JEE 13 (1) 2024: 51-60

Journal of Economic Education

http://journal.unnes.ac.id/sju/index.php/jeec

Exploring Public Sentiment on ChatGPT - An Analysis of Twitter Data

Nurul Arifah Wahyuni[1]✉ and Dhea Rizky Amelia, I'tishom Al Khoiry

[1)]Technology Management, Economics, and Public Policy, Seoul National University, [2,3)] Faculty of Economics and Business, Universitas PGRI Semarang

Article Info

Article History :
Received March 2024
Accepted Mei 2024
Published June 2024

Keywords:
ChatGPT, Twitter, Topic Modeling, Sentiment Analysis

Abstract

The rapid growth and widespread adoption of ChatGPT, an AI-driven natural language processing (NLP) tool developed by OpenAI, has garnered both positive and negative attention across various sectors. This study explores public opinion on ChatGPT using Twitter data. The methodology employs topic modeling techniques to identify the main topics discussed in relation to ChatGPT on the platform and conducts sentiment analysis to investigate public perceptions. This research is particularly relevant due to the recent release of the paid version of ChatGPT-4 by OpenAI. Understanding public sentiment and the main topics of conversation is essential to address potential risks, identify improvement opportunities, and ensure the ethical and responsible development of AI-driven NLP tools.

✉ Correspondence :
Postgraduate, Universitas Diponegoro
E-mail: dhearizkyamelia@upgris.ac.id

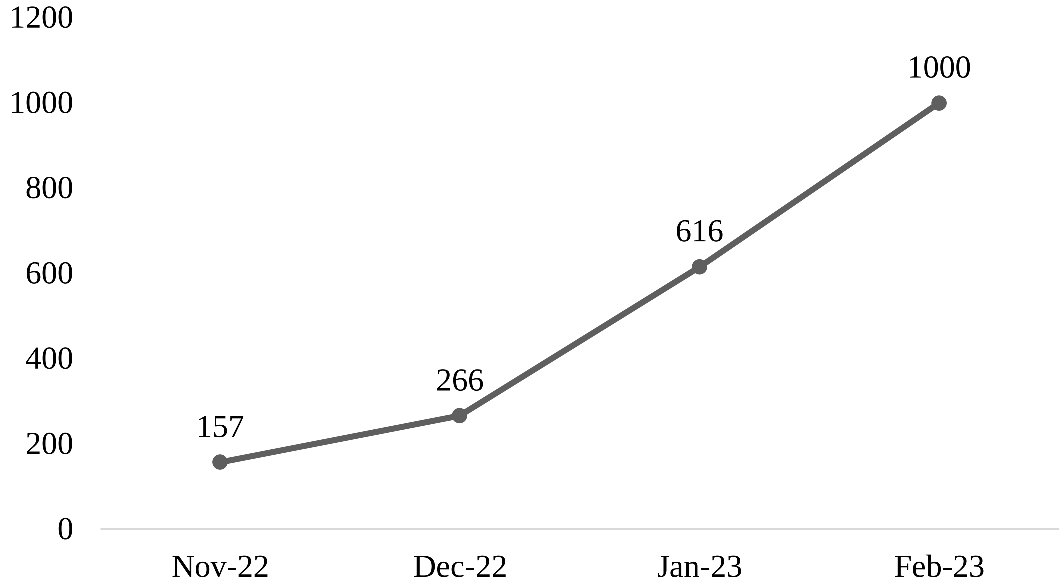
p-ISSN 2301-7341
e-ISSN 2502-4485

INTRODUCTION

The adoption of ChatGPT, developed by OpenAI, represents a significant advancement in AI-driven natural language processing tools (Taecharungroj, V., 2003). ChatGPT has attracted both positive and negative attention from various sectors (Reuters, 2023), becoming one of the fastest-growing user bases for a consumer application. Despite widespread discussion, there is limited research focusing on public sentiment towards ChatGPT (Schöbel, S. et al., 2003; Zierau, N., et al., 2020). This study seeks to explore this sentiment using data from Twitter, a platform known for real-time access to public opinion.

The rapid advancement of artificial intelligence (AI) technologies has introduced powerful tools that are transforming various aspects of human interaction (Haque, 2022). Among these, ChatGPT, an AI language model developed by OpenAI, has captured significant public attention for its ability to generate coherent and contextually relevant text. Since its release, ChatGPT has been employed in a range of applications, from customer service to creative writing, leading to widespread discussion about its capabilities, limitations, and ethical implications.

