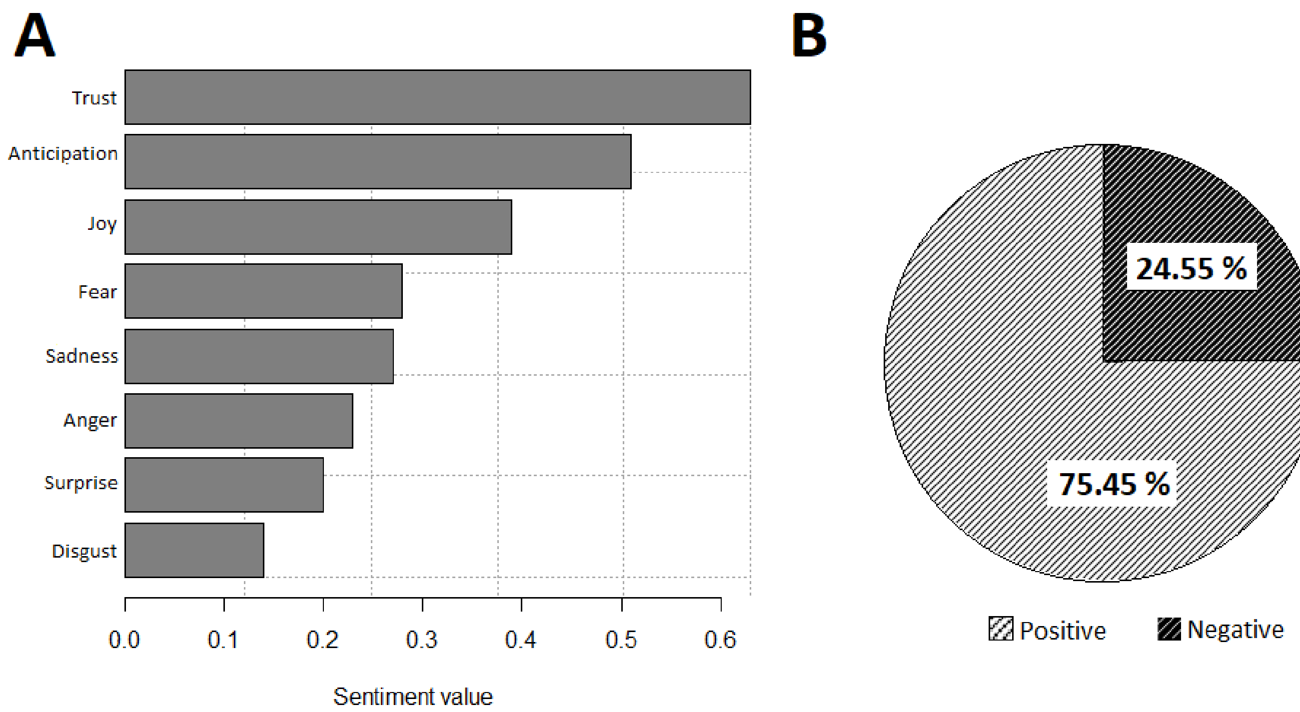

Fig. 3: Sentiment analysis of the whole corpus of tweets.

IV. DISCUSSION AND IMPLICATIONS

This study explored the role of GAI chatbots (exemplified by Bard) in higher education. Therefore, users' discussed topics and their sentiments were extracted from the Twitter messages. The result showed a high level of satisfaction when using Bard for getting assistance in learning. Our topic modelling results showed that Twitter users discussed three main topics ('Applications of Bard to students', 'Bard challenges in higher education', and 'The future of Bard in higher education'). Besides, the dominant emotions associated with users' experience are trust, anticipation, and joy.

Our result revealed that Bard enables students to engage in individual learning by asking questions and receiving quick accurate responses. Such experience helps the students to learn comfortably like learning with a personal tutor. This justifies the existence of high trust and anticipation in our result because the more trusted resources the leaners get help from, the more they rely on to handle their learning-related tasks (Lin, Chang, & Society, 2020; Okonkwo & Ade-Ibijola, 2021). Besides, Bard provides a flexible learning platform to the students since it has the potential to answer question as well as mange their tasks through an organized study schedule. Such a help could explain the existence of a high level of joy in our result because finding approach that helps students to track with their studies promotes a high level of satisfaction (Wu et al., 2020).

In spite the strong support provided by Bard to students in their leaning, our analysis showed that some of the tweets highlighted the importance of considering ethical-related aspects when using GAI chatbots. This is because ethical issues like usability, privacy, and security may influence the users' willingness to use the services (de Barcelos Silva et al., 2020; Okonkwo & Ade-Ibijola, 2021; Shumanov & Johnson, 2021). In fact, personal information is a sensitive matter, and users of Bard must be confident that their information is protected. This includes biographical details, location data, and any other information that users share with Bard. Therefore, Bard's team must guarantee the privacy and security of users' data. This explains the existence of negative sentiments (e.g., fear, sadness, and anger) in our tweets where these emotions are often triggered by ethics and security-related matters (Wall & Buche, 2017).

