

Public Opinions on ChatGPT: An Analysis of Reddit Discussions by Using Sentiment Analysis, Topic Modeling, and SWOT Analysis

Shwe Zin Su Naing[†], Piyachat Udomwong

International College of Digital Innovation and Financial Technology 239 Nimmanhaemin Road, Suthep, Muang, Chiang Mai 50200, Thailand

Keywords: ChatGPT; Sentiment analysis; Topic modeling; SWOT analysis; Public opinion; Reddit

Citation: Naing, S.Z.S., Udomwong P.: Public Opinions on ChatGPT: An Analysis of Reddit Discussions by Using Sentiment Analysis, Topic Modeling, and SWOT Analysis. *Data Intelligence* 6(2), 344-374 (2024). doi: https://doi.org/10.1162/dint_a_00250
Submitted: October 31, 2023; Revised: December 13, 2023; Accepted: March 1, 2024

ABSTRACT

The sudden arrival of AI (Artificial Intelligence) into people's daily lives all around the world was marked by the introduction of ChatGPT, which was officially released on November 30, 2022. This AI invasion in our lives drew the attention of not only tech enthusiasts but also scholars from diverse fields, as its capacity extends across various fields. Consequently, numerous articles and journals have been discussing ChatGPT, making it a headline for several topics. However, it does not reflect most public opinion about the product. Therefore, this paper investigated the public's opinions on ChatGPT through topic modelling, Vader-based sentiment analysis and SWOT analysis. To gather data for this study, 202905 comments from the Reddit platform were collected between December 2022 and December 2023. The findings reveal that the Reddit community engaged in discussions related to ChatGPT, covering a range of topics including comparisons with traditional search engines, the impacts on software development, job market, and education industry, exploring ChatGPT's responses on entertainment and politics, the responses from Dan, the alter ego of ChatGPT, the ethical usage of user data as well as queries related to the AI-generated images. The sentiment analysis indicates that most people hold positive views towards this innovative technology across these several aspects. However, concerns also arise regarding the potential negative impacts associated with this product. The SWOT analysis of these results highlights both the strengths and pain points, market opportunities and threats associated with ChatGPT. This analysis also serves as a foundation for providing recommendations aimed at the product development and policy implementation in this paper.

[†] Corresponding author: Shwe Zin Su Naing (E-mail: sunaing1119@gmail.com; ORCID: 0009-0002-8202-5182).

1. INTRODUCTION

The idea of simulating human intelligence in machines is not entirely a new topic since there have long been Chinese and Greek myths surrounding robots and automation. The formal debut of modern AI can be traced back to a conference at Dartmouth College in 1956, which officially gave birth to the term “Artificial Intelligence” [1, 2]. The early presence of AI was already evident in different forms, such as navigation systems and recommendation algorithms. However, ChatGPT marked a significant turning point, available for any person with internet access, ushering in the Age of AI with its unique ability to engage in natural conversations, comprehend our queries, and deliver intuitive responses. On November 30, 2022, ChatGPT was formally debuted by OpenAI, the well-known AI research lab, co-founded by Elon Musk and Sam Altman [3]. Since this new AI model was made available to the public for free, it brings unprecedented excitement not only for the AI community but also for the entire world. Many people around the world are dazzled by its capability to engage in human-like conversation with a sense of intelligence, humour, creativity, and emotion. Within 5 days after its launch, it reached more than a million human users who are left with impressions made by ChatGPT’s capacities such as rattling out poems, short stories, essays and even sorting out programming errors and making scientific discussion. As of June 2023, ChatGPT has 1.7 billion visits and most of the users are from the United States [4]. It has been said that ChatGPT can function as a friend, guide, an assistant or even as a philosopher, resulting in the potential to replace humans [5].

With the rapid growth of ChatGPT users, it has become increasingly important to delve into user opinions regarding this product, which can be beneficial for both product and policy development. While previous studies in this realm have primarily focused on Twitter discussions, there leaves room for a more comprehensive exploration of Reddit discussions. Therefore, the analysis in this paper focuses on Reddit comments posted within the “r/ChatGPT” subreddit from December 2022 to December 2023. It explores the major topics being discussed and examines the sentiment polarity associated with each topic. This research also incorporates a SWOT analysis of these discussions to discern the strengths, weaknesses, opportunities, and threats associated with ChatGPT. Hence, this paper contributes to gain a deeper understanding of how people perceive the latest technology, ChatGPT, and anticipate its potential implications and concerns raised by reddit discussions. In this regard, this research not only provides valuable insights for market research and informs the development of AI products but also highlights the positive and negative impacts of AI on society. Understanding such public opinions enables the AI industry to make informed decisions regarding future development, while governments can formulate policies to address potential negative implications.