Monthly Website Visitors of ChatGpt (in millions)



Source: OpenAI (2023), Retrieved https://similarweb.com/website/chat.openai.com/#overview

ChatGPT sets a new record for the fastest-growing customer application user base, as it continues to attract a significant number of users who visit the platform multiple times [2].

Public sentiment towards AI technologies like ChatGPT is crucial, as it influences both the acceptance and the development of such tools. Twitter, with its vast user base and real-time communication dynamics, offers a rich source of data for analyzing public opinion. Previous studies have demonstrated the value of Twitter data in capturing societal trends and sentiments, making it an ideal platform for examining reactions to emerging technologies (Sarlan, 2014)

This paper leverages Twitter data to conduct a comprehensive analysis of public sentiment towards ChatGPT. By collecting and analyzing tweets from diverse user demographics and geographic locations, this study aims to identify prevailing themes, sentiments, and concerns expressed by users. The analysis is grounded in data science methodologies, utilizing sentiment analysis tools and natural language processing (NLP) techniques to categorize and quantify public opinion. Additionally, references to key studies on sentiment analysis and AI adoption provide a contextual framework for interpreting the findings.

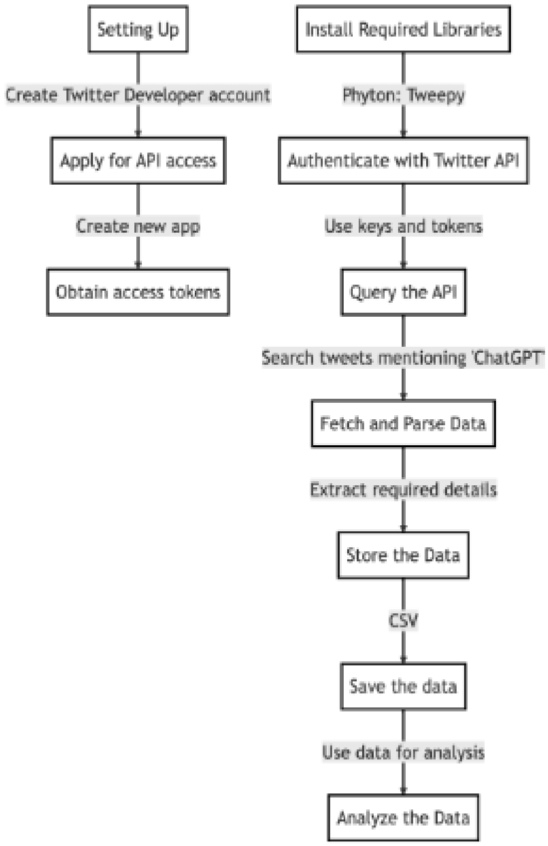
Through this analysis, the paper seeks to provide a nuanced understanding of the public's perception of ChatGPT, highlighting both the enthusiasm and apprehensions surrounding its use. By integrating data-driven insights with existing literature, this study contributes three research objective (1) Utilize topic modeling techniques to identify main topics discussed in relation to ChatGPT on Twitter (2) Conduct sentiment analysis to gauge public perceptions of ChatGPT (3) Understand public sentiment to address potential risks, identify opportunities for improvement, and ensure ethical development of AI-driven NLP tools.

RESEARCH METHODS

The study's methodology involves data collection, data preprocessing, sentiment analysis, and topic modeling using Latent Dirichlet Allocation (LDA). Latent Dirichlet Allocation (LDA) was used to identify latent topics in the dataset. LDA groups tweets into topics based on word co-occurrence and frequency, assigning probabilities to each tweet related to specific topics.

Data was sourced from Twitter, a microblogging platform where users post messages known as tweets. The data spans a three-month period from February 20 to April 19, 2023, and focuses on tweets containing the keyword "ChatGPT," excluding those in languages other than English. Data extraction utilized the Twitter API and Python for data scraping, collecting information such as the date, tweet content, username, likes, and retweets.

