Journal of Business and Digital Innovation (JBDI)
Volume 1, Issue 1, January 2026, PP 117-132
ISSN 2978-4522 (Print) & ISSN 2978-4530 (Online)
https://doi.org/10.19164/jbdi.v1i1.1773

Sentiment Analysis of Public Perceptions on ChatGPT and Generative Artificial Intelligence (GenAI): Model's Performance Evaluation and Examining Benefits and Risks in Education and Healthcare

[1]Mifta Uddin Khan & [2]Dilek Celik
[1]Independent Researcher, United Kingdom
[2]Northumbria University, London

Abstract
As GenAI systems become more integrated into daily activities, understanding how people react to these tools is critical for responsible design and governance. This study provides a large-scale, longitudinal analysis of public sentiment toward ChatGPT and GenAI by integrating transformer-based sentiment classification, temporal trend analysis, and sector-specific topic modeling for education and healthcare. Using over one million English-language posts collected between November 2022 and December 2023, we quantify sentiment patterns over time and identify domain-specific themes of perceived benefits and risks. A comparative evaluation of traditional machine-learning (ML) models (logistic regression, support vector machines, random forest), deep learning (DL) architectures (convolutional neural networks, long short-term memory), and Bidirectional Encoder Representations from Transformers (BERT) was conducted. DL models outperform classical ML, and BERT emerges as the most effective classifier, achieving 98% accuracy with a near-balanced profile across accuracy, precision, recall, and F1-score, outperforming traditional approaches. Using the best-performing model, the findings show that ChatGPT sentiment is predominantly positive, alongside a substantial minority of negative sentiments. Topic modeling reveals domain-specific benefits and risks in education and healthcare discourse. In education, ChatGPT promotes personalized learning, accessibility, and teacher support, but it also raises plagiarism, academic dishonesty, and data privacy concerns. In healthcare, GenAI improves patient information, diagnostics, and administrative efficiency, but it also raises concerns about misinformation, ethics, and empathy. Overall, the research provides evidence-based guidance for technology developers, educators, healthcare professionals, and policymakers taking advantage of GenAI while addressing its associated social and ethical issues.

Keywords: SA, ChatGPT, GenAI, Education, Healthcare
Wordcount: 244

1.0 Introduction

The spread of GenAI has intensified conversations about accuracy, ethics, data security, and the longer-term implications of machine-generated content. Since the public release of ChatGPT in late 2022, large language models (LLMs) have rapidly moved from experimental systems to tools embedded in education, healthcare, and everyday knowledge work. Recent evidence syntheses show that ChatGPT is already being trialled for clinical decision support, documentation, and patient communication in healthcare, as well as for feedback, explanation, and writing support in universities (Iqbal et al., 2025; Dos, 2025; Baxter et al., 2025). Tools such as ChatGPT, built on LLM, have drawn significant attention for their ability to simulate human-like conversation and produce coherent text across domains. This rapid adoption has generated both excitement and apprehension, as individuals and organizations question how such systems may influence work practices, learning environments, and decision-making processes (Koonchanok, Pan and Jang, 2024). Parallel to these advancements, SA is a branch of natural language processing focused on identifying emotions, attitudes, and opinions expressed in text has become a central tool for assessing public reactions to emerging technologies (Mao, Liu and Zhang, 2024).

However, examining sentiment regarding GenAI poses distinct challenges. Irony, sarcasm, quickly changing terms, and hidden emotional cues are all common in online discussions. These are difficult to spot with traditional lexicon-based methods (Qi and Shabrina, 2023). SA provides a way to quantify these evolving perceptions. Recent methodological reviews emphasize that transformer-based models now dominate sentiment and emotion analysis on large, noisy text corpora, and generally outperform traditional lexicon- or rule-based approaches (Bermudez-Sosa, Olarte-Henao and Rojas-Berrío, 2025).

ChatGPT has become a well-known example of SA due to its wide range of applications. According to Khan et al. (2024), the healthcare industry is the most active in terms of application-related publications, followed by education, computing, and research. Despite some concerns about academic integrity and the possibility of student overreliance, educational researchers have highlighted its ability to support personalized learning and increase access to resources (Chan and Hu, 2023). Concerns about false information, moral issues, and the decline of human judgment persist, despite discussions about ChatGPT as a tool to enhance diagnostics, administrative tasks, and patient communication in healthcare (Zhang and Kamel Boulos, 2023; Sallam, 2023). Given these contrasting views, understanding how the public interprets the benefits and risks of GenAI is essential. Social media platforms serve as a valuable barometer of public perception, offering real-time reflections of enthusiasm, caution, and emerging concerns. Analysing sentiment at scale can support policymakers, technology developers, educators, and healthcare professionals as they navigate the opportunities and challenges associated with GenAI.

1.1 Problem Statement

Several issues still impede a thorough understanding of public sentiment, despite the rapidly growing body of research on ChatGPT and GenAI. First, a lot of previous research is limited to brief time periods right after ChatGPT was released, which only records initial responses and ignores how public perceptions change as users are exposed to the technology over time (Liu and Lyu, 2024). Second, previous work's methodological approaches usually depend on lexicon-based sentiment tools or a single classical ML model, which frequently fails to identify complex, context-dependent expressions like ambivalence, sarcasm, or minority opinions (Qi and Shabrina, 2023; Tan, Lee and Lim, 2023). These restrictions limit the precision and dependability of SA, especially in dynamic online settings. Moreover, research within specific domains—such as education and healthcare—tends to focus on narrowly defined stakeholder groups, including students, educators, or clinicians, using qualitative methods such as interviews, surveys, or small-scale case studies (Chan and Hu, 2023; Sallam, 2023). While these studies offer valuable insights, they often fail to capture broader public discourse or the ways in which perceptions are shaped and circulated through open social media platforms. In parallel, large-scale analyses of social media discourse have tended to treat ChatGPT as a monolithic topic, without disaggregating public sentiment across critical sectors such as education and healthcare, thereby overlooking sector-specific attitudes and concerns (Khan et al., 2024).

1.2 Purpose and Objectives

This study aims to produce a comprehensive analysis of public perceptions of ChatGPT and GenAI through SA of large-scale social media data. This study has the following objectives:

1. To compare the performance of traditional ML, DL, and transformer-based models for sentiment classification of social media posts about ChatGPT and GenAI.
2. To determine the overall distribution of public sentiment (positive, negative, and neutral) towards ChatGPT and GenAI.
3. To analyze how public sentiment towards ChatGPT changes over time across the study period.
4. To identify the main perceived benefits and risks of ChatGPT within education and healthcare using topic modelling.

1.3 Research Questions

The study is guided by the following research questions:

1. How do traditional ML, DL, and transformer-based models compare in terms of accuracy and other performance metrics for sentiment classification of ChatGPT-related social media posts?
2. What is the overall distribution of sentiment (positive, negative, and neutral) towards ChatGPT and GenAI in the collected dataset?
3. How does public sentiment towards ChatGPT evolve over the study period?
4. What benefits and risks do users associate with ChatGPT in the contexts of education and healthcare, as reflected in social media discussions?

This study distinguishes itself from prior research in three key respects. First, it conducts a large-scale, longitudinal SA of ChatGPT and GenAI over fourteen months, capturing how public attitudes evolve beyond initial release reactions. Second, it systematically compares traditional ML, DL, and transformer-based models using a consistent evaluation framework, providing robust evidence of model performance differences in noisy social media data. Third, it integrates SA with sector-specific topic modelling to examine perceived benefits and risks in education and healthcare, moving beyond aggregate sentiment to domain-sensitive interpretation.

