Received: 29 August 2025 | Revised: 26 December 2025 | Accepted: 11 March 2026 | Published online: 25 March 2026

RESEARCH ARTICLE

Artificial Intelligence and Applications
2026, Vol. 00(00) 1–16
DOI: 10.47852/bonviewAIA62027447

BON VIEW PUBLISHING

Public Discussion of DeepSeek Large Language Model on Twitter: A Mixed-Methods Sentiment and Topic Modeling

Wei Chien Ng[1,3,*], Shiwen Chen[1], Quan Kai Ang[1] and Yu Qing Soong[2]

[1] *School of Management, Universiti Sains Malaysia, Malaysia*

[2] *Department of Accountancy and Business, Tunku Abdul Rahman University of Management and Technology, Malaysia*

[3] *College of Business Administration, International University of Business Agriculture and Technology, Bangladesh*

Abstract: Artificial intelligence (AI) has moved from research laboratories into everyday tools used by millions worldwide. In recent years, advances in natural language AI systems have sparked extensive public exploration and discussion. This study investigates overall public sentiment and key discussion topics related to the DeepSeek large language model (LLM) on Twitter (now rebranded as X) and examines sentiment differences across discussion topics during various DeepSeek-related events. After data collection, Python was used to perform preliminary cleaning and screening of English-language tweets. The Valence Aware Dictionary and Sentiment Reasoner (VADER) sentiment analysis tool was applied to classify tweet sentiment. Based on the VADER labels, the dataset was stratified to obtain a high-quality sample of 5000 tweets while preserving the original sentiment distribution. To further explore discussion themes, latent Dirichlet allocation combined with coherence score evaluation was employed for topic modeling. Topic-level sentiment analysis was then conducted across different DeepSeek-related events to assess public attitudes toward each discussion topic. Results indicate that overall public sentiment toward DeepSeek LLM is predominantly positive. Topic modeling identified 10 optimal discussion topics, covering areas such as technical performance, economic impact, political and cultural context, and international competition. The findings also reveal significant differences in sentiment distribution across topics, demonstrating the practical value of combining sentiment analysis and topic modeling for business intelligence and AI product optimization.

Keywords: DeepSeek, Twitter, large language model, natural language processing, sentiment analysis

1. Introduction

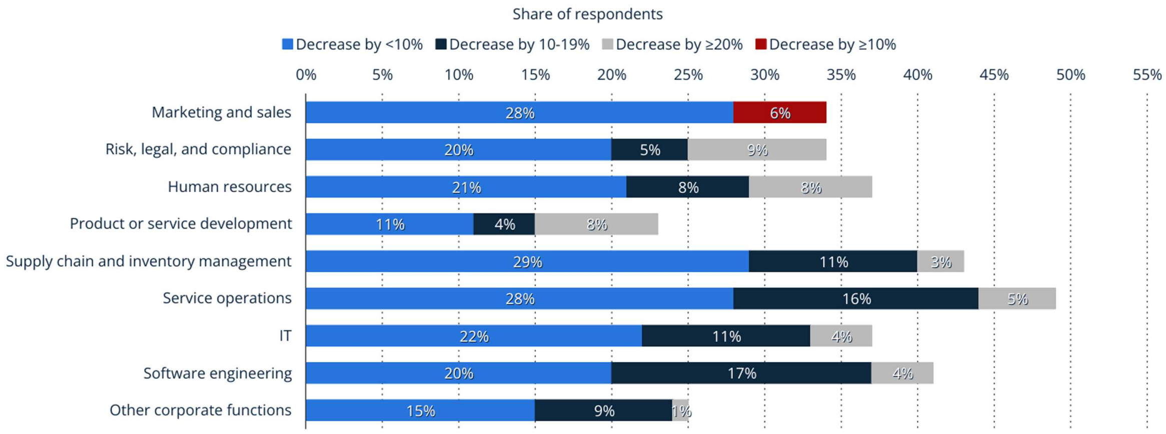
Artificial intelligence (AI) refers to computer systems and algorithms that perform tasks typically requiring human intelligence [1]. Hence, AI can learn from data and recognize patterns to make predictions, enabling businesses to simulate human decision-making and unlock actionable insights from large datasets. By applying AI to operations, businesses increase efficiency through process automation, which reduces errors and lowers costs while reallocating human talent to higher-value work. As shown in Figure 1 [2], the use of analytical AI in various fields has reduced the costs incurred in respective operations, particularly by more than 10% in marketing and sales. In research and development of products, generative models accelerate prototyping and innovation [3]. When responsibly governed, AI becomes a force multiplier that enhances productivity, decision quality, and sustainable growth potential.

Among the AI models, large language models (LLMs) are widely adopted by businesses, bringing significant value to their business performance [4]. LLMs are advanced machine learning systems trained on massive text bodies to predict and generate human-like language. For businesses, LLMs deliver tangible benefits by automating customer support and routine communications at scale, which saves labor power and time consumed in operations [5]. By extracting insights from documents and conversations, LLMs can produce draft reports through speed knowledge retrieval. Hence, LLM-driven tools enhance decision-making by rapidly integrating market intelligence and scenario generation, thereby improving personalization in sales and product recommendations.

Building on LLM capabilities, DeepSeek is a specialized language model optimized for enterprise search and context-aware information retrieval. One of the significant advantages of DeepSeek is its open-source nature, which allows for customization and transparency [6]. This nature makes it more accessible for integration into various systems. For the businesses, DeepSeek reduces time spent locating relevant knowledge by

*Corresponding author: Wei Chien Ng, School of Management, Universiti Sains Malaysia, Malaysia. Email: ngweichien@usm.my

Figure 1
Reduced costs in respective functions from the use of AI in enterprises worldwide in 2024



providing concise summaries and direct references [7]. It enhances the onboarding and customer support by making institutional knowledge accessible to employees and agents. By training on the DeepSeek LLM for months, these efficiency gains result in clear cost savings and a measurable competitive edge, as well as sustained strategic value [8].

To track the trend of AI, discussion on social media is crucial, as it has become a platform for users to share the performance of AI nowadays. Social media provides a public platform where diverse users share their opinions and information about trending topics. According to a survey conducted by Dixon [9], the number of social media users worldwide increased by approximately 64% from 2019 to 2025, as shown in Figure 2. These platforms enable the rapid dissemination of news and ideas, allowing conversations to reach broad audiences quickly.

As such, businesses may keep track of social media platforms, such as Twitter (currently rebranding as "X"), to explore the opportunities that can be leveraged in terms of AI integration in business. A study conducted by Kada et al. [10] demonstrated that Twitter can significantly influence public discussion and engagement. As of 2023, the number of Twitter users was 415.3 million, and it is expected to grow to 2028, as forecasted by Statista [11], as shown in Figure 3. Hence, businesses that adopt divergent content strategies on Twitter, compared to their traditional rivals, may benefit from higher engagement and faster follower acquisition.

The AI features that improve user engagement and suitability for tasks are crucial for enhancing the user experience in AI systems [12]. Users of various AI products demonstrate a strong interest in explainability features that help them understand AI decisions and actions, thereby bridging the gap between algorithmic work and user needs. However, AI systems may face ethical dilemmas such as privacy violations, bias, and discrimination, which can significantly reduce public trust [13]. The inability to explain AI decisions and the presence of algorithmic bias can lead to malfunctions and injustices.

Understanding the evolution of AI topics can provide deep insights into scientific innovations and help predict emerging fields, as a business must stay competitive and adapt to new developments. For example, AI applications in supply chain management are analyzed to identify key themes and trends, which can help the business enhance efficiency and manage risks [14]. Unfortunately, the rapid changes in AI technology and public conversation can lead to overlooked areas and unintended consequences [15]. Businesses may lack a strategic framework to align AI initiatives with business goals and public expectations, resulting in missed opportunities and increased public attention [16].

By gaining insights into public discussions on AI-related topics, businesses can align their products and roadmaps based on real comments and suggestions collected [17]. The participation

Figure 2
Number of social media users worldwide from 2019 to 2025

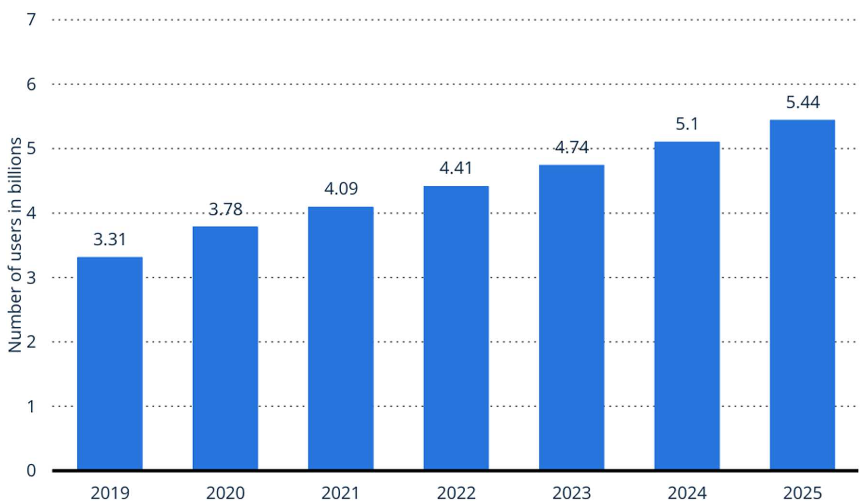
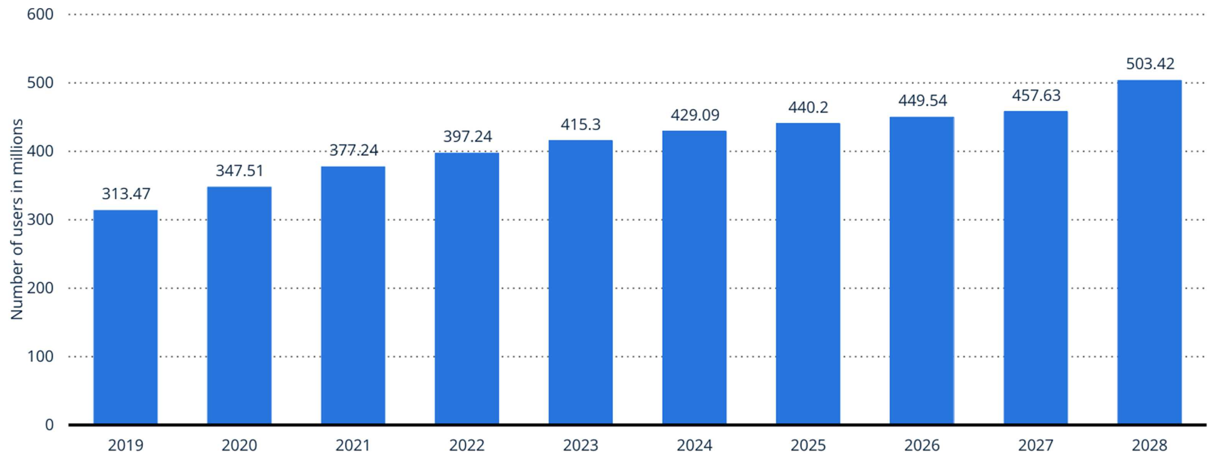


Figure 3
Number of Twitter users worldwide from 2019 to 2028



of potential customers early in the product development process can lead to more customized and relevant products. The different topics surrounding the same event may elicit opposite public reactions, which are often treated as uniform, misguided decisions. As shown in Choi's study [18], both near-future and distant-future contexts in the framing of AI can lead to different public perceptions of AI risks and benefits.