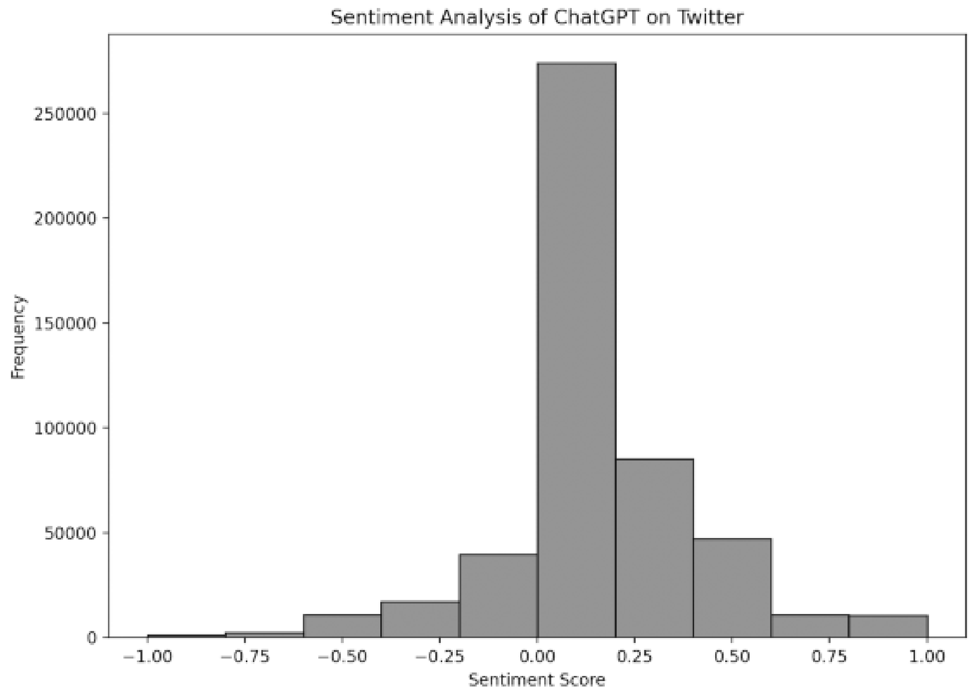
Preprocessing involved several steps: (1) Duplication Removal: Removed retweets and duplicates. (2) Lowercasing: Converted all text to lowercase for standardization, (3) Noise Removal: Removed URLs, emojis, handles, and punctuation, (4) Stop Words Removal: Removed common words using the NLTK English stop-word list, (5) Lemmatization: Transformed words to their root form using WordNet-based lemmatization in NLTK.



Sentiment analysis was conducted using TextBlob, which measures polarity (ranging from -1 for negative to +1 for positive sentiment) and subjectivity (ranging from 0 for objective to 1 for subjective content).

RESULTS AND DISCUSSION

Sentiment Analysis Result



The figure illustrates a histogram that captures the sentiment analysis of tweets about ChatGPT on Twitter. The x-axis represents the sentiment scores, ranging from -1.0 to 1.0. These scores indicate the polarity of the sentiments, with -1.0 representing extremely negative sentiment, 0 representing neutral sentiment, and 1.0 representing extremely positive sentiment. The y-axis represents the frequency, indicating the number of tweets corresponding to each sentiment score range.

At first glance, the histogram reveals a prominent peak at the sentiment score of 0.0. This peak suggests that a substantial number of tweets have a neutral sentiment toward ChatGPT. In fact, this is the most frequent sentiment category, indicating that many users express neutral or balanced opinions about the AI.

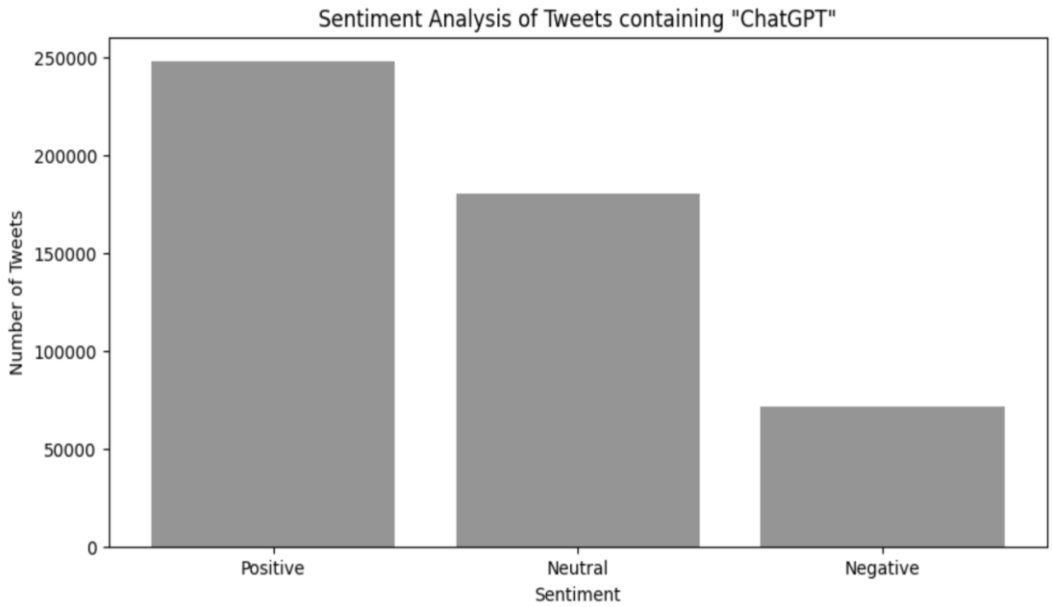
Moving towards the positive side of the sentiment scale, there is a significant number of tweets with sentiment scores ranging from 0.25 to 0.75. This range shows that many tweets express mildly to moderately positive sentiments about ChatGPT. There is a noticeable secondary peak around the 0.5 mark, highlighting a substantial volume of tweets that carry a positive tone.