2.0. Literature Review

2.1 Sentiment Analysis: Definitions and Techniques

SA, a core task within natural language processing, is widely used to infer attitudes, opinions, and emotions expressed in text, typically by classifying polarity as positive, negative, or neutral (Darraz et al., 2024). While early applications focused on document-level polarity, contemporary research has expanded SA to address more complex tasks such as aspect-based sentiment analysis, emotion recognition, conversational sentiment, irony detection, and misinformation-related sentiment (Mao, Liu and Zhang, 2024). Despite this diversification, the literature consistently identifies three dominant methodological paradigms—lexicon-based approaches, classical ML, and DL. Recent systematic reviews show a clear shift toward ML- and DL-based methods for analysing large-scale, unstructured social media data (Bermudez-Sosa, Olarte-Henao and Rojas-Berrío, 2025; Ramezani, 2025).

2.1.1. Lexicon-Based Approaches

Lexicon-based SA relies on predefined sentiment dictionaries to assign polarity scores to words or phrases, which are subsequently aggregated at the sentence or document level (Raees and Fazilat, 2024). Tools such as TextBlob and VADER remain popular due to their ease of deployment, transparency, and independence from labelled training data (Qi and Shabrina, 2023). These characteristics make lexicon-based approaches attractive for exploratory or resource-constrained studies. However, extensive prior research demonstrates that lexicon-based methods are poorly suited to the linguistic complexity of social media discourse. They struggle to handle sarcasm, implicit sentiment, rapidly evolving slang, and domain-specific terminology—features that are particularly prevalent in online discussions of artificial intelligence and emerging technologies (Qi and Shabrina, 2023; Mao, Liu and Zhang, 2024). Moreover, the assumption that word polarity remains stable across contexts limits their ability to capture nuanced or contested opinions. As a result, while lexicon-based approaches remain useful for coarse-grained analysis, they risk oversimplifying public sentiment in debates surrounding GenAI and ChatGPT.

2.1.2. Classical ML Approaches

Classical supervised ML models, including Support Vector Machines, Logistic Regression, Naïve Bayes, Decision Trees, Random Forests, and K-Nearest Neighbours, have been extensively applied to sentiment classification tasks (Alslaity and Orji, 2024). Operating primarily on bag-of-words or TF–IDF representations, these models have demonstrated strong performance on many benchmark datasets. For example, Alsemaree et al. (2024) report high accuracy for ensemble-based sentiment classification in Arabic social media data, while Wu and Gao (2024) achieve over 90% accuracy in airline-related tweet classification using SVMs. Comparable results have also been reported for e-commerce and pandemic-related SA (Tabany and Gueffal, 2024).

Despite these reported accuracies, comparative studies highlight important limitations of classical ML approaches when applied to complex and evolving domains. Their reliance on manual feature engineering makes them sensitive to domain shift, and their performance degrades in the presence of noisy language, sarcasm, and minority sentiment classes (Alslaity and Orji, 2024; Diri, Obiorah, and Du, 2025). These limitations are particularly problematic for analysing GenAI-related discourse, where negative or highly concerned opinions—often of greatest substantive interest—may be underrepresented and linguistically subtle.

2.1.3. DL and Transformer-Based Approaches

Deep learning models such as Convolutional Neural Networks and Long Short-Term Memory networks address some of the shortcomings of classical ML by enabling automatic feature extraction and improved modelling of sequential text data. Reviews of SA on social networks identify CNNs, LSTMs, and transformer architectures as the dominant approaches in large-scale applications (Ramezani, 2025). These models generally outperform classical ML methods by capturing local semantic patterns and temporal dependencies. Nevertheless, DL models introduce new challenges, including high computational requirements, dependence on large-labelled datasets, and limited interpretability of internal representations (Krishna et al., 2023). These issues are especially salient in sensitive domains such as healthcare, where transparency and accountability are critical considerations.

Transformer-based models, particularly Bidirectional Encoder Representations from Transformers (BERT), represent a further methodological advance. Pre-trained on large corpora and fine-tuned for downstream tasks, BERT has demonstrated strong performance across a wide range of NLP applications, including SA (Tan, Lee and Lim, 2023). Empirical studies show that transformer-based architectures can capture nuanced, context-dependent, and aspect-level sentiment that earlier models often miss, as demonstrated in social media analyses using RoBERTa-based and hybrid transformer-based frameworks (Jahanbin and Chahooki, 2023). Despite these advantages, transformer-based approaches remain underutilised in large-scale, applied SA of social media data (Rodríguez-Ibánez Sw., 2023).

2.2. Public Perception and Sentiment Towards ChatGPT

Following the emergence of ChatGPT, scholars examined its perception within public discourse. Su and Kabala (2023) utilized transfer learning and topic modelling to analyse approximately 500,000 tweets concerning ChatGPT, leading to the identification of a range of opinions. There exists optimism regarding work efficiency, mixed with concerns about misinformation and ethics. While their research displays advanced NLP techniques, it primarily focuses on methodology and offers minimal discussion regarding the evolution of sentiments over time.

Lian et al. (2024) used 56,769 Weibo comments to examine Chinese public attitudes towards ChatGPT, reporting most of the negative sentiment, particularly around job loss, misinformation, and dependency. By contrast, Koonchanok, Pan and Jang (2024) found that Twitter discussions were more neutral or positive overall, with topic salience varying significantly by occupation; educators focused on teaching and learning, while IT professionals highlighted cybersecurity and technical issues. Temporal studies further show that perceptions are not static. Liu and Lyu (2024) documented a shift from early enthusiasm to growing scepticism in Chinese social media, as ethical concerns and notions of technological overreach became more prominent. Demirel, Kahraman-Gokalp and Gündüz (2024) observed a similar pattern on Twitter: initial optimism was followed by increased anxiety about privacy, cybersecurity and misuse, with regional patterns shaped by differing levels of technological adoption.

Building on 2023–2024 work on temporal and occupational variation, recent studies examine public and academic discourse at scale. Williams and Burnap (2025) analyse early Twitter reactions and show that initial enthusiasm about ChatGPT gives way to more ambivalent emotions as hopes for productivity coexist with fears about job loss and misinformation. Chomiak-Orsa (2025) likewise find mostly positive but increasingly cautious sentiment, though their Twitter focus risks under-representing other publics. Beyond social media, Tao and Shen (2025) show that social science articles are broadly optimistic yet preoccupied with integrity and misinformation. Taken together with Miyazaki et al. (2024), these studies depict perceptions of ChatGPT as heterogeneous, unsettled, and sensitive to context and platform.

2.3. Applications of ChatGPT in Education and Healthcare

Domain-specific reviews in healthcare and higher education provide an important backdrop for analysing public sentiment about GenAI. In healthcare, Iqbal et al. (2025) synthesize 17 systematic reviews and meta-analyses on ChatGPT and related LLMs, concluding that these systems show promise as low-cost, always-available assistants for diagnosis support, triage, documentation, and patient education, but that the overall quality of evidence is uneven and risks around hallucinations, bias, and liability remain substantial (Iqbal et al., 2025).

2.3.1 Education

In higher education, multiple 2025 systematic reviews now offer a more comprehensive overview of the integration of ChatGPT in teaching and learning. Dos (2025), Abdallah et al. (2025), and Munaye et al. (2025) collectively demonstrate that many published studies characterize ChatGPT as a writing assistant, feedback mechanism, or conversational tutor. The reported advantages include enhanced perceived productivity, diminished anxiety regarding academic writing, and increased engagement in large classes. Critics often raise concerns about copying, lack of critical thinking, fabricated content, and unequal access. They advocate for clear communication about assessment and honesty in schools. Jin and Sercu's (2025) study focuses on the effects of a staged approach, showing that using ChatGPT with teacher support can improve quick performance and satisfaction. Nonetheless, these tools raise questions about academic integrity, ethical behaviour, and assessment fairness. Ali et al. (2024), focusing on dental education, show how ChatGPT can correctly answer knowledge-based questions but caution that unsupervised use may facilitate plagiarism and undermine skill development. The educational literature, therefore, portrays GenAI as a double-edged sword, requiring pedagogical redesign and stronger policies rather than simple adoption.