Hence, the research objectives in this study are to (i) determine the overall public sentiment toward DeepSeek LLM on Twitter (RO1), (ii) determine the key discussion topics around DeepSeek LLM on Twitter (RO2), and (iii) determine the public sentiment on respective discussion topics around DeepSeek LLM on Twitter (RO3).

2. Literature Review

The LLM of DeepSeek outperforms other LLMs. Hence, the public discussion of DeepSeek LLM on Twitter may help businesses to gain insights for improving product and strategy development. As a result, sentiment analysis (SA) and topic modeling are common analyses carried out to examine public discussions on social media.

2.1. DeepSeek LLM

DeepSeek LLM exhibits distinct advantages over many other LLMs due to its low-cost training, which substantially lowers the financial barriers to developing and operating advanced AI [18]. Its open-source nature provides transparency and extensive customization opportunities. This nature enables integrators to modify and adapt the model for specific field requirements, as it can assist in validating and ensuring compliance with new products. The current systems may limit flexibility and increase dependence on vendor support for updates and licensing if not assisted by DeepSeek LLM [19]. Moreover, the energy industry benefits from DeepSeek LLM's ability to handle complex queries and integrate vast amounts of data, which is essential for optimizing operations and enhancing research in fields such as petroleum engineering [20].

Additionally, DeepSeek LLM exhibits superior reasoning capabilities for complex tasks. DeepSeek-R1 exhibits advanced reasoning capabilities, achieving reasoning accuracy comparable to GPT-4 while producing more diverse judgments [21]. Its algorithms, such as Multi-head Latent Attention, Mixture-of-Experts, and Multi-Token Prediction, significantly improve memory efficiency and computational performance [8]. Supporting hardware-aware optimizations enables cost-efficient training at scale. It includes FP8 mixed-precision training and a Multi-Plane Network Topology designed to minimize the network overhead. These algorithmic and systems advancements allow DeepSeek to utilize hardware more effectively [8]. Businesses can reduce operational resource demands and support large-scale developments with lower infrastructure costs while maintaining comprehensive capacity.

Despite these noted strengths, the literature exhibits a pronounced optimism that warrants critical scrutiny. The discourse can benefit from engaging with persistent challenges, such as the need for independence, empirical validation of the claimed training efficiencies, and reasoning benchmarks in diverse, real-world industrial settings. Furthermore, the open-source model raises significant questions regarding long-term maintenance, model governance, and the potential hidden costs of customization. More balanced research must address this applicability gap and the evolving ethical and operational risks associated with deploying such systems in high-stakes domains, such as the energy sector.

2.2. Twitter data in business intelligence

Twitter data has become a crucial source of modern business intelligence, providing real-time insights into public discussion sentiment and emerging trends. The study conducted by Aramburu et al. [22] focused on the processing of Twitter data for Social Business Intelligence projects, emphasizing the importance of data exploration and profiling in building a high-quality data collection that aligns with the project's analysis objectives. By applying descriptive and content analysis techniques, businesses can derive insights into customer interactions and refine their engagement strategies. Another study by Al-Otaibi et al. [23] employed SA on Twitter data to categorize tweets as positive, negative, or neutral, providing insights into public discussions and supporting business decisions. By analyzing these Twitter-derived signals, businesses can help forecast demand and prioritize feature requests from their customers to shape systematic preparation for business opportunities.

As such, the reliance on public application programming interface raises critical questions about completeness and representativeness, which can potentially skew analysis. Furthermore, sentiment classification often overlooks linguistic nuance, such as sarcasm and context, which can compromise the validity of business insights derived. Confront the ethical dimensions of data extraction and the commercial use of public sentiment, issues frequently marginalized in techno-optimistic accounts of business intelligence.

2.3. Sentiment analysis

SA, also known as opinion mining, is a technique used to identify and categorize the opinions expressed in textual data [24]. It aims to determine the sentiment polarity of a given piece of text, which classifies it as positive, negative, or neutral. This process involves analyzing subjective information from various sources such as social media posts, product reviews, and customer feedback. Furthermore, SA research has gradually broken through the traditional polarity label, combining it with the dimension of emotional intensity [25]. This trend significantly enhances the expressive ability and practical value of the model, as SA is widely utilized in text analysis research on social media.

In fact, SA is crucial for businesses to understand consumer feedback and develop strategies accordingly. In the study by Sutedja and Hendry [26], this method enables businesses to identify customer needs and concerns, along with potential measures that can be taken to sustain and improve their performance. According to Murthy et al.'s research [27], this technique is used to support emotional marketing campaigns, enabling businesses to create more targeted and personalized messaging that resonates with their customers. In highly competitive industries, such as the automotive sector, businesses like Honda, Toyota, BMW, Audi, and Mercedes utilize SA to evaluate the customer sentiments expressed in social media posts, which helps them improve their marketing strategies and product offerings [28].

Therefore, businesses can utilize SA to monitor social media platforms, which have become essential tools for understanding customer emotions, preferences, and opinions. According to Singh et al.'s study [29], this analysis enables businesses to optimize their product offerings and improve customer satisfaction by analyzing user-generated content on platforms such as Facebook and Twitter. By analyzing Twitter data, Rao et al. [30] demonstrated that businesses can redesign their products based on identified areas for improvement to better meet customer needs and preferences. Moreover, Samuel and Krishna [31] demonstrated that Twitter SA provides real-time insights into public opinion, enabling businesses to respond quickly to customer feedback and complaints.

To carry out SA on Twitter Data, Valence Aware Dictionary and Sentiment Reasoner (VADER), which is a SA algorithm that leverages a lexicon and predefined rules, is particularly effective for short social media posts like tweets [32]. Furthermore, Mountstephens and Quen [33] demonstrated the multilingual capabilities of VADER by translating and adapting its sentiment lexicon while maintaining its rule-based structure. Moreover, Dhanalakshmi et al. [32] conducted SA on airline-related tweets. They found that the method is fast and the results are easy to interpret, making it suitable for sentiment classification of social media texts. According to a study by Abubakir et al. [34], VADER is specifically designed for analyzing sentiments expressed on social media platforms, such as Twitter, and is highly effective in analyzing the concise and informal language used in tweets.

While highlighting practical applications, it exhibits a pronounced techno-optimism that requires a critical counterpoint. It often overlooks fundamental methodological critiques, including the inherent subjectivity of sentiment categorization and the limitations of lexicon-based tools like VADER in capturing nuanced linguistic phenomena, such as sarcasm, context, and cultural nuances. The ethical dimensions of mass public opinion mining and the potential for algorithmic bias in shaping business strategies remain underexplored. These validity and ethical challenges are crucial to moving beyond instrumental descriptions of capability.

2.4. Topic modeling

Topic modeling is an unsupervised machine learning technique used to identify hidden topics within an extensive collection of documents [35]. It is beneficial for organizing large volumes of unstructured text data, which are viewed as mixtures of topics, where each topic is represented as a distribution over words. Hence, topic modeling utilizes hidden random variables to represent the latent topics in text and uncover these structures through subsequent analysis.

From past studies by Strydom et al. [36] as well as Yazıcı and Ozansoy Çadırcı [37], businesses utilize topic modeling to analyze social media data and customer reviews, gaining insights into customer opinions and experiences. For instance, Strydom et al. [36] applied various topic modeling algorithms to Reddit data to categorize the customer discussions into different topics related to product features and software application developments. Another study by Yazıcı and Ozansoy Çadırcı [37] employed topic modeling to analyze customer reviews from the Best Buy platform, comparing different algorithms to understand their semantic potential in marketing research. In terms of unstructured data sources, such as customer feedback, social media posts, and news articles, Nasereddin [38] demonstrated that topic modeling enables businesses to understand their customers, competitors, and market trends, facilitating informed, reasonable strategic decision-making. In the study by Vukanti and Jose [39], topic modeling was used to analyze business digital economy datasets, extracting emerging trends and generating word clouds.

As such, topic modeling has been applied to Twitter data across various domains. For example, Sharaff et al. [40] analyzed tweets related to the Indian farmers' protest using multiple topic modeling techniques to compare their coherence scores. According to the study by Cripps et al. [41], businesses can exchange high-quality information on Twitter to build customer relationships and promote innovation within their industries. Some research [42, 43] has compared different topic modeling methods, including latent Dirichlet allocation (LDA), Non-negative Matrix Factorization, and Correlation Explanation, to evaluate their effectiveness in extracting topics from Twitter data. As a result, LDA is widely recognized as the most common and popular technique in topic modeling.

LDA is a widely used probabilistic topic modeling technique in machine learning and natural language processing (NLP), which is designed to identify latent topics within a collection of documents by modeling each document as a mixture of topics and each topic as a distribution over words [44]. By adopting Twitter and utilizing LDA for analyzing firm-generated content, businesses that actively use Twitter to share product-related information can see an increase in their market value. For instance, a study by Azizi et al. [45] on COVID-19 tweets used LDA to identify topics and then performed SA to categorize these topics into positive and negative groups. Similarly, another study by Tan and Chia [46] analyzed that combining LDA with SA tools, such as VADER, allows businesses to evaluate the sentiment of topics discussed on Twitter. Yang et al. [47] examined media framing and public discourse on major global issues, including the representation of China in COVID-19 reporting by Western media. Hence, the combination of VADER and LDA can improve the efficiency and accuracy of text analysis [48].

Literature presents a largely uncritical adoption of LDA as a methodological standard, overlooking significant epistemological and practical constraints. Its application to sparse, short-form texts, such as tweets, is problematic, as the bag-of-words assumption often fails to capture nuanced semantic meaning. The heavy reliance on algorithmic coherence scores for validation is an internal metric that does not guarantee interpretability or thematic relevance to human analysts. This gap between statistical output and actionable human insight is frequently under-examined, revealing a need for more interdisciplinary validation frameworks.