On the negative side of the sentiment scale, the frequency of tweets decreases as the sentiment scores move towards -1.0. This indicates that fewer tweets express negative sentiments about ChatGPT. The histogram shows that negative sentiments are less prevalent compared to neutral and positive sentiments.

Overall, the histogram suggests that the general sentiment toward ChatGPT on Twitter is predominantly neutral to positive, with significantly fewer negative tweets. This distribution provides insight into how users perceive and discuss ChatGPT on social media, reflecting a generally favorable or unbiased view.

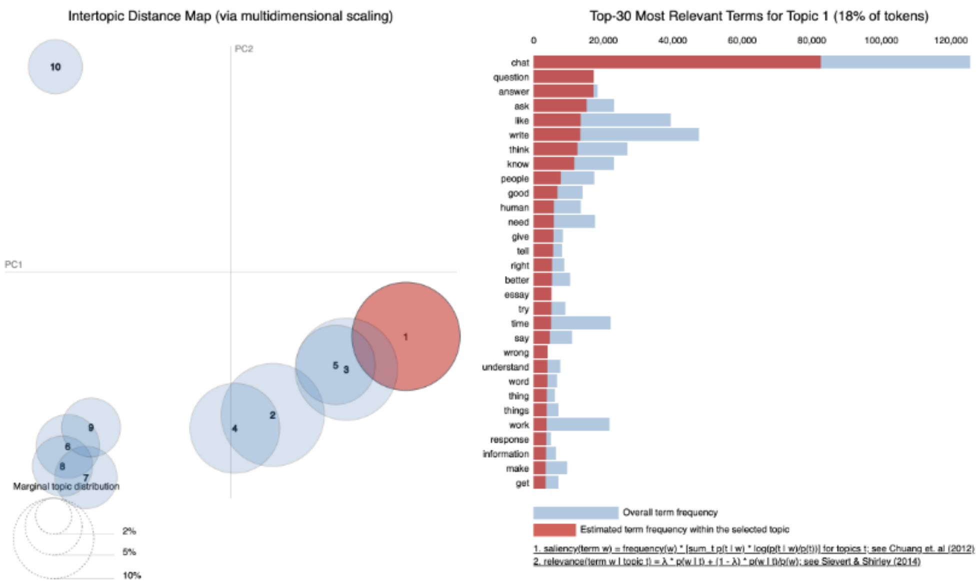
```
==============================================
Analyzing sentiment of tweets...
Analyzing sentiment distribution...
2023-05-08 07:40:17.264 python[59369:743889] +[CATransaction synchronize] called
 within transaction
Computing overall sentiment score...
Overall sentiment score: 0.12
(base) nurularifahwahyuni@Nuruls-MacBook-Air Downloads %
```

The sentiment analysis of Twitter data related to ChatGPT resulted in an overall sentiment score of 0.12.. A score of 0.12 is closer to neutral, suggesting that the general public sentiment on Twitter is neither strongly positive nor negative



The figure presents a sentiment analysis of tweets containing the keyword "ChatGPT." The bar chart shows the distribution of tweets across three sentiment categories: Positive, Neutral, and Negative. The majority of tweets fall under the Positive sentiment category, with approximately 250,000 tweets expressing a positive view towards ChatGPT. The Neutral sentiment category is the second most prevalent, with slightly fewer tweets than the Positive category, around 180,000 tweets. The Negative sentiment category has the least representation, with around 60,000 tweets expressing a negative sentiment. This suggests that public sentiment on Twitter regarding ChatGPT is predominantly positive, with a significant portion of tweets being neutral and a smaller fraction expressing negativity.

Topic Modelling Result 1



The figure consists of two main components: an intertopic distance map (left) and a bar chart of the top 30 most relevant terms for Topic 1 (right), likely generated from a topic modeling analysis such as Latent Dirichlet Allocation (LDA).

The left side of the figure displays an intertopic distance map, visualized through multidimensional scaling (MDS). The map shows the spatial relationships among ten topics identified in the dataset, with each circle representing a different topic. The positioning of the circles indicates the similarity between topics: circles that are closer together represent topics that are more similar, while those that are farther apart are less similar.