2.3.2 Healthcare

Healthcare uses GenAI and ChatGPT to enhance patient communication and decision-making. According to Sallam (2023), their application potential in health professions education, research, and practice is enormous. This can lead to benefits such as personalized learning, patient education, and clinical reasoning support. Zhang and Kamel Boulos (2023) similarly argue that GenAI will play an increasingly important role in medicine as regulatory frameworks mature. Madsen and Toston (2025) similarly note, in a broader narrative review of health, education, and the economy, that healthcare use cases are dominated by low-risk advisory and documentation tasks rather than fully automated decision-making. Complementing these syntheses, Baxter et al. (2025) analyse social media and academic texts about ChatGPT in health care and reveal a mixed sentiment pattern: clinicians and the public simultaneously emphasize efficiency gains and worries about safety, accountability, and the dehumanization of care.

However, both studies emphasize deep ethical and practical concerns. Key risks include misinformation, biased outputs, privacy violations, and reduced human empathy. Ainapure et al. (2023) demonstrate the value of SA for monitoring public attitudes to vaccination and health policies but also show that detecting implicit sentiment and sarcasm remains challenging in health-related discourse. Lee et al. (2024) further suggest that virtual education and AI tools can help meet resource constraints in healthcare training, while cautioning that insufficient preparation and infrastructure can blunt their effectiveness. Overall, the healthcare literature recognizes substantial potential for GenAI but insists on strong governance, careful integration and ongoing monitoring of public trust.

2.4 Identified Research Gaps

Across these strands, several gaps emerge. First, although transformer-based models such as BERT are widely recognized as state-of-the-art, they remain underutilized in empirical work on social media sentiment about ChatGPT, with many studies still relying on lexicon-based or classical ML approaches (Rodríguez-Ibánez et al., 2023; Tan, Lee and Lim, 2023). Second, much research operates on relatively small or geographically restricted datasets, limiting generalizability and failing to capture global, cross-platform dynamics (Lian et al., 2024; Demirel et al., 2024). Third, temporal analysis is often limited to short windows, leaving longer-term shifts in sentiment underexplored (Liu and Lyu, 2024).

Fourth, work in education and healthcare typically uses surveys, interviews or case studies to gauge stakeholder views (Chan and Hu, 2023; Ali et al., 2024; Sallam, 2023), while social media studies treat "ChatGPT" as a single aggregate topic. There is little integration of large-scale SA with topic modelling that focuses specifically on sector-based benefits and risks. Finally, many existing studies prioritize overall accuracy in model evaluation, paying less attention to precision, recall, F1-score, and ROC-AUC, or to the specific performance on negative or minority sentiment classes (Alslaity and Orji, 2024; Krishna et al., 2023).

This study differs from existing work in several important aspects. First, it conducts a large-scale, longitudinal analysis of public sentiment toward ChatGPT and GenAI over an extended fourteen-month period, capturing how perceptions evolve beyond early post-release reactions. Second, it systematically compares traditional ML, DL, and transformer-based models within a unified evaluation framework, enabling robust performance benchmarking in noisy social media data. Third, it combines SA with sector-specific topic modelling to examine perceived benefits and risks in education and healthcare, moving beyond aggregate sentiment toward domain-sensitive interpretation.

3.0. Methodology

3.1. Research Design

This study employs a quantitative, computational research design based on large-scale social media data. The overall approach is comparative and exploratory. Given the complexity and often undefined nature of public opinions on emerging AI tools, a structured analytical approach was necessary. To support this, the study employs the CRISP-DM methodology, which provides a systematic framework for data mining that begins with understanding the research context and proceeds through data preparation, modelling, evaluation, and deployment, with iterative refinement at each stage (Shimaoka, Ferreira and Goldman, 2024). The wide acceptance, flexibility, and robustness of the CRISP-DM framework in data science make it well suited for research environments that require adaptability and ongoing updates as new insights emerge. Initially, we trained and assessed various sentiment classification models—encompassing traditional ML, DL, and transformer-based architectures—on unlabelled social media posts pertaining to ChatGPT and GenAI. The best-performing model is then used to classify sentiments in a much larger unlabelled dataset. Finally, temporal analysis and topic modelling techniques are applied to explore how sentiment changes over time and how it relates to key themes in education and healthcare.

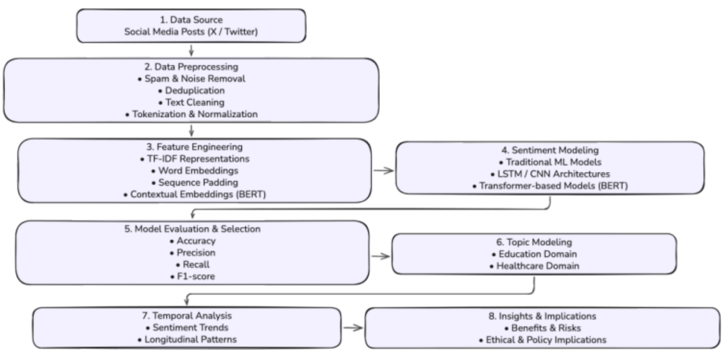

Figure 1- Overview of GenAI sentiment analysis pipeline.

Figure 1 presents an overview of the analytical pipeline adopted in this study, illustrating the sequence of data preprocessing, feature engineering, sentiment modelling, model evaluation, and downstream temporal and topic-based analyses.

3.2. Data Collection

Data for this study were collected from X (formerly Twitter), which was selected due to its widespread use for public discussion of emerging technologies and its suitability for large-scale SA. Posts related to ChatGPT and GenAI were retrieved using the platform's official API, ensuring compliance with data access policies and ethical research standards. Data collection queries were executed in accordance with API rate-limiting constraints. Automated delays and request scheduling were implemented to prevent throttling and ensure stable data retrieval throughout the collection period. Only publicly available posts were collected; no private, protected, or deleted content was accessed at any stage of the process. The final dataset consists of English-language posts published between November 2022 and December 2023, providing a longitudinal perspective on public sentiment toward ChatGPT and GenAI. This time span enables the analysis of evolving public attitudes beyond initial post-release reactions and supports robust temporal and sector-specific investigation.

3.3. Sentiment Annotation

A labelled dataset was constructed to train and evaluate the sentiment classification models. Posts in the sample were annotated into three categories: positive, negative, and neutral. The positive class captured expressions of approval, enthusiasm, or perceived benefits; the negative class included concerns, fears, frustration, or explicit criticism; and the neutral class covered factual descriptions or mixed views without a clear affective stance. Annotation followed a set of guidelines developed to ensure consistency. Ambiguous posts were discussed and resolved through agreement rules, and a portion of the data was double coded to confirm reliability.

3.4. Data Preprocessing

Prior to model training, the dataset underwent a series of preprocessing and filtering steps to improve data quality and reliability. Duplicate posts were identified and removed using exact text matching to prevent repeated content from biasing sentiment distributions. Spam-like content, including repeated promotional messages and posts with excessive hashtag usage, was excluded from reducing noise in the corpus. Additionally, very short posts with limited semantic content were removed, as they provide insufficient contextual information for reliable sentiment classification. Standard text normalization steps, including lowercasing and the removal of URLs, user mentions, and non-alphanumeric characters, were applied prior to analysis. The same preprocessing pipeline was applied consistently across all models to ensure fair and comparable evaluation. These preprocessing steps ensure a cleaner and more representative dataset while preserving the diversity of public discourse necessary for robust SA of ChatGPT and GenAI.