3. Research Methodology

The analysis process flow is demonstrated in Figure 4. At the first step, it involves collecting Twitter data. It proceeds with the data preprocessing of collected data, where SA, topic modeling, and topic SA are carried out in the next step to determine the overall public sentiment, key discussion topics, and public sentiment on respective discussion topics related to DeepSeek LLM on Twitter.

3.1. Data collection

Kaggle is recognized as the largest platform for learning and competition in data analysis and modeling [49]. The platform is known for its well-documented notebooks, shared by data scientists, which serve as exemplary of best practices in data science documentation. A dataset, "Tweets and Reactions on DeepSeek," is obtained from Kaggle, which contains Twitter data investigated in this study. The dataset comprises 364,295 tweets collected from January 20, 2025, to January 29, 2025. This dataset contains tweets and reactions about DeepSeek and its release model, as well as text data on other keywords closely related to DeepSeek, such as OpenAI, Meta, and Llama.

3.2. Data preprocessing

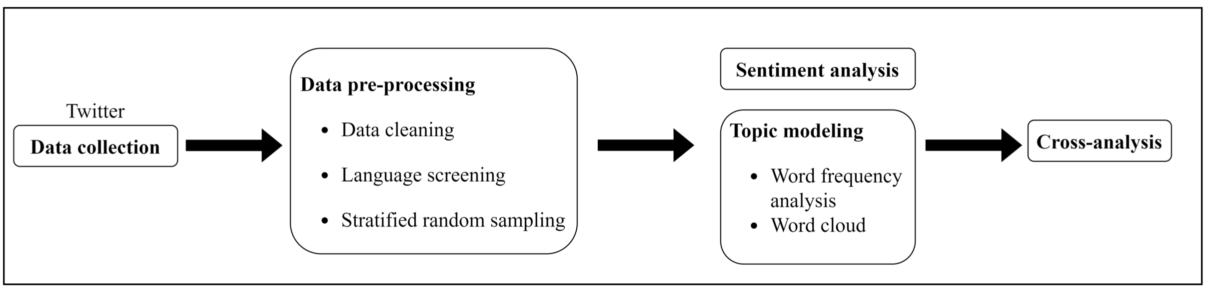
The raw data may contain errors, inconsistencies, and noise that can negatively impact the analysis [50]. Data preprocessing helps in cleaning the data by correcting noisy data [51]. It is vital for handling large-scale datasets efficiently, as data preprocessing can involve data reduction techniques that minimize the dataset's size without losing significant information. As such, the dataset from Kaggle is undergoing several stages of data preprocessing to generate clean and reliable data.

As mentioned by Timbers et al. [52], Python is widely used for data cleaning, as it offers a range of libraries specifically designed for this purpose, such as Pandas. Hence, this study uses Python to perform preliminary data cleaning. Given the requirements of the two analysis tools for text language, VADER is explicitly mentioned as an SA tool for English texts, particularly optimized for social media [53]. The performance of VADER in English is superior compared to its adaptations in other languages, as shown by lower results compared to VADER. On the other hand, LDA can be significantly disrupted by multilingual mixing as the model must account for multiple languages simultaneously, which can lead to challenges in accurately determining the topics. Therefore, this study retains only English tweets for language filtering, ensuring the language consistency required for sentiment classification and topic determination.

Text normalization is a fundamental preprocessing step in handling social media data, where it involves transforming noisy and unstructured text into a standard form to improve the performance of NLP applications, such as SA [54]. In the study by Ali et al. [55], duplicate records can significantly reduce data quality and lead to incorrect conclusions. Similarly, Agustini et al. [56] demonstrated that missing values can degrade the statistical power of analyses and lead to biased or incorrect models. To ensure accuracy, empty or duplicate values in text fields of datasets are removed to maintain data integrity. Social media texts are often noisy due to informal language and nonstandard grammatical structures [57]. Irrelevant content, such as hyperlinks, user mentions, hashtags, and punctuation, is removed from the dataset using regular expressions. Meanwhile, all text is converted to lowercase, as this helps reduce the variance caused by different capitalizations of the same word, ensuring that words like "Happy" and "happy" are treated as the same token [58]. Moreover, the leading and trailing spaces are removed from text data to eliminate unnecessary whitespace that can affect the analysis and lead to inconsistencies in the dataset [59].

After data preprocessing of the English tweets, this study used stratified random sampling based on the sentiment labels. The purpose of stratified sampling is to maintain the original proportion of each sentiment category [60] and reduce the bias in analysis results caused by the sample imbalance [61] after data cleaning. In the context of business, Pradhani et al. [62] demonstrated that stratified random sampling is used to evaluate the influence of social media on brand loyalty and brand equity, illustrating its application in marketing research. As such, it can improve the representativeness of sentiment visualization and topic modeling. As a result, the stratified random sampling produces a before and after sentiment distribution that stratifies three types of sentiment (positive, negative, and neutral) according to their proportion in the entire dataset and then uses the stratify parameter to extract a total of 5000 tweets in equal proportion from each layer to form the

Figure 4
Analysis process flow



analysis sample set, as shown in Figure 5. The sentiment proportion structure is retained after data cleaning, which effectively ensures the representativeness of the sample.

3.3. Sentiment analysis

In SA, VADER is used as a lexicon and rule-based tool to score text, which is particularly effective in capturing sentiments expressed on social media. The structure relies on a system that maps the lexical features to emotion intensities, such as the sentiment score [63]. The results of the VADER algorithm categorize sentiment polarity into three categories: positive, negative, and neutral. The sentiment scores are normalized between −1 (extremely negative) and +1 (extremely positive) to represent the overall sentiment [64]. The sentiment score, which is above zero, indicates a positive sentiment; the sentiment score, which is below zero, indicates a negative sentiment; and the sentiment score of zero indicates a neutral sentiment.

As a result, the sentiment extraction is done using RapidMiner. It supports various SA techniques, which include lexicon-based approaches and machine learning algorithms [65]. Therefore, the SA process flow is performed. First, the cleaned dataset is loaded using the Retrieve operator. Then, the target text column is extracted with the help of the Select Attribute operator to avoid interference from other fields. Next, the nominal attributes are converted to text with the Nominal-to-Text operator to enhance the readability of the data. After that, the Extract Sentiment operator is applied for SA. Finally, the Generate Attributes operator is used to create new fields that enrich the dataset by providing additional contextual information and facilitating a more comprehensive SA. To determine the confidence intervals of the result, bootstrap estimation is used as it simplifies the often-complicated calculations required by traditional statistical theory, making it easier to estimate confidence intervals for complex statistics [66]. Specifically, 1000 resampling are performed based on the original sentiment classification results, and the 95% confidence intervals for the proportion of each sentiment type are calculated.

3.3.1. Model validation and accuracy metrics

Since most DeepSeek-related tweets are collected from natural sources and lack manually annotated sentiment labels, the tweets are difficult to use directly for performance testing of sentiment classification models [67]. The tweets need to be aggregated, cleaned, and manually annotated, which is a time-consuming process and is not suitable for the rapid model performance verification stage. Therefore, a dataset "COVID-19 NLP Text Classification" is chosen from Kaggle, as it has manually annotated sentiment labels to verify the model's performance and the classification performance of VADER on the Twitter corpus. As shown in Table 1, this dataset contains English tweets posted by Twitter users during the COVID-19 pandemic. It has been manually annotated into five emotion categories: extremely positive, positive, neutral, negative, and extremely negative. Both datasets are informal text data from the Twitter platform. Second, both datasets address the current emerging social topics, where COVID-19 Tweets NLP focused on global public health events, while "Tweets and Reactions on DeepSeek" focused on DeepSeek

Figure 5
Stratified random sampling based on sentiment analysis

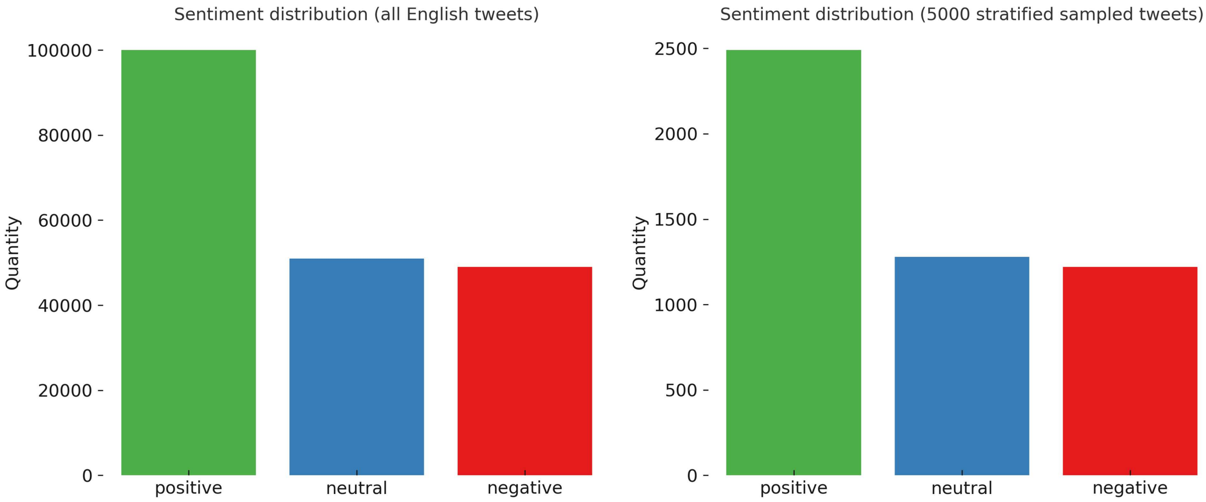


Table 1
Comparison of datasets

Attribute	Tweets and reactions on DeepSeek	COVID-19 tweets NLP text classification
Data source	Twitter	Twitter
Topic	Discussions and reactions to DeepSeek (AI)	COVID-19 pandemic-related discourse
Type of data	Raw tweet text (short text)	Raw tweet text (short text)
Language	English	English
Text length	Standard Twitter length (≤ 280 characters)	Standard Twitter length (≤ 280 characters)
Sentiment label	No label	Human-annotated (5 classes)

(*Continued*)

Table 1
(*Continued*)

Attribute	Tweets and reactions on DeepSeek	COVID-19 tweets NLP text classification
Label classes	No predefined classes	Extremely positive, positive, neutral, negative, extremely negative
Label source	Unlabeled	Manually labeled
Use of case	Real-world prediction and inference	Model training and performance evaluation
Academic use	Limited academic references	Widely used as a benchmark dataset in the literature

around recent hot models in the AI field. The data text is similar in length to original Twitter tweets, with a character length of 280 or fewer. Both reflect the emotional attitude of the public toward the significant events or technologies on social media and belong to the emotional corpus of time-sensitive topics. The emotional tendencies of the two-text data are obvious. Users often express strong positive, negative, or neutral opinions, such as appreciation, concern, and questions, in both datasets. Therefore, both are suitable as inputs for three categories of emotion classification tasks (positive/neutral/negative) and are highly adapted to the VADER model.