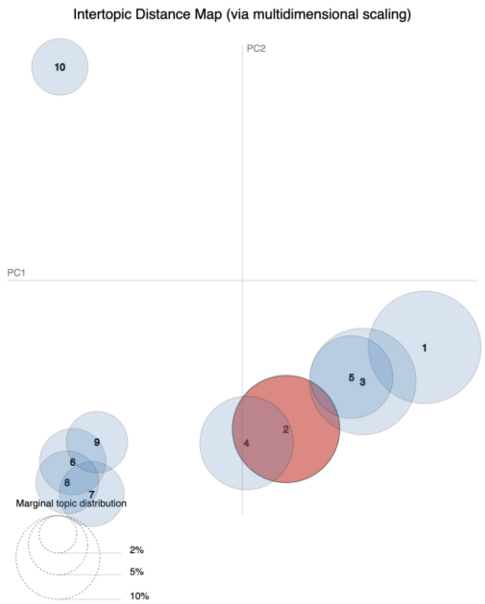
- Topic 1, represented by the largest circle, occupies a central position, suggesting it has a significant presence in the dataset and is related to several other topics.
- Topics 2 through 6 are clustered together at the bottom-left, indicating that they are closely related to each other.
- Topic 10 is distinctly separate from the other topics, implying it is quite different in content.
- The marginal topic distribution is shown as dotted lines in the bottom-left, indicating the proportion of the overall dataset that each topic occupies, with Topic 1 being the most dominant.

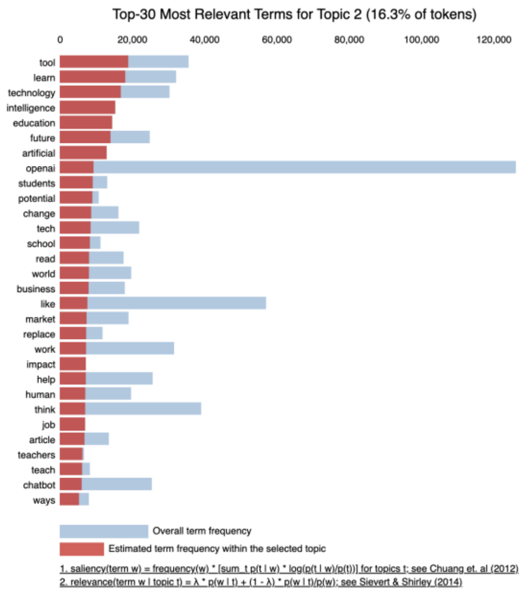
On the right side of the figure, a bar chart lists the top 30 most relevant terms for Topic 1, which accounts for 18% of the total tokens in the dataset. The terms are ordered by relevance to the topic, with the most relevant term at the top.

The term "chat" is the most relevant, with a high estimated frequency within Topic 1, indicating that this topic likely revolves around discussions related to chat or communication. Other top terms include "question," "answer," "ask," "like," "write," and "think," suggesting that Topic 1 may be focused on dialogues or interactions, possibly involving Q&A formats or discussions about writing and thinking processes.

The bar chart uses two colors: red represents the estimated term frequency within the selected topic, while blue represents the overall term frequency within the entire dataset. This differentiation helps in understanding how uniquely relevant these terms are to Topic 1 compared to the overall corpus. Overall, the figure provides insights into the thematic structure of the data, with Topic 1 being a central and dominant theme, characterized by terms related to conversation and communication.

Topic Modelling Result 2





The figure comprises two key visualizations: an intertopic distance map on the left and a bar chart detailing the top 30 most relevant terms for Topic 2 on the right. This analysis likely stems from a topic modeling technique such as Latent Dirichlet Allocation (LDA), which uncovers latent themes within a corpus of text data.

The intertopic distance map visualizes the spatial relationships among ten identified topics using multidimensional scaling (MDS). Each circle represents a distinct topic, with its size corresponding to the prevalence of that topic within the dataset. The position of the circles relative to each other indicates the semantic similarity between topics—closer circles suggest more closely related topics, whereas those positioned further apart imply greater thematic divergence.

a) Topic 2, shown as a prominent and relatively large circle, occupies a central position on the map, suggesting it is a significant and well-integrated theme within the overall dataset.
b) Other topics, notably Topics 1, 3, 4, and 5, are clustered around Topic 2, indicating a strong interconnection between these thematic elements.
c) Topic 10 stands alone, far from the other topics, implying it captures a distinct and less related theme.
d) The marginal topic distribution depicted in the bottom-left corner indicates the proportion of the corpus that each topic occupies, with Topic 2 comprising 16.3% of the tokens, making it one of the dominant themes.