3.5. Modelling Approach

The modelling strategy used in this study was intended to assess the three main categories of sentiment classifiers. The first group included traditional machine-learning models such as logistic regression, Naïve Bayes, SVM, Random Forest, Decision Tree, and K-Nearest Neighbours. We trained these models using TF-IDF representations of the text. To ensure that the comparison was fair and consistent, we used grid search or a similar optimization method on the validation set. The third group consisted of deep-learning (DL) models like Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN). For these models, posts were converted into word indices and displayed using dense embeddings. The architecture decisions, such as the number of layers, hidden units, dropout rates, and learning rates to use, were based on preliminary experiments and standard SA research practices.

The final model category included a transformer-based approach using a BERT classifier. A pre-trained BERT model was fine-tuned on the labelled dataset, where the [CLS] token representation served as input to a classification layer responsible for predicting sentiment categories. Fine-tuning used mini-batch training alongside a suitable learning rate schedule and early stopping to minimize overfitting and improve generalization. All models were trained and tested on the same labelled dataset, which was divided into training, validation, and test subsets. This consistent evaluation framework ensured that any performance differences observed were the result of modelling choices rather than variations in the underlying data.

3.6. Evaluation Metrics

In order to assess and compare the performance of various models, several metrics were used. Accuracy served as a broad measure of the frequency with which each model made a correct sentiment classification. Nevertheless, since accuracy by itself can conceal that the models perform unevenly across classes, supplementary metrics—precision, recall, and F1-score were also computed for each sentiment category. These indicators gave a more comprehensive picture of the models' ability to detect correct positive, negative, and neutral posts as well as their capability to reduce the incorrect classification to the minimum. Macro-averaged metrics were also applied to ensure that each class was given equal weight. This was important because the dataset was not perfectly balanced, and relying solely on overall accuracy could obscure weaker performance on minority classes. In cases where more profound insight was needed, confusion matrices were examined to identify recurring misclassification patterns. This was especially relevant for distinguishing between negative and neutral posts, which are often challenging to separate due to subtle or implicit sentiment. Beyond quantitative metrics, execution time and model complexity were considered qualitatively to assess the practicality of deploying each model at scale. Taking all factors into account, the model that demonstrated the strongest and most balanced performance, especially in accurately identifying negative sentiment was selected as the final classifier for labelling the full unlabelled dataset.

3.7 Temporal Analysis of Sentiment

After classifying sentiment on the entire dataset using the most effective model, researchers conducted a temporal analysis to determine how public sentiment changed over time. Posts were grouped by period (by month), and the share of positive, negative, and neutral posts was determined for each time interval. As a result, it was possible to identify sentiment trends, peaks, and changes throughout the study period. Significant changes, if any, were also matched with the occurrences outside, e.g., a major product announcement, policy debates, or an incident involving GenAI, which was talked about extensively. The analysis examined whether the changes in sentiment were more positive, more negative, or relatively stable over time.

3.8 Topic Modelling for Education and Healthcare

To investigate domain-specific perceptions, posts about education and healthcare were selected from the dataset using additional keyword filters. Healthcare posts discussed doctors, patients, hospitals, diagnoses, treatment, and similar terms, while education posts mentioned teaching, learning, students, exams, universities, and schools. Topic modelling was then performed independently on these subsets to identify the most discussed topics. A probabilistic topic model, such as Latent Dirichlet Allocation (LDA), identified sets of words related to posts. Most representative terms and example posts were used to label themes. Within each domain, the topics were analysed in conjunction with sentiment data to ascertain the positive or negative framing of themes such as personalized learning, academic integrity, patient support, privacy, and misinformation. Employing sentiment classification alongside topic modelling provided a more nuanced understanding of discussions surrounding ChatGPT within the education and healthcare sectors.

4.0. Results

This section presents the empirical results of the study in four parts. First, it reports the comparative performance of the sentiment classification models and the selection of the final classifier. Second, it summarizes the distribution of public sentiment towards ChatGPT and GenAI across the full dataset. Third, it examines temporal changes in sentiment between November 2022 and December 2023. Finally, it presents topic-modelling results for education and healthcare, highlighting key benefits and risks identified in public discourse.

4.1. Model Performance and Selection

The first stage of analysis compared traditional ML models, DL architectures, and a transformer-based model using the manually labelled subset of social media posts. Models were evaluated on a held-out test set using accuracy, precision, recall, and F1-score.

Table 1- Model comparison results

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	0.79	0.80	0.79	0.78
SVM	0.93	0.93	0.93	0.93
Decision Tree	0.77	0.76	0.77	0.76
Logistic Regression	0.91	0.91	0.91	0.91
KNN	0.48	0.73	0.48	0.40
Naïve Bayes	0.77	0.81	0.77	0.75
LSTM	0.97	0.98	0.99	0.98
BERT	0.98	0.99	0.99	0.98
CNN	0.96	0.96	0.97	0.96

Across all metrics, the BERT-based classifier outperformed the other models. It achieved an accuracy of approximately 98%, and an F1-score of 0.98, indicating excellent discrimination between positive, negative, and neutral classes. Traditional models such as Logistic Regression, SVM and Random Forest performed reasonably well but showed lower F1-scores, particularly for the negative class, and were less robust when distinguishing negative from neutral posts. DL models (LSTM and CNN) improved some of these shortcomings but still fell short of the transformer-based approach.

Although the overall performance metrics of the LSTM and BERT models are numerically close, BERT consistently demonstrates slightly higher and more balanced results across accuracy, precision, recall, and F1-score. This pattern suggests a practical advantage rather than a purely marginal difference. In contrast to sequence-based LSTM architectures, BERT's bidirectional transformer design enables more effective modeling of contextual and semantic relationships within text, which is particularly important for SA of complex and informal social media discourse. Prior research has similarly shown that transformer-based models tend to be more robust to linguistic variability, sarcasm, and rapidly evolving terminology than recurrent neural networks. Accordingly, BERT was selected as the final model for downstream analysis based on its contextual modeling capability, consistency across evaluation metrics, and suitability for large-scale SA of GenAI–related discussions, rather than on claims of statistical superiority.

4.2. Sentiment Distribution in the Full Dataset

The selected BERT model was then applied to the complete dataset of over one million social media posts referring to ChatGPT and GenAI. Each post was assigned a sentiment label—positive, negative, or neutral—using the final fine-tuned classifier.

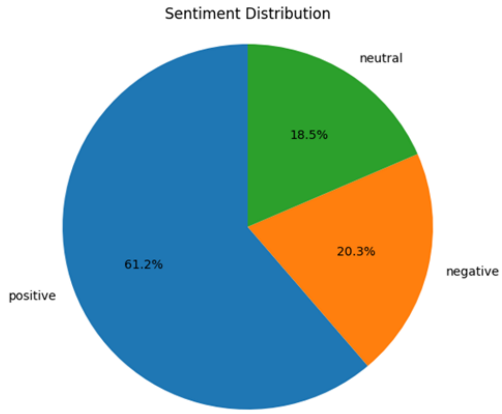

Figure 2- Distribution of sentiment classes.

The classification results indicate that public sentiment was predominantly positive: Positive: 61.2%, Negative: 20.3%, and Neutral: 18.5%. Thus, almost two-thirds of posts expressed favorable views or optimism regarding ChatGPT and GenAI, while roughly one-fifth conveyed concerns, criticism, or negative experiences. The remaining posts contained neutral, descriptive, or ambivalent content. These proportions suggest that, within the period studied, public discourse leaned clearly towards positive sentiment, even though a substantial minority of posts articulated skepticism or worry.

To complement the quantitative sentiment distribution, word cloud visualizations were generated for positive, neutral, and negative sentiment categories (Figure 3). The word clouds provide an illustrative overview of frequently occurring terms within each sentiment class, highlighting common themes discussed by users. While positive sentiment is characterized by terms related to usefulness, learning, and productivity, negative sentiment emphasizes concerns surrounding accuracy, reliability, and trust. Neutral sentiment primarily reflects informational and exploratory discussions. These visualizations are intended as descriptive aids and are used to support, rather than replace, statistical SA.