Both datasets demonstrate the extensive value of NLP applications in public discussion analysis, technology trend prediction, crisis management, and other fields, regardless of whether it involves the public discussion of the COVID-19 pandemic or the outcomes of AI-related technologies. To simplify the classification task and focus on the model's ability to recognize basic emotions, this study merges the five labels into three categories: positive, neutral, and negative. Consequently, this study uses standard evaluation indicators to verify the accuracy of the sentiment model in "COVID-19 NLP Text Classification," including accuracy [68], precision [69], recall [69], and F1-score [70].

The accuracy of a sentiment model measures the proportion of correct predictions made by the model and is found using Equation (1).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

where TP is a true example in which the model correctly predicts the number of positive classes, TN is a true example in which the model correctly predicts the number of negative classes, FP is a false example in which the model incorrectly predicts the number of positive classes, and FN is a false example in which the model incorrectly predicts the number of negative classes.

Continuously, the precision of a sentiment model indicates the proportion of positive samples predicted as positive, as determined by Equation (2). The recall of a sentiment model demonstrates the proportion of positive samples that are correctly predicted, as shown in Equation (3). Ultimately, the F1-score is the harmonic average of precision and recall, which aims to comprehensively evaluate model performance and is calculated using Equation (4).

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - score = 2 \times \frac{Precison \times Recall}{Precision + Recall} \tag{4}$$

3.4. Topic modeling

Topic modeling is a text mining technique used to automatically identify hidden topics from large amounts of text data [35]. The topic modeling method employed in this study is LDA, a generative model used to automatically discover potential topic structures from a large amount of unlabeled text. LDA is a commonly used tool in text analysis, NLP, and social media mining. Hence, the flowchart of topic modeling in this study is performed using RapidMiner. Initially, the preprocessed data is imported into the RapidMiner database. Next, the research review text is selected using the Select Attribute operator, and then it is converted into text format using the Nominal-to-Text operator. After that, a Data-to-Document operator is used to convert the cleaned data into a document collection. Finally, a Loop Collection operator is implemented to preprocess the text.

Through the Loop Collection process, the quality and consistency of the data can be significantly improved by unifying the cases, removing meaningless stop words, extracting stems, and filtering tokens by length [44]. It simplifies the data processing process, improving the accuracy and reliability of the analysis results. Ultimately, these steps help better understand the core content of the text data and support effective decision-making.

To enhance the quality of topics, coherence scores are critical for evaluating the quality of topics generated by topic models. They measure the semantic relatedness of words within a topic, which is essential for ensuring that the topics are meaningful and interpretable by humans [71]. In terms of social media data, they help in validating the quality of these topics and ensure that the insights derived are reliable and actionable. For example, coherence scores were used by Valencia et al. [72] to analyze topics from Facebook posts, revealing key themes related to personal relationships and school-related concerns. Finally, this study utilizes Gensim in Python to perform a topic consistency assessment by determining the optimal number of topics (k value). It conducts LDA modeling analysis in RapidMiner to effectively measure the semantic similarity between popular keywords in each topic. A series of models is evaluated with different k values, and the coherence score is plotted for comparison.

3.4.1. Word frequency analysis

Word frequency analysis is a frequently used tool for analyzing social media user discussions, which helps identify high-frequency words and topics. It provides a better understanding of the main content and emotions expressed in the dataset [73]. Furthermore, word frequency analysis can help businesses gain insight into the concerns of potential customers and their overall sentiment, enabling them to improve customer experience and adjust their business strategies accordingly. As such, the process flows for word frequency analysis in RapidMiner are performed to determine the frequency of the respective topics. The flow starts with Retrieve, which loads the input dataset, then continues with Select Attributes to narrow the examples to the textual fields of interest. The selected attributes feed into the Nominal-to-Text operator to convert any nominal-coded values into plain text examples. The text then enters Process Documents from Data operator, where standard text preprocessing is applied to produce a corpus-level word list, which is passed to the WordList-to-Data operator to convert the word list into a structured example set of word frequency attributes. Finally, the processed result is written out by the Store for further analysis.

3.4.2. Word cloud analysis

Word cloud analysis is a text visualization technique that highlights the most frequently used words in a body of text by displaying them in a visually prominent manner [74]. The size of each word in the cloud typically represents its frequency or importance within the text. It can help identify recurring themes, concerns, and emotions in qualitative data. By distilling text down to the most frequent words, word clouds offer a summarized representation of the document, which makes it easier to extract crucial information quickly. In the study of Xu et al. [75], an analysis of tweets about smartwatches used word clouds to extract and visualize product attributes and sentiments over time, which provided valuable insights into brand sentiment. Therefore, the word cloud analysis is carried out. The process begins with loading the input dataset into the process and sorting the examples by a frequency attribute, so the most relevant tokens rise to the top. Continuously, the Filter Example Range operator trims the sorted set to a defined slice that will be used in the cloud. Finally, the filtered list is exported to an Excel file for visualization or further analysis.

3.5. Topic sentiment analysis

The integration of topic information with SA can significantly improve the accuracy of sentiment classification. For instance, Wang et al. [76] demonstrated that topics can enhance SA by providing fine-grained sentiment information, resulting in improved prediction performance. The analysis is valuable in multiple fields such as marketing, customer care, and financial analysis, as it helps businesses understand consumer behavior to improve services by making informed decisions [77]. Therefore, this study conducts topic SA combining topic modeling with SA to show the variation of discussions and emotions by different topics. To carry it out in this study, each topic is assigned to a separate document, allowing a sentiment classifier to be run per document in Python. The flexibility of Python allows for the development of customized SA programs that can be tailored to specific needs, such as analyzing customer feedback [78]. As a result, the respective sentiment score can be computed to determine the sentiment according to different topics. Meanwhile, a heat map is used as a visualization tool for topic SA, which can quickly identify changes and trends in public discussion and is more effective than textual representations, particularly when using Python. Ultimately, a chi-square test is used to determine the significance level of the relationship between public sentiment and discussion topics, providing direct insight for the businesses in this study. The chi-square test is crucial for validating the significance of experimental results in SA, ensuring that improvements in accuracy and performance are statistically significant [79].

3.6. Libraries in Python

Python is widely used for SA on social media platforms, such as Twitter. Ilyas et al. [80] demonstrated that Python libraries, such as VADER and TextBlob, have been employed to analyze public sentiment toward government decisions and events, including Brexit. These tools help to quantify daily public sentiment and correlate it with real-world events. The libraries of Python facilitate the extraction and analysis of sentiments and topics from large datasets efficiently [81]. As a summary, the libraries used in this study are listed in Table 2.

Table 2
Summary of libraries used in Python

Analysis	Library	Usage
Data preprocessing	pandas	Reading, filtering, deduplicating tweet data from CSV files
	re	Text normalization via regex (e.g., URLs, mentions, hashtags)
	langdetect	Language detection using “detect” to retain English tweets
	numpy	Array-based data transformation, value aggregation, and statistical computation
Sentiment analysis	vadersentiment	Compound score-based sentiment classification into positive/neutral/negative
	sklearn.model_selection	Stratified sampling of tweets to ensure class balance in subsampling
	scipy.stats	Performing statistical significance tests (e.g., chi-square tests) on sentiment distributions
	numpy	Bootstrapping to compute confidence intervals for sentiment class proportions
Topic modeling	nltk	Tokenization and stop word removal for topic modeling preprocessing
	gensim	Building and evaluating LDA topic models with coherence scoring
	pyLDAvis	Interactive topic visualization of LDA model results
	tqdm	Progress bar tracking for loops in preprocessing and modeling
Visualization	matplotlib.pyplot	Bar charts for sentiment distribution and coherence evaluation
	seaborn	Sentiment comparison before and after sampling using count plots
	wordcloud	Creating positive/negative word clouds from tweet corpora
	seaborn.heatmap	Generating correlation or association heatmaps from sentiment/topic statistics

4. Results and Discussion

4.1. Sentiment analysis

By analyzing the 5000 tweets of datasets, as shown in Figure 6, the results of SA indicate that most Twitter users expressed positive sentiment, with 2585 of them being positive comments, which suggests that Twitter users are highly satisfied with DeepSeek-related events. Additionally, 1219 comments are classified as negative, indicating that DeepSeek has some shortcomings. In other words, the launch of DeepSeek has a negative impact and pressure on some users. By analyzing negative comments, businesses can gain valuable insight into the specific factors that led to user dissatisfaction with DeepSeek. By addressing these negative sentiments, the businesses can improve product development on certain AI-related features to sustain long-term growth. Among these comments, 1196 are classified as neutral. These neutral comments can be understood in the following three perspectives: (1) users have a general experience with the release of DeepSeek, (2) users do not care about the development of the LLM in the AI field, (3) it needs a more extended period to observe the performance of DeepSeek, where further understanding and experience may be needed. For the LLM environment in AI, such as DeepSeek, these perspectives show that there is an opportunity for continued performance and improvement. As an overview, the positive tweets account for 49.76%, neutral tweets account for 25.84%, and negative tweets account for 24.40%.

To further statistically verify the results of SA, this study employs the Bootstrap resampling method to estimate the 95% confidence interval for each sentiment ratio. The results in Table 3 indicate that the confidence intervals for positive sentiment are from 48.5% to 51.2%, negative sentiment is from 23.3% to 25.6%, and neutral sentiment is from 24.6% to 27.0%. Since the confidence intervals for each sentiment category are narrow, it means that the sample size is sufficient, and the sentiment results are statistically significant and reliable. Therefore, this study can reasonably conclude that DeepSeek has triggered positive user comments on social media.