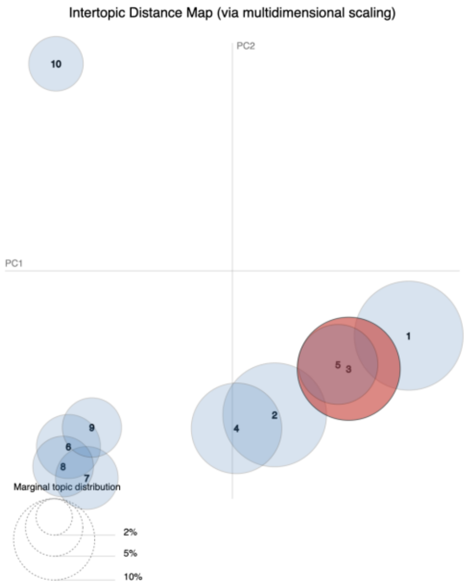
The bar chart on the right lists the top 30 terms most relevant to Topic 2, which accounts for 16.3% of the tokens in the dataset. The terms are ordered by their relevance, a measure combining their frequency within the topic and their uniqueness compared to other topics.

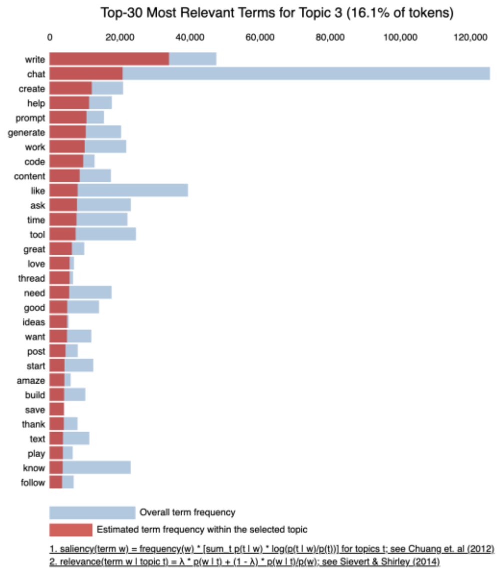
a) The term "tool" emerges as the most relevant, reflecting the centrality of discussions about tools or instruments within this topic. This suggests that Topic 2 may be focused on the practical applications or utility of technology, particularly in the context of learning and education, as evidenced by the presence of other high-ranking terms such as "learn," "technology," "intelligence," and "education."
b) The prominence of terms like "artificial," "openai," "students," and "future" further underscores the topic's thematic emphasis on the intersection of artificial intelligence with education and its future implications.
c) Terms related to business, market, and impact—such as "business," "market," "replace," and "impact"—indicate that Topic 2 also touches upon the broader economic and societal ramifications of these technologies.
d) The chart visually distinguishes between the overall term frequency in the dataset (represented in blue) and the estimated term frequency within Topic 2 (represented in red), highlighting the unique significance of these

terms within the topic compared to the corpus as a whole.

In summary, the figure offers a comprehensive view of the thematic landscape within the analyzed dataset, with Topic 2 emerging as a central theme, heavily focused on discussions surrounding technological tools, artificial intelligence, and their implications for education and society. This insight into the structure and content of Topic 2 can inform further exploration of how these discussions are framed and understood within the broader discourse.

Topic Modelling Result 3





The provided figure consists of two primary components: an Intertopic Distance Map on the left and a bar chart detailing the Top-30 Most Relevant Terms for Topic 3 on the right.

This plot visualizes the relationships and distances between different topics generated from a corpus, using multidimensional scaling (MDS). Each circle represents a distinct topic, with its size corresponding to the marginal distribution of that topic in the corpus (i.e., how prevalent the topic is). The axes, PC1 and PC2, denote the principal components derived from the MDS, facilitating the visualization of inter-topic distances in a two-dimensional space. Notably, Topic 10 is significantly isolated, suggesting it has distinct terms that do not overlap much with other topics. In contrast, Topics 1, 3, and 5 show considerable overlap, indicating they share many common terms or thematic elements.