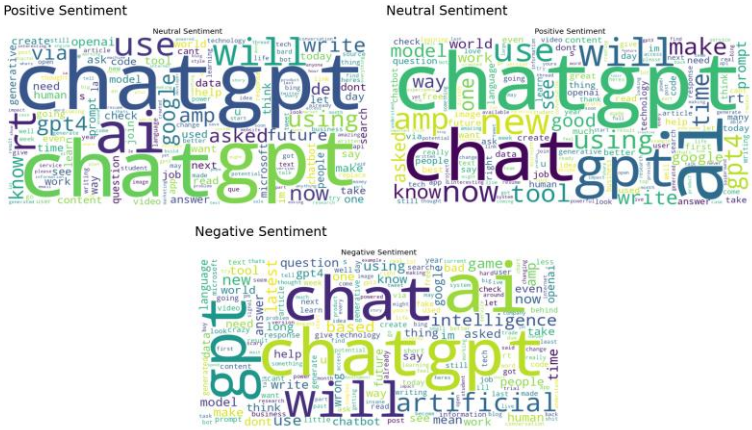

Figure 3- Word cloud visualizations for positive, neutral, and negative sentiment categories.

4.3 Temporal Dynamics of Sentiment

To investigate how sentiment evolved over time, sentiment labels were aggregated by month from November 2022 to December 2023. For each month, the proportion of positive, negative, and neutral posts was calculated, and a sentiment score was derived to provide a continuous measure of overall attitude.

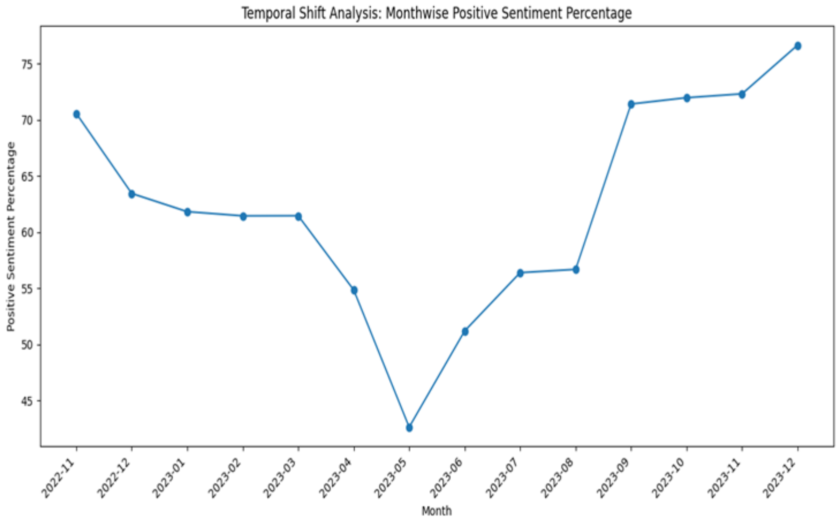

Figure 4- Monthly sentiment trends for proportion of positive posts.

The monthly trend reveals three broad phases:

The sentiment towards ChatGPT and GenAI evolved significantly over the course of 2023. In the initial phase, from late 2022 to early 2023, positive sentiment was high, with a dominant share of posts reflecting excitement, curiosity, and enthusiasm following the release of ChatGPT. However, as the year progressed, the number of positive posts gradually declined, reaching a low point around mid-2023. This decline coincided with increased scrutiny of GenAI, growing concerns about misinformation, and heightened ethical debates. In the second half of 2023, sentiment began to recover steadily. By the end of the year, the proportion of positive posts once again approached the levels seen early on, suggesting a renewed, though more measured, optimism. Contributing factors to this shift likely included system updates, wider adoption, regulatory developments, and more visible real-world applications of GenAI.

A statistical comparison of the first three months and last three months of the study period was conducted using a t-test on monthly sentiment scores (Table 2).

Table 2- t-test results comparing early and late-period sentiment scores.

Comparison	t-statistic	p-value	Interpretation
Early vs late-period sentiment scores	−4.52	0.011	Significant difference; sentiment higher in the late period than early period

The analysis showed a significant difference between the two periods, with sentiment in the final months being higher on average. The t-test produced a statistic of -4.52 and a p-value of 0.011, indicating that the improvement in sentiment over time was unlikely to be due to random variation.

4.4. Topic Modelling: Education

To examine domain-specific perceptions, posts related to education were extracted and analysed using topic modelling. The Latent Dirichlet Allocation (LDA) algorithm identified several recurrent topics, which were interpreted based on their most representative keywords and exemplar posts. Across the education subset, topics clustered around two main dimensions: perceived benefits and perceived risks.

4.4.1 Perceived benefits

Posts on social media highlighted several key benefits of ChatGPT in education. Many users described it as a valuable tool for personalized learning and student support, noting its ability to tailor explanations, assist with practice questions, and encourage independent learning. Additionally, ChatGPT was recognized for enhancing accessibility and inclusivity, offering around-the-clock guidance and helping non-native speakers, as well as learners who may not have access to traditional tutoring. Users also pointed out its role in supporting teachers and educational administration, with applications such as drafting materials, generating lesson ideas, and reducing routine administrative tasks. Together, these discussions frame ChatGPT as a tool that enables more flexible, responsive, and accessible educational experiences.

4.4.2 Perceived risks

Many social media posts raised concerns about the risks associated with ChatGPT in education. A significant worry was the potential for students to misuse the tool for plagiarism, generating assignments or exam answers that undermine the fairness of assessments. There were also fears that heavy reliance on AI-generated responses could reduce critical thinking and independent problem-solving, as students might become overly dependent on the tool. Additionally, concerns about the quality, ethics, and data privacy of AI-generated content were prevalent. Users questioned the reliability of the content produced by ChatGPT, voiced discomfort over the handling of their personal data, and called for clearer guidelines and safeguards to address these issues.

Table 3- Summary of key education-related topics and associated sentiment.

Topic	Type	Brief description	Associated sentiment
Personalized learning and student support	Benefit	ChatGPT used to tailor explanations, help with practice questions and support independent learning.	Mainly positive / optimistic
Improved accessibility and inclusivity	Benefit	AI seen as providing around-the-clock guidance and help for learners who lack access to traditional support.	Positive, with equity focus
Support for teachers and administration	Benefit	Used to draft materials, generate ideas for lessons and reduce routine administrative workload.	Positive / efficiency-oriented
Plagiarism and academic *integrity*	Risk	Concern that students may misuse ChatGPT to generate assignments or exam answers.	Strongly negative
Reduced critical thinking and over-reliance	Risk	Fear that heavy dependence on AI responses may discourage independent reasoning and problem-solving.	Negative / cautionary
Quality, ethics and data privacy	Risk	Doubts about reliability of AI content, ethical use of data and the need for clear safeguards.	Negative, with ethical concern

4.5 Topic Modelling: Healthcare

A similar procedure was applied to posts relating to healthcare, using keywords associated with medical practice, patients, and health systems. Topic modelling again revealed a mixture of perceived benefits and concerns.

4.5.1 Perceived benefits

Many discussions on social media highlighted the potential of ChatGPT to improve various aspects of healthcare. One key benefit mentioned was its ability to explain complex medical concepts in simple, understandable language, making it easier for patients to grasp important health information. In addition, users saw the potential for ChatGPT to assist with routine tasks like drafting documentation, summarizing medical information, and supporting clinical decision-making, which could improve efficiency in everyday healthcare settings. The tool was also seen as valuable for managing large amounts of data and potentially aiding in remote patient monitoring and telehealth services.

4.5.2 Perceived risks

While many users acknowledged the potential benefits of ChatGPT in healthcare, there were significant concerns about the risks involved. Overall, while users recognized the advantages, they emphasized the need for robust safeguards to address these serious concerns.