Figure 6
Results of sentiment analysis

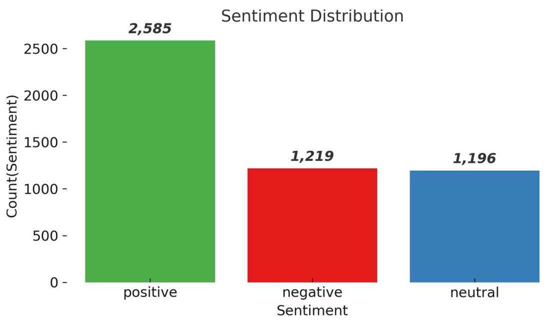


Table 3
Bootstrap resampling with 95% confidence level

Sentiment	Proportion %	95% confidence level
Positive	49.76	(48.50%, 51.20%)
Negative	24.40	(23.30%, 25.60%)
Neutral	25.84	(24.60%, 27.00%)

In terms of the accuracy of the sentiment model, the accuracy of the VADER model in the "COVID-19 NLP Text Classification" dataset is 94.42%, which indicates that the rule-based VADER model performs well in sentiment classification. The negative category performs best with an F1-score of 0.9519, as the negative sentiment vocabulary may be more concentrated and easier to identify. The neutral category performs relatively weakly (F1-score = 0.9138), which is attributed to the ambiguity of the VADER model when classifying neutral emotions.

These sentimental results provide businesses with a reliable reference point for public reaction to DeepSeek and a quantifiable basis for prioritizing decisions. Nearly half of the tweets expressed positive reactions, which shows healthy initial market acceptance, while almost a quarter expressed dissatisfaction, and another quarter remained neutral or undecided. The narrow bootstrap confidence intervals indicate that these proportions are stable and suitable for guiding resource allocation, rather than relying on anecdotes. Practically, this enables the product teams to extract recurring negative themes and trace them back to specific features or user journeys that need improvement. Marketing can responsibly focus on authentic positive commentary to build credibility, and customer success can be achieved by engaging neutral users with targeted education and demonstrations to accelerate adoption. The recurring negative themes are identified and mapped to support workflows and documentation, ensuring that feedback loops close quickly. Since the neutral detection in VADER is weaker, manual review samples may be added, and a fine-tuned domain classifier may be explored to capture the nuance. A dashboard can be created to unify sentiment, topics, and usage for weekly leadership reviews, where social sentiment can inform prioritization and investment decisions. Through these discussions, businesses can address the most frequent complaints while continuing to work on successful features.

4.2. Topic modeling

As shown in Figure 7, when $k = 10$, C_V ≈ 0.386; when $k = 50$, C_V ≈ 0.389. Although the coherence score is the largest when there are 50 topics, fewer topics can improve interpretability, and smaller k values are usually suitable for summary analysis with high requirements for human readability. Therefore, 10 topics are selected for LDA topic modeling analysis in this study.

As a result, there are five keywords associated with each topic in Table 4. The topic encompasses multiple dimensions, including model performance, national technological competition, capital market dynamics, and data privacy, which illustrates the multi-level discussion among users on DeepSeek and the AI ecosystem in which it is situated.

Topics 0 and 5 reveal the public interest in comparing DeepSeek to other LLMs. The keywords "train," "cost," and "code" highlight that the users focus on training expenses, efficiency, and output quality as the key concerns as AI models transition from research to market. Topic 1 reveals that the public is particularly concerned about the Chinese background and technological sovereignty issues associated with DeepSeek. Under the dominance of OpenAI and other models, discussions on DeepSeek as a model with strong Chinese processing capabilities have triggered comparisons with models from other countries due to its local positioning in China. This theme reflects the intersection between technology and national competition, which manifests the public concern about the strategic level of AI development.

Figure 7
Number of topics and coherence score

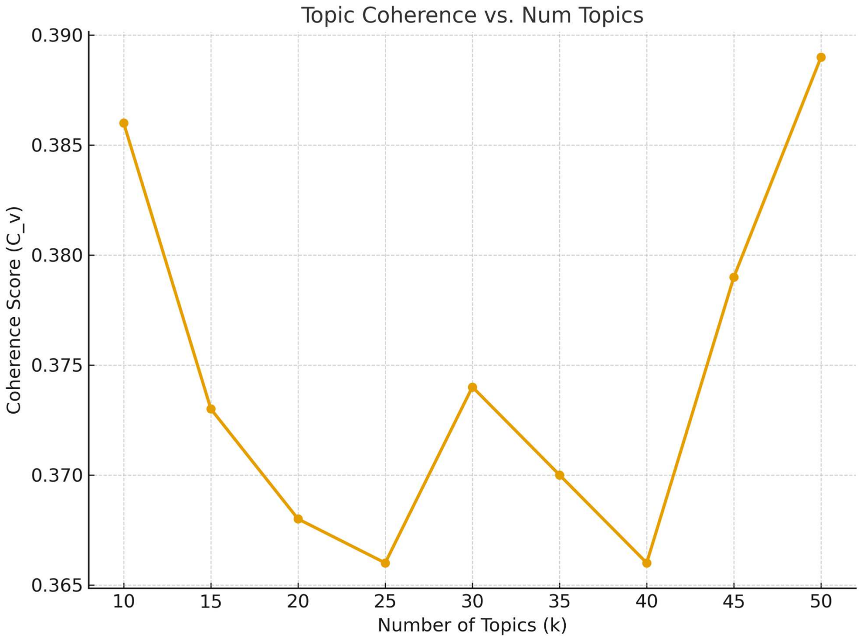


Table 4
Topics determined through topic modeling

Topic	Top keywords
0	model, DeepSeek, train, OpenAI, cost
1	DeepSeek, chines, China, OpenAI, tech
2	NVIDIA, game, card, core, Altman
3	NVIDIA, market, stock, DeepSeek, NVDA
4	DeepSeek, OpenAI, NVIDIA, people, think
5	ChatGPT, code, Claud, OpenAI, gener
6	data, secur, potenti, concern, user
7	comput, decentr, power, network, Aethir
8	ChatGPT, llama, stockmarketcrash, YouTube, Disney
9	crypto, Trump, Bitcoin, coin, Solana

Importantly, this study proves that multiple topics reflect the market impact of DeepSeek and NVIDIA. Topic 2 suggests that people associate the development of AI with the demand for hardware, especially NVIDIA graphics processing units (GPUs). Topic 3 focuses more on the reaction of the stock market and investment market, which reveals the collateral impact of releasing DeepSeek on NVIDIA's stock price fluctuations. Topic 4 reflects the views of several major AI giants. These topics demonstrate that the public focuses on both AI models from a technical perspective and interprets them within a broader industrial and financial ecosystem.

Topic 6 highlights public concerns about user data privacy and the potential risks associated with AI. The keywords such as "secure," "concern," and "potential" clearly convey that the public is attentive to information leakage and abuse that models like DeepSeek may cause. Negative emotions may accompany this theme, as evidenced by a higher proportion of neutral-to-negative comments in SA, which highlights the sensitivity and importance of this topic.

Topic 7 involves less popular but forefront topics, which are decentralized computing power and Web3 network infrastructure. The keyword "Aethir" refers to a distributed computing power platform, indicating that some users are exploring the relationship between decentralized computing power resources and AI model training and deployment.

Topic 8 contains multiple popular words, such as ChatGPT, Llama, YouTube, and Disney. It may reflect that some users discuss DeepSeek alongside other popular models, entertainment content, or network trends. This topic suggests that discussions about AI models on social media are not always focused on the technology itself, where users are more inclined to explore a broader pop culture context.

Topic 9 involves cryptocurrencies and politicians, which may have an indirect relationship with AI itself. It may indicate that AI is widely utilized in the cryptocurrency market to enhance security, predict price movements, and optimize trading strategies. Meanwhile, the rise of cryptocurrencies necessitates the development of regulatory frameworks to ensure their safe and effective use. Hence, the connection between these keywords may be more profound in the future.

These 10 topics reveal that the public discussion has become more diversified after the release of DeepSeek LLM. Twitter users are not only concerned with technical performance and economic benefits but also extend to political and cultural aspects, as well as international competition. LLM such as DeepSeek has gradually been integrated into the mainstream discourse system in the field of AI, which has received widespread attention.

The 10-topic LDA identified provides businesses with a clear and actionable view of the market frames of DeepSeek across technical, industrial, political, and cultural dimensions. The recognition that users debate model performance, training cost, and code quality allows engineering and product managers to prioritize improvements that address the most common complaints and enhance perceived value. The hardware and Nvidia-linked discussions may support strategic vendor partnerships and capacity planning, where supply and pricing risks can be managed

Table 5
Results of word frequency analysis

Word	Frequency	Word	Frequency
DeepSeek	2287	Company	324
NVIDIA	1500	Make	322
OpenAI	925	Data	316
Model	839	Cost	299
ChatGPT	814	Think	275
Chinese	629	Chip	268
Market	609	Train	265
China	568	Time	255
Stock	456	People	245
Tech	394	Dont	238

proactively. In the realm of national competition and sovereignty themes, they guide public relations and localization strategies that help the legal and policy teams in businesses prepare compliance and communications tailored to the relevant audiences. Since there are many conversations taking place in broader pop culture or platform contexts, it suggests that marketing should position the case studies and demos where those audiences actively engage, thereby keeping up with the trends of customers. Additionally, the clear presence of privacy and safety topics indicates where investment in governance, transparency, and safety tooling will most directly reduce neutral-to-negative sentiment. This alignment is designed to reduce time-to-market and enhance customer retention. Continuously, the exploration of decentralized computing power and Web3 indicates potential partnerships or pilot programs to reduce infrastructure costs and increase resilience. The product marketing of businesses can increase the positive conversations by seeding influencer and platform-specific content while monitoring coin- and politics-related threads to avoid brand entanglement.

4.2.1. Word frequency and word cloud analysis

From the results of topic modeling, the data is sorted in descending order from high to low to identify the most frequently used words. The most frequently used word is "DeepSeek," which appears 2,287 times, and the least frequently used word is "dont," which appears 238 times. The top 20 highest frequency words are determined, and the top five most frequently used words are "DeepSeek," "NVIDIA," "OpenAI," "model," and "ChatGPT."

As shown in Table 5 and Figure 8, the highest frequency of the word "DeepSeek" shows that the users pay great attention to the model itself. Second, "NVIDIA" and "OpenAI" frequently emerge as key companies in the current AI industry, indicating that users are keen to explore the relationship between DeepSeek and these prominent technological companies. This reflects public interest in the relationship between technological ecosystems and the industry's competitive landscape. In addition, the words "model" and "ChatGPT" show that the users focus on both the model structure and the technical framework adopted by DeepSeek by comparing it with ChatGPT, a representative product of OpenAI. As such, it demonstrates that users are focused on the performance of different LLMs to stay up-to-date with the latest AI features.