On the other hand, our LDA result revealed that a significant number of users expressed their enthusiastic for the future application of Bard in the academic field, such as research, knowledge dissemination, and educational interactions. This would be recognized from the positive sentiment in our tweets that is associated with the perceived value and potential benefits that Bard can bring to the academic community in the future. From the previous information, it can be observed that our suggested techniques using LDA and sentiment analysis are very helpful to provide insightful information on assessing the significance of new technologies via people's opinions existing on social medial platforms like Twitter platform. Consequently, this study facilitates the understanding of technology professionals and researchers regarding the role of such technologies in society. It also suggested methods for discovering the existing challenges and provides valuable insights toward future improvements in related domains.

V. LIMITATIONS AND FUTURE WORKS

While the findings of this study provide valuable insights, it is important to address certain limitations that may affect the generalizability and interpretation of the results. Our first limitation is that our corpus was a collection of English-language tweets because this is one of the most popular languages in the world. Second, to extract critical topics from the data, LDA was used as an example unsupervised learning scheme. This method was extensively applied in the literature to accomplish similar knowledge extraction tasks. In the future, scholars could apply other machine learning algorithms to extract hidden themes from their textual data.

Finally, in this study, we obtained different types of emotions (disgust, surprise, sadness, anger, fear, joy, anticipation, and trust) because these are popular types in emotion studies. Future work could consider obtaining additional emotions from users' tweets.

VI. Conclusion

This study demonstrated the potential of using Twitter to understand the role of GAI chatbots in Higher Education. Therefore, several English-language tweets related were collected using the Twitter free streaming Application Programming Interface (API). Hidden topics were extracted from these tweets using the LDA algorithm. Besides, users' sentiments toward the utilized technology were obtained using the lexicon-based method. Despite the limitations, the findings from this study provide valuable insights into the public's perceptions of GAI chatbots in higher education. This helps technology professionals and researchers to understand the role of such technologies in modern learning systems, fix the tackle the existing challenges, and suggest proper directions for improvement.

References

[1] Murugesan, S. and A.K.J.C. Cherukuri, *The Rise of Generative Artificial Intelligence and Its Impact on Education: The Promises and Perils.* 2023. **56**(5): p. 116-121.

[2] Wang, F.-Y., et al., *What does ChatGPT say: The DAO from algorithmic intelligence to linguistic intelligence.* 2023. **10**(3): p. 575-579

[3] Rahaman, M.S., et al., *The AI race is on! Google's Bard and OpenAI's ChatGPT head to head: an opinion article.* 2023

[4] Plevris, V., G. Papazafeiropoulos, and A.J.J.a.p.a. Rios, *Chatbots put to the test in math and logic problems: A preliminary comparison and assessment of ChatGPT-3.5, ChatGPT-4, and Google Bard.* 2023.

[5] Di Minin, E., H. Tenkanen, and T.J.F.i.E.S. Toivonen, Prospects and challenges for social media data in conservation science. 2015. **3**: p. 63.

[6] Giachanou, A. and F.J.A.C.S. Crestani, Like it or not: A survey of twitter sentiment analysis methods. 2016. **49**(2): p. 1-41.

[7] Tang, Y., K.F.J.C. Hew, and education, Using Twitter for education: Beneficial or simply a waste of time? 2017. **106**: p. 97-118.

[8] Ostrowski, D.A. Using latent dirichlet allocation for topic modelling in twitter. in Proceedings of the 2015 IEEE 9th international conference on semantic computing (IEEE ICSC 2015). 2015. IEEE.

[9] Zhou, S., et al., A guided latent Dirichlet allocation approach to investigate real-time latent topics of Twitter data during Hurricane Laura. 2023. **49**(2): p. 465-479.

[10] Jelodar, H., et al., Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. 2019. **78**: p. 15169-15211.

[11] Kim, Y. and K.J.I.S. Shim, TWILITE: A recommendation system for Twitter using a probabilistic model based on latent Dirichlet allocation. 2014. **42**: p. 59-77.

[12] Waheeb, S.A., N.A. Khan, and X.J.E. Shang, Topic Modeling and Sentiment Analysis of Online Education in the COVID-19 Era Using Social Networks Based Datasets. 2022. **11**(5): p. 715.

[13] Paul, D., et al. Compass: Spatio temporal sentiment analysis of US election what twitter says! in Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining. 2017.

[14] Liao, S., et al., CNN for situations understanding based on sentiment analysis of twitter data. 2017. **111**: p. 376-381.

[15] Mohammad, S.M., *Word affect intensities.* arXiv preprint arXiv:1704.08798, 2017.

[16] Thelwall, M.J.C.C.e.i.c., The Heart and soul of the web? Sentiment strength detection in the social web with SentiStrength. 2017: p. 119-134.

[17] Lin, M.P.-C., D.J.J.o.E.T. Chang, and Society, *Enhancing post-secondary writers' writing skills with a chatbot.* 2020. **23**(1): p. 78-92.

[18] Okonkwo, C.W. and A. Ade-Ibijola, *Chatbots applications in education: A systematic review.* Computers and Education: Artificial Intelligence, 2021. **2**: p. 100033.

[19] Wu, E.H.-K., et al., Advantages and constraints of a hybrid model K-12 E-Learning assistant chatbot. IEEE Access, 2020. **8**: p. 77788-77801.