Table 4- Summary of key healthcare-related topics and associated sentiment.

Topic	Type	Brief description	Associated sentiment
Enhanced patient information and support	Benefit	ChatGPT used to explain medical concepts in understandable language and answer basic health questions.	Mainly positive /supportive
Efficiency and decision support	Benefit	Assists with drafting documentation, summarizing information and supporting routine clinical decision-making.	Positive / efficiency-oriented
Data management and remote care	Benefit	Discussed as improving handling of large data volumes and supporting remote monitoring and telehealth workflows.	Positive, with cautious optimism
Data privacy and security	Risk	Concerns about confidentiality of sensitive health data and potential misuse or leakage of information.	Strongly negative / high concern
Misinformation and unsafe advice	Risk	Fear that inaccurate or misleading medical guidance might be followed without professional oversight.	Negative / safety-focused
Bias and ethical dilemmas	Risk	Worries about biased diagnostic suggestions and broader ethical issues in delegating aspects of care to AI.	Negative, ethically framed

5.0. Discussion

5.1 Interpretation of Key Findings

5.1.1 Model performance and methodological contribution

The paper found that BERT performed exceptionally well, with an accuracy of 98% and an F1-score of 0.98, outperforming traditional and DL models. The result confirms the assertion in the preceding paper that one of the most promising applications of transformer-based architectures lies in SA within intricate, contextually rich environments, such as social media (Su and Kabala, 2023; Rodríguez-Ibánez et al., 2023). In comparison, classical ML models and simpler DL architecture struggled more with distinguishing negative from neutral posts, especially where sentiment was implicit or mixed. This has two implications. First, it suggests that studies of public perceptions of GenAI that rely solely on lexicon-based or basic ML methods may under-detect subtle negativity or ambivalence. Second, it demonstrates that the additional computational cost of transformer-based models can be justified in applications where nuance and misclassification of minority classes are consequential.

5.1.2 Overall sentiment and its dynamics

The second major finding is that sentiment towards ChatGPT and GenAI was largely positive—about 61.2% of posts were classified as positive, compared with 20.3% negative and 18.5% neutral. This aligns with survey-based evidence that users recognize substantial benefits of GenAI, including convenience, productivity, and support for learning or work tasks (Chan and Hu, 2023). At the same time, the presence of a sizeable negative minority indicates that enthusiasm coexists with scepticism and concern. Temporal analysis indicated a decrease in positive sentiment from late 2022 to mid-2023, followed by a recovery in the latter half of 2023. This pattern mirrors earlier observations that initial excitement around ChatGPT gave way to heightened ethical debate and anxiety about misinformation, then stabilized as regulatory discussions advanced and real-world uses became more familiar (Liu and Lyu, 2024; Demirel et al., 2024). The t-test evidence that sentiment in the past three months of the period was significantly higher than in the first three months suggests that, over time, public opinion became more favourable or at least more confident. Taking together, these results indicate that public reactions to GenAI should not be treated as fixed. Instead, sentiment appears to respond to external events, media narratives, and policy signals, which supports the case for continuous monitoring rather than one-off measurements.

5.1.3 Sector-specific perceptions: education and healthcare

The topic-modelling analysis for education and healthcare revealed a consistent pattern of "dual-edged" perceptions. In education, the benefits identified included personalized learning, improved accessibility, support for teachers and students, and the potential to "revolutionize" aspects of teaching and assessment. Risks were primarily associated with plagiarism, academic integrity, loss of critical thinking skills, data privacy, and over-dependence on AI tools. Such results are in line with the research that has been done earlier, which indicates that students value GenAI as a learning assistant, but at the same time, they worry about the ethical aspects and the diminishing of human skills (Chan and Hu, 2023). Healthcare highlighted patient care, precision medicine, data management, and workflow efficiency. Many posts expressed concerns about data security, misinformation, and the ethical issues of using AI to make diagnostic and treatment decisions. ChatGPT has clinical potential but significant risks in health and medical education, according to earlier reviews (Sallam, 2023; Zhang and Kamel Boulos, 2023). According to sector-specific analysis, public discourse is not simply pro- or anti-AI. When used carefully and with clear safeguards, ChatGPT is valuable, especially in contexts with high error consequences.

5.1.4 Ethics and employment

There was a lot of talk about ethics in almost all the areas that were talked about. These problems were brought up most often and included bias, privacy breaches, and false information. When this was talked about, the healthcare sector got the most attention because bad advice can have very bad results there. The mentioned trend fits with the general debate that says ethical rules for AI shouldn't come after GenAI's release if people are to keep trusting it (Naing and Udomwong, 2024; Lee et al., 2024). Job displacement was another strong concern. Both the SA and topic modelling pointed to clear negativity around the idea that GenAI might replace human workers, especially in creative and knowledge-based roles. This mirrors the findings of Miyazaki et al. (2024), who report deep unease among illustrators and other creative professionals about the potential exploitation and devaluation of their work. Together, these concerns help explain the negative minority sentiment and indicate that worries about labour-market impacts are a major source of resistance to GenAI.

5.1.5 Theoretical Contributions

Beyond methodological comparison, this study contributes conceptually to understanding public perceptions of GenAI by framing sentiment as a dynamic and context-dependent phenomenon. The longitudinal analysis demonstrates that public attitudes toward ChatGPT are not static but evolve in response to increased familiarity, media discourse, and emerging ethical concerns. Furthermore, the sector-specific findings show that optimism and risk perception coexist and vary systematically across domains, with education and healthcare exhibiting distinct patterns of trust, utility, and concern. These insights extend existing research on technology acceptance by highlighting the importance of temporal dynamics and institutional context in shaping public sentiment toward GenAI systems. Rather than treating public perception as a single aggregate measure, the study underscores the need to account for domain-specific expectations and ethical sensitivities when evaluating societal responses to AI technologies.

5.2. Theoretical Implications

Theoretically, this study contributes to research on both SA and public perceptions of GenAI. First, it shows that transformer-based models are better equipped than traditional sentiment classifiers to handle large-scale, noisy, and context-dependent data, particularly when the language is fast-changing and emotionally charged. This lends weight to the view that BERT and similar architectures should not be seen as optional enhancements but as core tools when analysing high-stakes and contested technologies. Second, by combining sentiment classification with temporal analysis and topic modelling, the study offers a framework for tracking how attitudes toward emerging technologies develop over time and across different domains. Rather than treating sentiment as a static snapshot, this approach encourages future work to view it as a process that unfolds alongside regulatory change, media narratives, and peoples lived experiences. Third, by looking at both education and healthcare, the study provides cross-sector insights into how GenAI is positioned. It shows that ethical concerns such as privacy, misinformation, and bias appear in multiple settings, but that their prominence and framing depend on local stakes and professional norms. This points towards more nuanced theories of "AI acceptance" that take sector-specific logics seriously, instead of assuming a single, uniform adoption pattern.

5.3. Practical Implications

Practically speaking, the results can be interpreted as a signal to people working in different fields like educators, healthcare professionals, technology developers, and media organizations. With regard to the educational sector, the findings lead to the conclusion that ChatGPT could be a means of achieving more adaptive and personalized learning; however, this is the case only when institutions have established robust frameworks for maintaining academic integrity and fostering a critical attitude towards AI-generated content. To accomplish this, it might be necessary to rethink assessment design, teach AI literacy more explicitly, and set up clear rules about the time and way generative tools can be utilized.

In healthcare, GenAI applications need to be introduced cautiously and in clearly defined support roles, rather than as substitutes for professional judgment. The study underlines the importance of robust data privacy safeguards, systematic auditing of AI outputs, and careful integration into clinical workflows so that tools such as ChatGPT assist, rather than override, clinicians' decisions. For technology developers, the findings emphasize that they must address bias through more diverse training data, improve the transparency of model behaviour, and maintain ongoing monitoring and feedback mechanisms to detect drift or misuse. Public sentiment data can play a useful role here by flagging emerging concerns before they escalate.