The high-frequency terms indicate that users focus on the product itself, as well as major industry players and comparative models, where attention should be concentrated and acted

Figure 8
Result of word cloud analysis

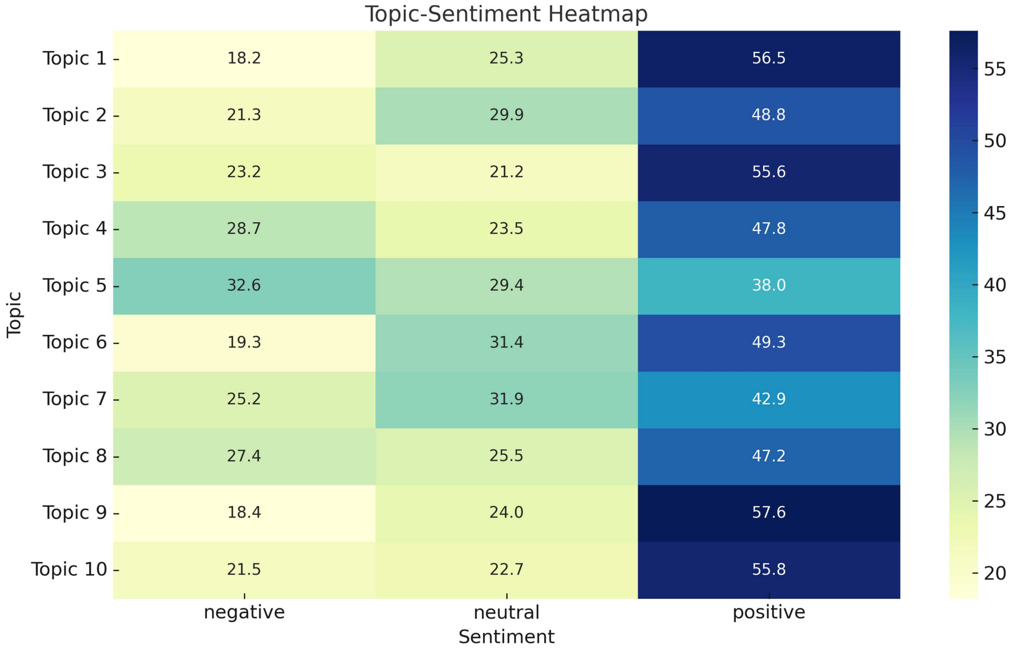
upon by businesses. By observing that users emphasize the product name and compare it to established platforms, product teams can prioritize comparative benchmarks, publish performance metrics, and create concise migration guides that reduce adoption friction. Partnerships with hardware providers should be revisited, as the supply and performance of GPUs are important to public discourse. Marketing and investor relations can leverage these keywords for targeted campaigns while crafting messages that directly address competitive positioning and ecosystem relationships. By optimizing documentation and community support around noted technical concerns, it can convert interest into trials and sustained usage, while reducing misinformation and improving product adoption, thereby strengthening its competitive position over time.

4.3. Topic sentiment analysis

Figure 9 presents the sentiment of users across various topics. In overview, the dominant sentiment across all topics is positive, as the public generally holds a positive attitude during discussions about DeepSeek. Particularly, topic 8 has the highest proportion of positive sentiment (57.6%), where users are confident in the performance of LLMs in fields other than AI, such as entertainment. Besides, topic 6 has the highest neutral emotion (31.9%). It may indicate that users continue to observe security and privacy issues, possibly caused by DeepSeek-like LLMs. On the other hand, the negative sentiment is lower than other sentiments on average, with the highest negative sentiment at topic 4 (32.6%). However, topic 4 has the lowest positive sentiment (38%), which may indicate that users are more divided on this topic or have more dissatisfied feedback, such as comparisons with competitors or performance doubts.

Ultimately, the chi-square test results in a p-value of 0.0000, indicating a statistically significant correlation between "topic" and "sentiment" as it is smaller than the significance level of 0.05. This study demonstrates that the topics discussed in tweets significantly affect the emotional tendencies and expressions of users across different topics. The emotions of users are influenced by various topics. This result indicates that the distribution of user emotions across different topics in DeepSeek-related tweets is not random, but rather closely tied to the topics being discussed. For example, some technical topics, such as model performance and computing power deployment, are more likely to be accompanied

Figure 9
Heat map of topic sentiment analysis



by neutral or positive comments. In contrast, topics involving market competition and capital relations are more likely to cause negative emotions.

The strong statistical link between topic and sentiment provides businesses with a reliable basis to tailor their products, communications, and risk strategies. By mapping the topics with their respective attractive positive, neutral, or negative reactions, it allows them to prioritize solutions quickly where sentiment is most negative and to highlight growth where the positive signal is strongest. For instance, the controversy and performance concerns associated with topic 4 may be addressed by promoting entertainment and cross-platform use cases related to topic 8. Marketing can create topic-specific campaigns and investor relations to proactively address capital market concerns, ensuring messaging aligns with stakeholder expectations. Operationally, businesses can route topic-tagged feedback to expert teams to accelerate improvement and measure the impact through follow-up SA. The combined topic sentiment tracking with retention and conversion metrics will reveal the interventions that move both perception and business outcomes.

5. Conclusion

This study has made a significant contribution to the theoretical framework of social media research, specifically in the context of tweet analysis and text mining methods for AI products. This study presents a transparent and replicable Python workflow for preprocessing tweet datasets, specifically for SA and topic modeling, which encompasses data cleaning and text normalization of user mentions. The data preprocessing flow is platform-independent and readily transferable to comment data from other social networks and e-commerce sites, such as Amazon and Taobao, allowing practitioners to extract customer opinions across various domains. By improving model input quality and interpretability, practitioners are not limited to an AI perspective, but are applicable to wider fields, such as discussions on pricing, product design, and loyalty initiatives. The product teams, marketers, and researchers are supported in making evidence-based decisions for strategic business outcomes. This study proposes a methodological framework that combines VADER SA and LDA topic modeling, suitable for the multidimensional evaluation of AI products in public discussions on social media. The results of SA indicate that most of the public sentiment toward DeepSeek-related events is positive. The topic modeling identifies a total of 10 major discussion topics, including model performance, competitors, and data security. Finally, the chi-square test results ($p < 0.05$) indicate a significant statistical correlation between topics and sentiment.

These findings have important implications for the development and public communication strategies of AI products. This study enables businesses to strategically allocate resources by prioritizing enhancements in areas with concentrated negative sentiment, such as data privacy and model performance, while actively promoting positive narratives around entertainment and cross-platform utility to bolster public perception. From a policy perspective, the identified concerns regarding technological sovereignty, data security, and market competition underscore the need for clear regulatory frameworks that foster innovation while ensuring the ethical deployment of AI and transparent user data governance. By aligning product development with sentiment-driven insights and advocating for policies that address public apprehensions, organizations can not only improve product-market fit but also contribute to the development of a trustworthy and sustainable AI ecosystem.

However, this study poses limitations, as the data obtained from comments posted by Twitter users may contain some commercial marketing accounts or discussions dominated by specific countries and regions, which can cause biases in the results. This study only analyzes a subset of the available Twitter data. Although the dataset contains temporal information, this study does not examine the evolution of topics or sentiment over time, as the timestamps of the dataset are likely irregular to support reliable trend analysis. This study is limited to English-language tweets, as multilingual analysis is not performed through language-specific annotation, model validation, and native-speaker review. Moreover, the VADER SA model is based on dictionary rules, which can lead to misjudgments when dealing with sarcasm, irony, or expressions with strong contextual dependence. This study primarily focuses on the public discussion on Twitter regarding DeepSeek-related events, excluding other external factors that may influence public sentiment, such as the market performance of related companies or the platform's algorithm in public discussion mechanisms.

To sustain the research, future studies may focus on overcoming these limitations and incorporate other complex sentiment models, such as RoBERTa-large, DeBERTa-v3-base, and other models to account for various possible factors. To enhance validity, future research should integrate graph-based network analysis to filter out non-human and spam accounts and employ multilingual transformer models, such as mBERT, rather than relying on lexicon translation. To address model limitations, hybrid approaches that combine deep contextual models, such as RoBERTa, with pragmatic rule-based filters are recommended to improve the detection of sarcasm and irony. Additionally, moving beyond ternary classification to fine-grained, aspect-specific sentiment scales with confidence metrics would more accurately capture neutral and ambiguous expressions, improving practical utility.

Funding Support

This work was supported by the Universiti Sains Malaysia, Research University Transdisciplinary (RUTrans) Grant Scheme (Grant Number: R502-KR-RUT002-0000000252-K134).

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/datasets/bwandowando/tweets-and-reaction-on-deepseek-models and https://www.kaggle.com/datasets/datatattle/covid-19-nlp-text-classification.

Author Contribution Statement

Wei Chien Ng: Conceptualization, Methodology, Resources, Writing – original draft, Supervision, Project administration, Funding acquisition. **Shiwen Chen:** Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft. **Quan Kai Ang:** Writing – review & editing, Visualization. **Yu Qing Soong:** Validation, Writing – review & editing.

References

[1] Darda, P., & Pendse, M. K. (2025). The impact of artificial intelligence (AI) transformation on the financial sector from the trading to security operations. In A. Khang (Ed.), *Shaping cutting-edge technologies and applications for digital banking and financial services* (pp. 322–339). Productivity Press. https://doi.org/10.4324/9781003501947-20

[2] Human-Centered Artificial Intelligence. (2025). *The AI Index 2025 Annual Report*. Stanford University. https://hai-production.s3.amazonaws.com/files/hai_ai_index_report_2025.pdf

[3] Shaikh, Z. P. (2025). Generative artificial intelligence in productivity and quality conformances: Modelling of generative adversarial networks (GANs). In A. Venaik, V. Chaturvedi, S. Saproo, & S. Bansal (Eds.), *Generative AI for business analytics and strategic decision making in service industry* (pp. 33–64). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-7026-1.ch002

[4] Raza, M., Jahangir, Z., Riaz, M. B., Saeed, M. J., & Sattar, M. A. (2025). Industrial applications of large language models. *Scientific Reports*, *15*(1), 13755. https://doi.org/10.1038/s41598-025-98483-1

[5] Raj, R., Singh, A., Kumar, V., & Verma, P. (2023). Analyzing the potential benefits and use cases of ChatGPT as a tool for improving the efficiency and effectiveness of business operations. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, *3*(3), 100140. https://doi.org/10.1016/j.tbench.2023.100140

[6] Xiong, L., Wang, H., Chen, X., Sheng, L., Xiong, Y., Liu, J., . . . , & Tang, Y. (2025). DeepSeek: Paradigm shifts and technical evolution in large AI models. *IEEE/CAA Journal of Automatica Sinica*, *12*(5), 841–858. https://doi.org/10.1109/JAS.2025.125495

[7] Yan, W., Liu, J., & Liang, W. (2025). DeepSeek fù néng quán kē yīxué: Qiánzài yìngyòng yǔ zhǎnwàng [DeepSeek empowers general medicine: Potential application and prospect]. *Chinese General Practice*, *28*(17), 2065. https://doi.org/10.12114/j.issn.1007-9572.2025.0023