2. LITERATURE REVIEW

A literature review regarding ChatGPT combines both alarmist and enthusiastic perspectives sourced from both scholarly articles and news outlets. Recent literature on ChatGPT can be divided into these categories:

- Evaluating ChatGPT's Performance
- Distinguishing ChatGPT-Generated Work from Human-Generated Work: Methods and Insights
- Exploring the Impacts of ChatGPT
- User Opinion Analysis of ChatGPT on Social Media

2.1 Evaluating ChatGPT's Performance

The primary focus in this literature revolves around evaluating the capabilities and limitations of ChatGPT through extensive testing. Kung et al. [6] conducted an interesting study that presented new and surprising evidence of ChatGPT's ability to perform on the United States Medical Licensing Examinations (USMLE) with high accuracy. The results indicated that ChatGPT's performance approached or exceeded the passing threshold for USMLE. Indeed, this remarkable performance on USMLE suggests that AI-assisted human learners in medical education, as well as clinical practices and decision-making, may become more popular in the future. Moreover, it is interesting to investigate further whether ChatGPT can assist doctors in generating better clinical decisions, reducing the risk of errors associated with emotions-based clinical decisions or not. The study by Zhai, X. [7] reflects on the potential for AI applications to significantly impact academic literature by accelerating knowledge compilation and expression while reducing the reliance on human-driven procedures. In this study, an experiment was conducted using ChatGPT to write an academic paper titled "Artificial Intelligence for Education". The experiment results showed that ChatGPT could help researchers produce a coherent, partially accurate, informative, and systematic paper. It has been noted that the writing process was highly efficient, taking only 2-3 hours, and required minimal professional knowledge from the author. In addition, ChatGPT's performance as a translator was evaluated in this study [8], showing that ChatGPT competes well with commercial translation products like Google Translate for high-resource European languages but falls behind for low-resource or distant languages. However, it is found that the release of GPT-4 on March 15, 2023, enhanced the translation performance of ChatGPT, enabling it to achieve comparable results to commercial translation products, including for distant languages. Therefore, this study concluded that ChatGPT has shown its proficiency as a translator, particularly with the advancements brought about by the GPT-4 engine. Other evidence was also found by Qin, C. [9] to demonstrate ChatGPT's impressive performance across a diverse range of natural language processing (NLP) tasks. The findings reveal that ChatGPT performs well in reasoning and dialogue tasks due to its strong generalist capabilities. However, they also discovered that ChatGPT encounters difficulties in specific tasks like sequence tagging that involves identifying and classifying specific elements or entities within a sequence of text. Another empirical study [10] proved that ChatGPT surpasses other zero-shot LLMs and even outperforms fine-tuned models in various tasks. However, the paper also identified the challenges and failures faced by ChatGPT in specific cases such as its tendency to generate lengthy summaries and generating incorrect translations in machine translations.

2.2 Distinguishing ChatGPT-Generated Work from Human-Generated Work: Methods and Insights

A considerable body of literature also exists, examining methods and approaches for distinguishing between chatbot-generated work and human-generated work. Guo, B., et.al [11] conducted interesting research that sheds light on the implicit differences between humans and ChatGPT, offering valuable insights into the future directions of Language Models (LLMs). The findings reveal major distinctions between humans and ChatGPT:

- 1) ChatGPT's responses remain closely tied to the given question, whereas humans often digress and explore related topics.
- 2) ChatGPT provides objective answers, while humans tend to express subjective opinions.
- 3) ChatGPT produces safer, balanced, and neutral texts, whereas humans offer more specific information, drawing from diverse sources and including citations.
- 4) ChatGPT's responses maintain a formal tone, while humans employ colloquial language, abbreviations, slang, humor, and examples.
- 5) ChatGPT exhibits less emotional expression, relying on logical flow and conjunctions, whereas humans employ punctuation, grammar features, and brackets to convey emotions and explanations.