Finally, for media and communication practitioners, the prominence of misinformation-related topics suggests that GenAI should be treated both as an object of coverage and as a potential tool in countering false information. How the public feels about ChatGPT will likely be shaped by clear, responsible communication about its capabilities.

5.4. Policy Implications

The results emphasize the necessity of detailed governance frameworks for GenAI, especially in areas like education and healthcare, which are emotionally sensitive. The prevalence of concerns about privacy, bias, and misinformation highlights the need for regulation that addresses various aspects. Firstly, ethical governance is indispensable, with explicit provisions regarding data collection, consent, storage, and usage aimed at ensuring responsible practices. Moreover, transparency is equally important, implying that models should clarify how they were trained, the data they used, and their limitations.

On top of that, there should be implemented accountability measures, for example, entities entrusted with independent oversight, which would be responsible for not only supervising the GenAI deployment but also investigating any detrimental effects resulting from it. Furthermore, engaging the public in these matters is crucial, as it ensures that people's concerns and expectations are at the forefront of the regulatory decision-making process, rather than being a mere afterthought. Such policies, which are perceived as being responsive and inclusive, have a higher probability of strengthening and sustaining the positive feeling, noticed towards the end of the studyperiod.

5.5. Ethical Considerations

The study itself raises several ethical questions. Working with public social media data still demands careful attention to privacy, even when posts are technically open and accessible. Anonymizing and aggregating the data mitigates certain risks; however, continuous efforts are required to establish responsible utilization of public data in research. More broadly, the ethical issues that appear in public discourse include bias, data privacy, misinformation, and uneven impacts on labour markets—show that SA should not be used simply as a way of measuring acceptance level. Instead, it ought to feed into critical reflection on whether, and under what conditions, GenAI is aligned with social values, and on who is likely to bear the costs and who stands to gain from its deployment.

5.6. Key Contributions

This study makes several key contributions. (1) It provides one of the largest longitudinal sentiment analyses of ChatGPT and GenAI, examining over one million social media posts collected across a fourteen-month period. (2) It demonstrates the superior and statistically robust performance of transformer-based models compared with traditional ML and DL approaches for SA of GenAI-related discourse. (3) It integrates sentiment classification, temporal trend analysis, and sector-specific topic modelling into a unified analytical framework for studying public perceptions of emerging technologies. (4) It reveals domain-specific benefits and risks of GenAI in education and healthcare, highlighting ethical, social, and institutional concerns reflected in public discourse. (5) Finally, it offers methodological and theoretical guidance for future research on public perceptions of GenAI and other rapidly evolving digital technologies.

6.0. Conclusion, Limitations and Future Work

The findings indicate that the BERT-based classifier was significantly more effective than both the traditional and DL models, especially in differentiating subtle differences between negative and neutral sentiment. In general, the public sentiment towards ChatGPT and GenAI was positive most of the time throughout the study period, however, a significant minority of posts expressing concern or scepticism. Sentiment also evolved over time: it was lower in the first half of 2023 and then picked up in the second half, with statistical tests showing an overall recovery at the end of the year. Topic modelling uncovered a set of advantages like personalized learning, improved access to information, and time savings and a set of risks, such as plagiarism, data privacy breaches, misinformation, bias, and loss ofjobs.

6.1. Limitations

Several limitations should be acknowledged. The analysis is based on publicly available social media posts and therefore reflects the views of people who are active on those platforms. It may underrepresent groups who rarely use social media, who post in other languages, or who discuss GenAI in closed communities or offline settings. Although care was taken in cleaning and annotating the data, sentiment classification is still a probabilistic process. Misclassification will always be a part of the process, especially in cases of highly ambiguous or sarcastic posts, even if a strong model is used. Since keyword-based filters have been utilized for education and healthcare, it implies that some very relevant posts that use less typical language might not have been detected, and a small number of posts that are only loosely related may have beenincluded. Lastly, the study's scope is relatively narrow, spanning from November 2022 to December 2023. While this offers a more detailed temporal view than many early studies, it still does not capture longer-term shifts that may emerge as GenAI becomes more deeply embedded in everyday systems and as regulatory responses continue to develop.

6.2. Directions for Future Research

Future research could extend this framework in several directions to capture richer and more nuanced public perceptions of GenAI. First, multimodal SA incorporating textual, visual, and audiovisual content—such as images and videos—could be employed to better interpret sentiment expressed through memes, screenshots, and short-form media. Second, fine-grained emotion classification may complement polarity-based SA by distinguishing affective responses such as trust, fear, enthusiasm, or anxiety toward GenAI systems. Third, stance detection techniques could be integrated to identify whether users support, oppose, or remain neutral toward the adoption of GenAI, particularly in sensitive domains such as education and healthcare. Finally, extending the analysis across multiple social media platforms would enable comparative insights into how platform-specific affordances shape public discourse and engagement with GenAI technologies.

References

Abdallah, N., Katmah, R., Khalaf, K. & Jelinek, H.F. (2025) 'Systematic review of ChatGPT in higher education: Navigating impact on learning, wellbeing, and collaboration', *Social Sciences and Humanities Open*, 12, 101866. Available at: https://doi.org/10.1016/j.ssaho.2025.101866.

Ainapure, B.S. et al. (2023) 'Sentiment analysis of COVID-19 tweets using deep learning and lexicon-based approaches', *Sustainability*, 15, 2573. Available at: https://doi.org/10.3390/su15032573.

Ali, K., Barhom, N., Tamimi, F. & Duggal, M. (2024) 'ChatGPT—A double-edged sword for healthcare education? Implications for assessments of dental students', *European Journal of Dental Education*, 28, pp. 206–211. Available at: https://doi.org/10.1111/eje.12937.

Alslaity, A. & Orji, R. (2024) 'Machine learning techniques for emotion detection and sentiment analysis: Current state, challenges, and future directions', *Behaviour & Information Technology*, 43(1), pp. 139–164. Available at: https://doi.org/10.1080/0144929X.2022.2156387.

Alsemaree, O., Alam, A.S., Gill, S. & Uhlig, S. (2024) 'Sentiment analysis of Arabic social media texts: A machine learning approach to deciphering customer perceptions', *Heliyon*, 10(9), e27863. Available at: https://doi.org/10.1016/j.heliyon.2024.e27863.

Baxter, P., Li, M-H., Wei, J. & Koizumi, N. (2025) 'Public versus academic discourse on ChatGPT in health care: Mixed methods study', *JMIR Infodemiology*, 5(1). Available at: https://doi.org/10.2196/64509.

Bermudez-Sosa, H.J., Olarte-Henao, J. & Rojas-Berrío, S.P. (2025) 'Sentiment and emotion analysis from textual information: A systematic literature review', *Journal of Information Science*. Available at: https://doi.org/10.1177/01655515251353170.

Chan, C.K.Y. & Hu, W. (2023) 'Students' voices on generative AI: Perceptions, benefits, and challenges in higher education', *International Journal of Educational Technology in Higher Education*, 20(1), pp. 1–18. Available at: https://doi.org/10.1186/s41239-023-00411-8.

Chomiak-Orsa, I. (2025) 'Sentiment and emotion analysis of public discourse on ChatGPT using VADER sentiment analysis', *Journal of Digital Society*, 1(1). Available at: https://doi.org/10.63913/jds.v1i1.1.

Darraz, N. et al. (2024) 'Integrated sentiment analysis with BERT for enhanced hybrid recommendation systems', *Expert Systems with Applications*, 261, 125533. Available at: https://doi.org/10.1016/j.eswa.2024.125533

Demirel, S., Kahraman-Gokalp, E. & Gündüz, U. (2024) 'From optimism to concern: Unveiling sentiments and perceptions surrounding ChatGPT on Twitter', *International Journal of Human–Computer Interaction*. Available at: https://doi.org/10.1080/10447318.2024.2392964.