[8] Zhang, H. (2025). How DeepSeek-R1 was created? *Journal of Shenzhen University Science and Engineering*, *42*(2), 226–232. https://doi.org/10.3724/SP.J.1249.2025.02226

[9] Dixon, S. J. (2025). *Number of social media users worldwide from 2019 to 2029 (in billions)*. *Statista Research Department*. https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

[10] Kada, A., Chouikh, A., Mellouli, S., Prashad, A. J., Straus, S. E., & Fahim, C. (2022). An exploration of Canadian government officials' COVID-19 messages and the public's reaction using social media data. *PLoS One*, *17*(9), e0273153. https://doi.org/10.1371/journal.pone.0273153

[11] Statista Research Department. (2025). *Number of Twitter users worldwide from 2019 to 2028 (in millions)*. https://www.statista.com/forecasts/1146722/twitter-users-in-the-world

[12] Zender, A., Humm, B. G., & Holzheuser, A. (2024). Successfully improving the user experience of an artificial intelligence system. In *Conference on Computer Science and Intelligence Systems*, *39*, 253–258. https://doi.org/10.15439/2024F2707

[13] Dirgová Luptaková, I., Pospíchal, J., & Huraj, L. (2023). Beyond code and algorithms: Navigating ethical complexities in artificial intelligence. In *Software Engineering Methods in Systems and Network Systems: Proceedings of 7th Computational Methods in Systems and Software 2023*, *2*, 316–332. https://doi.org/10.1007/978-3-031-54813-0_30

[14] Olasiuk, H., Kumar, S., Singh, S., Nagpal, G., & Ganushchak, T. (2023). Exploring the thematic clusters of artificial intelligence applications in supply chains using topic modelling and text mining: A machine learning insight. In

International Conference on Smart Technologies in Computing, Electrical and Electronics, 1–6. https://doi.org/10.1109/ICSTCEE60504.2023.10584847

[15] Yang, P., Han, K., & Diesner, J. (2024). Topics, temporal patterns, and network characteristics of AI-related discourse on Reddit. In *Social Networks Analysis and Mining: 16th International Conference*, 33–344. https://doi.org/10.1007/978-3-031-78554-2_22

[16] Gudigantala, N., Madhavaram, S., & Bicen, P. (2023). An AI decision-making framework for business value maximization. *AI Magazine*, *44*(1), 67–84. https://doi.org/10.1002/aaai.12076

[17] Witkowski, A., & Wodecki, A. (2025). Where does AI play a major role in the new product development and product management process? *Management Review Quarterly*. https://doi.org/10.1007/s11301-025-00533-5

[18] Choi, S. (2024). Temporal framing in balanced news coverage of artificial intelligence and public attitudes. *Mass Communication and Society*, *27*(2), 384–405. https://doi.org/10.1080/15205436.2023.2248974

[19] Gopalakrishnan, S., & Zhang, H. (2019). Client dependence: A boon or bane for vendor innovation? A competitive mediation framework in IT outsourcing. *Journal of Business Research*, *103*, 407–416. https://doi.org/10.1016/j.jbusres.2017.12.023

[20] Gharieb, A., Gabry, M. A., & Soliman, M. Y. (2024). The role of personalized generative AI in advancing petroleum engineering and energy industry: A roadmap to secure and cost-efficient knowledge integration: A case study. In *SPE Annual Technical Conference and Exhibition*. https://doi.org/10.2118/220716-MS

[21] Chan, L., Xu, X., & Lv, K. (2025). DeepSeek-R1 and GPT-4 are comparable in a complex diagnostic challenge: A historical control study. *International Journal of Surgery*, *111*(6), 4056–4059. https://doi.org/10.1097/JS9.0000000000002386

[22] Aramburu, M. J., Berlanga, R., & Lanza-Cruz, I. (2021). Quality management in social business intelligence projects. In *Proceedings of the 23rd International Conference on Enterprise Information Systems*, *1*, 320–327. https://doi.org/10.5220/0010495703200327

[23] Al-Otaibi, S., Alnassar, A., Alshahrani, A., Al-Mubarak, A., Albugami, S., Almutiri, N., & Albugami, A. (2018). Customer satisfaction measurement using sentiment analysis. *International Journal of Advanced Computer Science and Applications*, *9*(2), 106–117. https://doi.org/10.14569/IJACSA.2018.090216

[24] Sindhu, S., Kumar, S., & Noliya, A. (2023). A review on sentiment analysis using machine learning. In *2023 International Conference on Innovative Data Communication Technologies and Application*, 138–142. https://doi.org/10.1109/ICIDCA56705.2023.10099665

[25] Suzuki, H., Miyauchi, Y., Akiyama, K., Kajiwara, T., Ninomiya, T., Takemura, N., . . . , & Nagahara, H. (2022). A Japanese dataset for subjective and objective sentiment polarity classification in micro blog domain. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 7022–7028.

[26] Sutedja, I., & Hendry. (2025). Sentiment analysis: An insightful literature review. *International Journal of Advanced Computer Science and Applications*, *16*(3), 517–522. https://doi.org/10.14569/IJACSA.2025.0160351

[27] Murthy, A., Deshpande, P., Gakhreja, S., Garwal, Y., Jain, N., & Katare, R. (2024). Natural language processing for sentiment analysis in marketing. In *2024 Second International Conference Computational and Characterization Techniques in Engineering & Sciences*, 1–5. https://doi.org/10.1109/IC3TES62412.2024.10877630

[28] Asghar, Z., Ali, T., Ahmad, I., Tharanidharan, S., Nazar, S. K. A., & Kamal, S. (2018). Sentiment analysis on automobile brands using Twitter data. In *International Conference on Intelligent Technologies and Applications: First International Conference*, 76–85. https://doi.org/10.1007/978-981-13-6052-7_7

[29] Singh, K. U., Kumar, A., Kumar, G., Choudhury, T., Singh, T., & Kotecha, K. (2023). Sentiment analysis in social media marketing: Leveraging natural language processing for customer insights. In *ICT: Smart Systems and Technologies: Proceedings of ICTCS 2023*, *4*, 457–467. https://doi.org/10.1007/978-981-99-9489-2_40

[30] Rao, J. N., Malleswari, Y. D., Krishna, K. T., Sai, C. V., & Sri, Y. R. (2024). Twitter sentiment analysis for different business applications. *Journal of Advanced Research in Dynamical and Control Systems*, *12*(2), 1048–1055.

[31] Vijay, K., Samuel, P., Krishna, B. V., & Manikandan. (2024). Exploration of sentiment analysis in Twitter propaganda: A deep dive. *Multimedia Tools and Applications*, *83*(15), 44729–44751. https://doi.org/10.1007/s11042-023-17383-6

[32] Dhanalakshmi, P., Kumar, G. A., Satwik, B. S., Sreeranga, K., Sai, A. T., & Jashwanth, G. (2023). Sentiment analysis using VADER and logistic regression techniques. In *International Conference on Intelligent Systems for Communication, IoT and Security*, 139–144. https://doi.org/10.1109/ICISCoIS56541.2023.10100565

[33] Mountstephens, J., & Quen, M. T. Z. (2022). Mixed-language sentiment analysis on Malaysian social media using translated VADER and normalisation heuristics. In *Third Congress on Intelligent Systems: Proceedings of CIS 2022*, *2*, 185–196. https://doi.org/10.1007/978-981-19-9379-4_15

[34] Abubakir, S., Baimbet, A., Kabieva, A., Sarsenov, B., Seidilabek, A., Akhmetov, I., & Gelbukh, A. (2024). Sentiment analysis for religious tweets. *Computación y Sistemas*, *28*(4), 1911–1918. https://doi.org/10.13053/CyS-28-4-5208

[35] George, M. (2018). Unsupervised topic detection based on 2D vector space model using Apriori algorithm and NLP. In *International Conference on Digital Information Management*, 279–283. https://doi.org/10.1109/ICDIM.2018.8846982

[36] Strydom, I. F., Grobler, J., & Vermeulen, E. (2023). Investigating the use of topic modeling for social media market research: A South African case study. In *Computational Science and Its Applications: 23rd International Conference*, 305–320. https://doi.org/10.1007/978-3-031-36808-0_20

[37] Yazıcı, G., & Ozansoy Çadırcı, T. (2024). Creating meaningful insights from customer reviews: A methodological comparison of topic modeling algorithms and their use in marketing research. *Journal of Marketing Analytics*, *12*(4), 865–887. https://doi.org/10.1057/s41270-023-00256-0

[38] Nasereddin, A. Y. (2023). A business analytics approach to strategic management using uncovering corporate challenges through topic modeling. *Information Sciences Letters*, *12*(5), 1717–1733. http://dx.doi.org/10.18576/isl/120518

[39] Vukanti, V., & Jose, A. (2021). Business analytics: A case-study approach using LDA topic modelling. In *2021 5th International Conference on Computing Methodologies and Communication*, 1818–1823. https://doi.org/10.1109/ICCMC51019.2021.9418344

[40] Sharaff, A., Papatla, S. R., & Rajput, N. (2024). Analyzing farmers' protest tweets using topic modelling techniques. In *OPJU International Technology Conference on Smart Computing for Innovation and Advancement in Industry 4.0*, 1–6. https://doi.org/10.1109/OTCON60325.2024.10687790

[41] Cripps, H., Singh, A., Mejtoft, T., & Salo, J. (2020). The use of Twitter for innovation in business markets. *Marketing Intelligence and Planning*, *38*(5), 587–601. https://doi.org/10.1108/MIP-06-2019-0349

[42] D, S. J. (2024). Cross-media topic modeling: A comparative analysis of NMF, LDA and CoRex combined with BERT on Quora and Twitter. In *International Conference on Computing and Data Science*, 1–6. https://doi.org/10.1109/ICCDS60734.2024.10560395

[43] Espinoza, R., Kishimoto, K., Liu, C., Santos, L. H. O., Mori, Y., Yamamoto, G., & Kuroda, T. (2025). Leveraging social media for public health: NLP implementations for blood donation data analysis in Japan. *Social Network Analysis and Mining*, *15*(1), 32. https://doi.org/10.1007/s13278-025-01437-8

[44] Tekin, Y. (2024). Initialization in Gibbs sampling implementation of LDA. In *Signal Processing and Communications Applications Conference*, 1–4. https://doi.org/10.1109/SIU61531.2024.10600919

[45] Azizi, F., Hajiabadi, H., Vahdat-Nejad, H., & Khosravi, M. H. (2023). Detecting and analyzing topics of massive COVID-19 related tweets for various countries. *Computers and Electrical Engineering*, *106*, 108561. https://doi.org/10.1016/j.compeleceng.2022.108561