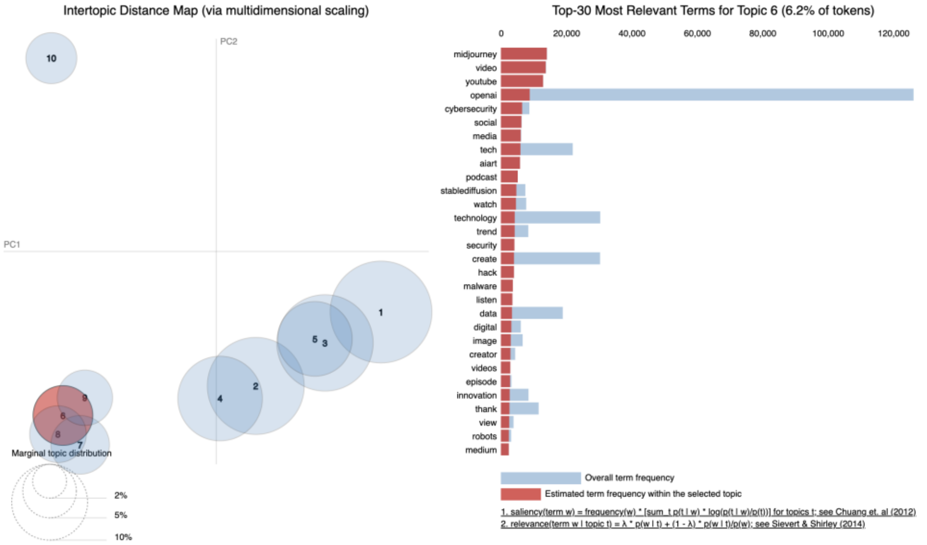
The bar chart on the right side highlights the top 30 terms most relevant to Topic 3, accounting for 16.1% of the tokens in the corpus. Each term is displayed with two bars: the blue bar represents the overall frequency of the term in the entire corpus, while the red bar indicates the estimated term frequency within the selected topic (Topic 3). For example, the term "write" shows a high overall frequency but is particularly salient in Topic 3. The chart provides a clear view of the terms that characterize Topic 3, with words such as "write," "chat," "create," and "help" being among the most significant. This suggests that Topic 3 is likely centered around activities related to content creation and interaction, such as writing and chatting.

The Intertopic Distance Map provides a visual summary of topic coherence and separation. Topics that are closer together share more terms and thematic content, while those that are farther apart are more distinct. The distribution and size of the circles help in understanding the prominence and spread of each topic across the corpus.

The bar chart complements this by offering a detailed look at the specific terms that define one of the topics. The dual bars for each term enable a comparative analysis of its general prevalence and its significance within Topic 3, thereby helping in pinpointing the thematic core of the topic.

Topic Modelling Result 4

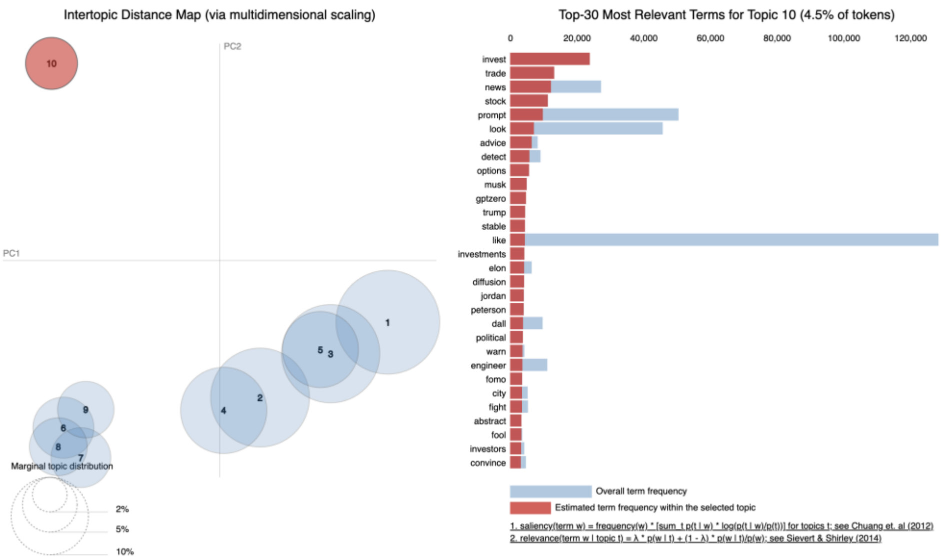
The provided figure contains two main visualizations: an Intertopic Distance Map on the left and a bar chart showcasing the Top-30 Most Relevant Terms for Topic 6 on the right.



This plot illustrates the spatial relationships among various topics derived from a corpus, employing multidimensional scaling (MDS) to represent these distances. Each circle signifies a distinct topic, with its size corresponding to the marginal distribution of that topic within the corpus, indicating its relative prevalence. The plot is anchored by two principal components, PC1 and PC2, derived from the MDS, which help in visualizing the topics in a two-dimensional space. Topic 10 appears notably isolated from the rest, suggesting a unique set of terms distinct from other topics. In contrast, Topics 6, 8, and 9 display significant overlap, indicating a considerable shared vocabulary or thematic content among them.