Diri, G., Obiorah, P. & Du, H. (2025) 'Comparative study of sentiment analysis techniques: Traditional machine learning vs deep learning approaches', in *InSITE 2025: Informing Science + IT Education Conferences*. Hiroshima: Informing Science Institute. Available at: https://doi.org/10.28945/5490.

Dos, I. (2025) 'A systematic review of research on ChatGPT in higher education', *The European Educational Researcher*, 8(2), pp. 59–76. Available at: https://doi.org/10.31757/euer.824.

Iqbal, U. et al. (2025) 'Impact of large language model (ChatGPT) in healthcare: An umbrella review and evidence synthesis', *Journal of Biomedical Science*, 32(1), 45. Available at: https://doi.org/10.1186/s12929-025-01131-z.

Jahanbin, K. & Ali, M. (2023) 'Aspect-based sentiment analysis of Twitter influencers to predict cryptocurrency trends using hybrid deep transfer learning models', *IEEE Access*, pp. 1–1. Available at: https://doi.org/10.1109/access.2023.3327060.

Jin, Y. & Sercu, L. (2025) 'ChatGPT interventions in higher education: A systematic review of experimental studies', *Journal of Computer Assisted Learning*, 41(4). Available at: https://doi.org/10.1111/jcal.70072.

Khan, N. et al. (2024) 'Global insights and the impact of generative AI-ChatGPT on multidisciplinary fields: A systematic review and bibliometric analysis', *Connection Science*, 36(1). Available at: https://doi.org/10.1080/09540091.2024.2353630.

Koonchanok, R., Pan, Y. & Jang, H. (2024) 'Public attitudes toward ChatGPT on Twitter: Sentiments, topics, and occupations', *Social Network Analysis and Mining*, 14(1). Available at: https://doi.org/10.1007/s13278-024-01260-7.

Krishna, K.M. et al. (2023) 'A CNN-LSTM-based hybrid deep learning approach for sentiment analysis on Monkeypox tweets', *New Generation Computing*. Available at: https://doi.org/10.1007/s00354-023-00227-0.

Lee, J., Kim, H. & Kron, F. (2024) 'Virtual education strategies in sustainable healthcare and medical education: A topic modelling analysis of four decades of research', *Medical Education*, 58, pp. 47–62. Available at: https://doi.org/10.1111/medu.15202.

Li, L. et al. (2022) 'Key factors in MOOC pedagogy based on NLP sentiment analysis of learner reviews', *Computers & Education*, 176, 104354. Available at: https://doi.org/10.1016/j.compedu.2021.104354.

Lian, Y. et al. (2024) 'Public attitudes and sentiments toward ChatGPT in China: A text mining analysis based on social media', *Technology in Society*, 76, 102442. Available at: https://doi.org/10.1016/j.techsoc.2023.102442.

Liu, Y. & Lyu, Z. (2024) 'Changes in public perception of ChatGPT: A text mining perspective based on social media', *International Journal of Human–Computer Interaction*. Available at: https://doi.org/10.1080/10447318.2024.2406966.

Madsen, D.Ø. & Toston, D.M. II (2025) 'ChatGPT and digital transformation: A narrative review of its role in health, education, and the economy', *Digital*, 5(3), 24. Available at: https://doi.org/10.3390/digital5030024.

Mao, Y., Liu, Q. & Zhang, Y. (2024) 'Sentiment analysis methods, applications, and challenges: A systematic literature review', *Journal of King Saud University – Computer and Information Sciences*, 36(4), 102048. Available at: https://doi.org/10.1016/j.jksuci.2024.102048.

Miyazaki, K. et al. (2024) 'Public perception of generative AI on Twitter: An empirical study based on occupation and usage', *EPJ Data Science*, 13(2). Available at: https://doi.org/10.1140/epjds/s13688-023-00445-y.

Munaye, Y.Y. et al. (2025) 'ChatGPT in education: A systematic review on opportunities, challenges, and future directions', *Algorithms*, 18(6), 352. Available at: https://doi.org/10.3390/a18060352.

Naing, S. & Udomwong, P. (2024) 'Public opinions on ChatGPT: An analysis of Reddit discussions using sentiment analysis, topic modelling, and SWOT analysis', *Data Intelligence*, pp. 1–50. Available at: https://doi.org/10.1162/dint_a_00250.

Qi, Y. & Shabrina, Z. (2023) 'Sentiment analysis using Twitter data: A comparative application of lexicon- and machine-learning-based approaches', *Social Network Analysis and Mining*, 13(1). Available at: https://doi.org/10.1007/s13278-023-01030-x.

Raees, M. & Fazilat, S. (2024) 'Lexicon-based sentiment analysis on text polarities with evaluation of classification models', *arXiv*. Available at: https://doi.org/10.48550/arxiv.2409.12840.

Ramezani, E.B. (2025) 'Sentiment analysis applications using deep learning advancements in social networks: A systematic review', *Neurocomputing*, 634, 129862. Available at: https://doi.org/10.1016/j.neucom.2025.129862.

Rodríguez-Ibánez, M. et al. (2023) 'A review on sentiment analysis from social media platforms', *Expert Systems with Applications*, 223, 119862. Available at: https://doi.org/10.1016/j.eswa.2023.119862.

Sallam, M. (2023) 'ChatGPT utility in healthcare education, research, and practice: Systematic review on promising perspectives and valid concerns', *Healthcare*, 11, 887. Available at: https://doi.org/10.3390/healthcare11060887.

Shaik, T. et al. (2023) 'Sentiment analysis and opinion mining on educational data: A survey', *Natural Language Processing Journal*, 2, 100003. Available at: https://doi.org/10.1016/j.nlp.2022.100003.

Shimaoka, A.M., Ferreira, R.C. & Goldman, A. (2024) 'The evolution of CRISP-DM for data science: Methods, processes and frameworks', *SBC Reviews on Computer Science*, 4(1), pp. 28–43. Available at: https://doi.org/10.13140/RG.2.2.22493.42721.

Su, Y. & Kabala, Z.J. (2023) 'Public perception of ChatGPT and transfer learning for tweets sentiment analysis using Wolfram Mathematica', *Data*, 8(12), 180. Available at: https://doi.org/10.3390/data8120180.

Tabany, M. & Gueffal, M. (2024) 'Sentiment analysis and fake Amazon reviews classification using SVM supervised machine learning model', *Journal of Advances in Information Technology*, 15(1), pp. 49–58. Available at: https://doi.org/10.12720/jait.15.1.49-58.

Tan, K.L., Lee, C.P. & Lim, K.M. (2023) 'A survey of sentiment analysis: Approaches, datasets, and future research', *Applied Sciences*, 13(7), 4550. Available at: https://doi.org/10.3390/app13074550.

Tao, Y. & Shen, Q. (2025) 'Academic discourse on ChatGPT in social sciences: A topic modelling and sentiment analysis of research article abstracts', *PLOS ONE*, 20(10). Available at: https://doi.org/10.1371/journal.pone.0334331.

Williams, L. & Burnap, P. (2025) 'The emotional landscape of technological innovation: A data-driven case study of ChatGPT's launch', *Informatics*, 12(3), 58. Available at: https://doi.org/10.3390/informatics12030058.

Wu, S. & Gao, Y. (2024) 'Machine learning approach to analyse the sentiment of airline passengers' tweets', *Transportation Research Record*, 2678(2), pp. 48–56. Available at: https://doi.org/10.1177/03611981231172948.

Zhang, P. & Kamel Boulos, M.N. (2023) 'Generative AI in medicine and healthcare: Promises, opportunities and challenges', *Future Internet*, 15(9), 286. Available at: https://doi.org/10.3390/fi15090286.