[46] Tan, J. S., & Chia, W. C. (2022). Research output to industry use: A readiness study for topic modelling with sentiment analysis. In *8th International Conference on Computational Science and Technology*, 13–25. https://doi.org/10.1007/978-981-16-8515-6_2

[47] Yang, Z., Men, H., & Ingham, R. (2022). Media representation of China in COVID-19 reports: Text mining of the language of New York Times. In *European Conference on Natural Language Processing and Information Retrieval*, 101–107. https://doi.org/10.1109/ECNLPIR57021.2022.00030

[48] Osmani, A., Mohasefi, J. B., & Gharehchopogh, F. S. (2020). Enriched latent Dirichlet allocation for sentiment analysis. *Expert Systems*, *37*(4), e12527. https://doi.org/10.1111/exsy.12527

[49] Irawan, M. I., & Jamhuri, M. (2022). State of the art of machine learning: An overview of the past, current, and the future research trends in the era of quantum computing. *AIP Conference Proceedings*, *2641*(1), 040009. https://doi.org/10.1063/5.0131848

[50] Singla, S., Gupta, G., & Bhathal, G. S. (2023). Potential pre-processing techniques of data mining on mutual funds. *AIP Conference Proceedings*, *2916*(1), 130001. https://doi.org/10.1063/5.0178820

[51] Sharma, R., Sandhu, J., & Bharti, V. (2024). Experimental analysis of a multimodal biometric system using preprocessing and feature extraction techniques and their impact on analytical results. In *2024 6th International Conference on Computational Intelligence and Communication Technologies*, 212–219. https://doi.org/10.1109/CCICT62777.2024.00043

[52] Timbers, T., Campbell, T., Lee, M., Ostblom, J., & Heagy, L. (2024). *Data science: A first introduction with Python*. USA: CRC Press. https://doi.org/10.1201/9781003438397

[53] Chavan, R., Latthe, S., Dhorepati, M., Suryawanshi, A., Sharma, N., & Salge, A. (2024). Sentiment analysis using VADER & word cloud techniques. *AIP Conference Proceedings*, *3217*(1), 020012. https://doi.org/10.1063/5.0234543

[54] Veliz, C. M., de Clercq, O., & Hoste, V. (2021). Is neural always better? SMT versus NMT for Dutch text normalization. *Expert Systems with Applications*, *170*, 114500. https://doi.org/10.1016/j.eswa.2020.114500

[55] Ali, A., Emran, N. A., & Asmai, S. A. (2021). Missing values compensation in duplicates detection using hot deck method. *Journal of Big Data*, *8*(1), 112. https://doi.org/10.1186/s40537-021-00502-1

[56] Agustini, M., Fithriasari, K., & Prastyo, D. D. (2023). The impact of clustering-based sequential multivariate outliers detection in handling missing values. In *Data Science and Emerging Technologies: Proceedings of DaSET 2023*, 221–235. https://doi.org/10.1007/978-981-97-0293-0_17

[57] Sangeetha, J., & Anu, V. M. (2023). A conceptual Aquila Merged Arithmetic Optimization (AIAO) integrated Auto-Encoder Based Long Short Term Memory (AUE-LSTM) for sentiment analysis. *Journal of Applied Engineering and Technological Science*, *5*(1), 213–228. https://doi.org/10.37385/jaets.v5i1.2825

[58] Liya, B. S., Indumathy, P., Hemlathadhevi, A., & Dharaniya, R. (2026). Cascaded adaptive dilated temporal convolution network-based efficient sentiment analysis model from social media posts. *International Journal of Image and Graphics*, *26*(03), 2650015. https://doi.org/10.1142/S0219467826500154

[59] Emaduddin, S. M., Ullah, R., Mazahir, I., & Uddin, M. Z. (2022). Enhancing information preservation in social media text analytics using advanced and robust pre-processing techniques. *International Journal of Media and Information Literacy*, *7*(1), 60–70. https://doi.org/10.13187/ijmil.2022.1.60

[60] Amalia, R. N., Sadik, K., & Notodiputro, K. A. (2023). A preliminary study of sentiment analysis on COVID-19 news: Lesson learned from data acquisition, pre-processing, and descriptive analytics. *Barekeng: Jurnal Ilmu Matematika dan Terapan*, *17*(4), 1901–1914. https://doi.org/10.30598/barekengvol17iss4pp1901-1914

[61] de Haan-Ward, J., Woolford, D. G., & Bonner, S. J. (2025). Predicting rare events using training data from stratified sampling designs, with application to human-caused wildfire prediction. *Canadian Journal of Statistics*, *53*(3), e70008. https://doi.org/10.1002/cjs.70008

[62] Pradhani, R. A., Khair, F., Mirdad, A. J., Rumangkit, S., & Ferlin, C. B. L. (2025). Tech-driven strategies for brand loyalty: The impact of social media marketing on trust and equity in fashion brands. In *International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation*, *3*, 1–5. https://doi.org/10.1109/IATMSI64286.2025.10985600

[63] Htait, A., & Azzopardi, L. (2021). Sentiment intensity prediction using neural word embeddings. In *Proceedings of the 2021 ACM SIGIR International Conference on Theory of Information Retrieval*, 93–102. https://doi.org/10.1145/3471158.3472254

[64] Nurcahyawati, V., & Mustaffa, Z. (2023). Vader lexicon and support vector machine algorithm to detect customer sentiment orientation. *Journal of Information Systems Engineering and Business Intelligence*, *9*(1), 108–118. https://doi.org/10.20473/jisebi.9.1.108-118

[65] Md Saad, N. H., Mei Nie, F., & Yaacob, Z. (2025). Exploring sentiment analysis of online food delivery services post

COVID-19 pandemic: GrabFood and foodpanda. *Journal of Foodservice Business Research*, *28*(5), 1087–1111. https://doi.org/10.1080/15378020.2023.2294406
[66] Moslim, N. H., Zubairi, Y. Z., Hussin, A. G., Hassan, S. F., & Mokhtar, N. A. (2020). Construction of confidence interval of the parameter in von Mises distribution using bootstrap methods. *ASM Science Journal*, *13*, 1–7. https://doi.org/10.32802/asmscj.2020.sm26(1.6
[67] Chen, Z., Zhang, J., & Yuan, H. (2024). Aspect target sentiment classification with instance retrieval-based prompt template selection. In *International Conference on Electrical Engineering, Big Data and Algorithms*, 371–375. https://doi.org/10.1109/EEBDA60612.2024.10486026
[68] Alsagri, H. S., & Sohail, S. S. (2024). Fractal-inspired sentiment analysis: Evaluation of large language models and deep learning methods. *Fractals*. http://doi.org/10.1142/S0218348X24400565
[69] Wang, Y. (2024). A curriculum system development method of embedded curriculum innovation and entrepreneurship based on sentiment analysis model. In *2024 8th Asian Conference on Artificial Intelligence Technology*, 24–28. https://doi.org/10.1109/ACAIT63902.2024.11021838
[70] Zulfadli, Ilham., A, A., & Indrabayu. (2023). Sentiment analysis with soft-voting method on customer reviews for purchasing transactions of e-commerce. In *International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology*, 295–299. https://doi.org/10.1109/IAICT59002.2023.10205954
[71] Pereira, A., Viegas, F., Gonçalves, M. A., & Rocha, L. (2023). Evaluating the limits of the current evaluation metrics for topic modeling. In *Proceedings of the 29th Brazilian Symposium on Multimedia and the Web*, 119–127. https://doi.org/10.1145/3617023.3617040
[72] Valencia, J. D. M., Laure, A. J. T., Centino, N. M. R., Fabito, B. S., Imperial, J. M. R., Rodriguez, R. L., . . . , & Jamis, M. N. (2019). Understanding anonymous social media posts using topic modeling. In *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management*, 1–4. https://doi.org/10.1109/HNICEM48295.2019.9072791
[73] Putra, D. T., & Setiawan, E. B. (2023). Sentiment analysis on social media with glove using combination CNN and RoBERTa. *Jurnal RESTI: Rekayasa Sistem dan Teknologi Informasi*, *7*(3), 457–563. https://doi.org/10.29207/resti.v7i3.4892
[74] Feng, K. J. K., Gao, A., & Karras, J. S. (2022). Towards semantically aware word cloud shape generation. In *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, 35. https://doi.org/10.1145/3526114.3558724
[75] Xu, Z., Kohli, A., Nkhalamba, S., & Gai, L. (2025). Analyzing product attributes and brand sentiment of smartwatches using Twitter/X data from a time series perspective. *Journal of Marketing Analytics*, *13*(3), 885–901. https://doi.org/10.1057/s41270-024-00349-4
[76] Wang, S., Zhou, G., Lu, J., Chen, J., & Xia, Y. (2023). Topic enhanced sentiment co-attention BERT. *Journal of Intelligent Information Systems*, *60*(1), 175–197. https://doi.org/10.1007/s10844-022-00749-x
[77] Zhou, T., Law, K., & Creighton, D. (2022). A weakly-supervised graph-based joint sentiment topic model for multi-topic sentiment analysis. *Information Sciences*, *609*, 1030–1051. https://doi.org/10.1016/j.ins.2022.07.126
[78] Amrutphale, Y., Vijayvargiya, N., & Malviya, V. (2020). A novel adaptive approach for sentiment analysis on social media data. In *Social Networking and Computational Intelligence: Proceedings of SCI-2018*, 717–726. https://doi.org/10.1007/978-981-15-2071-6_60
[79] Zheng, L., Wang, H., & Gao, S. (2018). Sentimental feature selection for sentiment analysis of Chinese online reviews. *International Journal of Machine Learning and Cybernetics*, *9*(1), 75–84. https://doi.org/10.1007/s13042-015-0347-4
[80] Ilyas, S. H. W., Soomro, Z. T., Anwar, A., Shahzad, H., & Yaqub, U. (2020). Analyzing Brexit's impact using sentiment analysis and topic modeling on Twitter discussion. In *Proceedings of the 21st Annual International Conference on Digital Government Research*, 1–6. https://doi.org/10.1145/3396956.3396973
[81] Zhang, S., Ly, L., Mach, N., & Amaya, C. (2022). Topic modeling and sentiment analysis of Yelp restaurant reviews. *International Journal of Information Systems in the Service Sector*, *14*(1), 1–16. https://doi.org/10.4018/IJISSS.295872

How to Cite: Ng, W. C., Chen, S., Ang, Q. K., & Soong, Y. Q. (2026). Public Discussion of DeepSeek Large Language Model on Twitter: A Mixed-Methods Sentiment and Topic Modeling. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA62027447