The bar chart on the right side delineates the top 30 terms most relevant to Topic 6, which constitutes 6.2% of the tokens in the corpus. For each term, there are two bars: the blue bar indicates the overall frequency of the term across the entire corpus, while the red bar shows the estimated term frequency within Topic 6. This dual representation allows for a nuanced comparison of term usage. Terms like "midjourney," "video," "youtube," and "openai" are highly relevant to Topic 6, suggesting that this topic is centered around themes related to digital media, technology platforms, and cybersecurity. The presence of terms such as "cybersecurity," "social," "media," and "tech" further emphasizes the technological and digital focus of Topic 6.

Topic Modelling Result 5



The figure presented comprises two integral visual elements: the Intertopic Distance Map on the left and a bar chart displaying the Top-30 Most Relevant Terms for Topic 10 on the right.

This map offers a spatial representation of the relationships among various topics extracted from a textual corpus using multidimensional scaling (MDS). Each circle on the map denotes a unique topic, with the circle's size reflecting the marginal distribution of that topic within the corpus, thereby indicating its relative prevalence. The map is structured along two principal components, PC1 and PC2, which facilitate the visualization of topics in a two-dimensional space. Notably, Topic 10 is positioned distinctly apart from the cluster of other topics, indicating a significant thematic divergence. The other topics (1 through 9) are more closely grouped, suggesting overlapping vocabularies and thematic content.

The bar chart to the right lists the top 30 terms that are most relevant to Topic 10, which encompasses 6.9% of the tokens in the corpus. Each term is represented by two bars: the blue bar indicates the overall frequency of the term in the entire corpus, whereas the red bar shows the estimated frequency of the term within Topic 10. This comparison highlights terms that are particularly salient to Topic 10. For instance, terms like "crypto," "invest," "blockchain," and "trading" have high relevance, suggesting that Topic 10 predominantly revolves around financial and cryptocurrency themes. The inclusion of terms such as "stock," "market," "price," and "trade" further corroborates this focus on financial trading and investment.

CONCLUSION

The slight positivity in sentiment towards ChatGPT on Twitter suggests that users find some value in the tool, though the score's proximity to zero indicates the presence of neutral viewpoints or concerns. This moderate sentiment reflects a need for improvement in user perception, presenting opportunities to address these concerns and enhance positive sentiment. Current sentiment analysis tools, however, face challenges such as handling ambiguity, understanding context, and detecting sarcasm or irony areas where future research could focus on developing models that better account for the larger context of words and their multiple meanings. Additionally, while much progress has been made in English sentiment analysis, there is considerable scope for improving tools for other languages. Research into real-time sentiment analysis could prove invaluable for live events, offering timely insights for social media platforms and news agencies. Furthermore, longitudinal sentiment analysis, tracking changes over time, could yield valuable insights into shifting public opinions. Lastly, improving the Latent Dirichlet Allocation (LDA) model by fine-tuning its hyperparameters could enhance its effectiveness in identifying topics within sentiment analysis, further refining the accuracy and relevance of the results.

REFERENCES

ChatGPT Statistics 2023. Retrieved from (https://www.similarweb.com/website/chat.openai.com/overview).

Haque, M. U., et al. (2022). "I think this is the most disruptive technology": Exploring Sentiments of ChatGPT Early Adopters using Twitter Data. arXiv preprint arXiv:2212.05856.

Haque, M. U., Iwaya, L. H., & Babar, M. A. (2020). Challenges in Docker development: A large-scale study using Stack Overflow. In Proceedings of the 14th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM) (pp. 1–11).

Reuters. (2023). ChatGPT sets record for fastest-growing user base. Retrieved from (https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/).

Sarlan, A., Nadam, C., & Basri, S. (2014). Twitter sentiment analysis. In Proceedings of the 6th International Conference on Information Technology and Multimedia (pp. 212–216). IEEE. doi:10.1109/ICIMU.2014.7066632

Schöbel, S. et al. (2023). Charting the Evolution and Future of Conversational Agents: A Research Agenda Along Five Waves and New Frontiers. Information Systems Frontiers. doi:10.1007/s10796-023-10375-9.

Taecharungroj, V. (2023). “What Can ChatGPT Do?” Analyzing Early Reactions to the Innovative AI

Chatbot on Twitter. Big Data and Cognitive Computing, 7(35).
Zierau, N., et al. (2020). A review of the empirical literature on conversational agents and future research directions. International Conference on Information Systems (ICIS).

.