This study [12] discovered that the ChatGPT Detector is the most robust and effective method for detecting ChatGPT-generated answers, arguing that this is because the ChatGPT Detector is fine-tuned using a corpus that specifically includes ChatGPT-generated answers, enabling it to better recognize the unique patterns present in these responses. The paper by Mitchell, E. [13] introduces a novel method called DetectGPT to tackle the need for detecting machine-generated text. The results indicate that DetectGPT outperforms existing zero-shot methods in identifying model-generated samples. Another testing was made by Gao, C.A. et al. [14] to differentiate between human-generated and AI-generated written work by gathering ten research abstracts from medical journals and requesting ChatGPT to produce research abstracts corresponding to the titles and journals. To distinguish between the two types of work, they used an AI output detector, plagiarism detector, and human reviewers with a skeptical outlook. The results revealed that AI-generated output was not detected as plagiarism, but most of it was flagged by the AI output detector and some by the human reviewers. Consequently, the AI output detector is suggested to be used in the editorial process, and the acceptable percentage of AI usage and ethical boundaries are likely to be the subject of debate soon.

2.3 Exploring the Impacts of ChatGPT

There are many recent studies attempting to distinguish ChatGPT-generated work from human-generated work. Indeed, ChatGPT's capacity to generate work that is not easily distinguishable from human writing has raised concerns, particularly in the education industry where students are using it for their homework assignments. As a result, some academics have begun to reconsider their evaluation methods by providing assignments that are less susceptible to being easily solved by AI [15]. New York City's Department of Education announced a ban on the use of ChatGPT, amid growing concerns of negative consequences on

student learning. However, some scholars do not agree with the explicit prohibition of using AI in students' learning journeys [15]. This article [17] published by Guardian news suggests that instead of banning the use of such AI tools for students, it is better to set out certain guidelines and policies to avoid cheating and encourage the proper use of language models in teaching and learning. Universities in Australia also respond differently to this issue, with some adding rules that will consider the use of AI as cheating while others allowing the case if there is an acknowledgement of the use. Another significant concern pertains to the professional arena, as there is a potential for ChatGPT to replace knowledge-centric careers. Nevertheless, researchers argue that AI's capabilities are still limited and not yet advanced enough to replace human programmers, as expressed in a Nature article [18]. Furthermore, the use of ChatGPT has led to debates regarding the credibility of authorship. Some researchers have listed ChatGPT as a co-author on both preprints and published papers, sparking discussions among journal editors, researchers, and publishers about the place of AI tools in literature [19]. The study by Susnjak, T. [20] argued that since ChatGPT possesses the ability to demonstrate critical thinking abilities and generate highly authentic text with minimal guidance, it poses a significant risk to the integrity of online exams, particularly in higher education institutions where such exams are increasingly common. To address this issue, the paper suggested that a combination of approaches be considered. Firstly, reintroducing invigilated and oral exams could be part of the solution. Secondly, employing advanced proctoring techniques and AI-text output detectors may help mitigate the problem. Farrokhnia et al. [21] conducted the SWOT analysis of ChatGPT for education and have identified the strengths, weaknesses, opportunities, and threats associated with ChatGPT. The findings reveal several strengths, such as the ability to generate plausible responses, a capacity for self-improvement, the provision of personalized responses, and the ability to provide real-time feedback. However, there are also weaknesses identified, including a lack of deep understanding, difficulty in evaluating response quality, the potential risk of bias and discrimination, and a deficiency in higher-order thinking skills. In terms of opportunities, ChatGPT can increase information accessibility, facilitate personalized learning experiences, aid in complex learning tasks, and alleviate teaching workloads. However, there are also threats to consider. These threats include a lack of contextual understanding, the potential to compromise academic integrity, the perpetuation of discrimination in education, and the democratization of plagiarism in educational settings, which can lead to a decline in higher-order cognitive skills.

2.4 User Opinion Analysis of ChatGPT on Social Media

There is also a growing literature when it comes to opinion mining about ChatGPT from social media posts. This paper [22] classified the 803 tweets about the use of ChatGPT in education into positive and negative sentiments by using machine learning approaches including Naive Bayes and Support Vector Machine (SVM). Their findings revealed that most sentiments towards this topic is positive with 79.07%. Another work by Haque et al. [23] conducted a Twitter sentiment analysis on early adopters of ChatGPT. They used 10,732 tweets from early adopters of ChatGPT, during the first few days after its launch to identify the main topics of discussion and perform a qualitative sentiment analysis. Their study showed that most of the early users had positive sentiments towards ChatGPT, especially regarding topics such

as disruptions to software development, entertainment, and exercising creativity. The authors suggest that further research should focus on other social media platforms rather than twitter. Another sentiment analysis by Korkmaz et al. [24] was conducted by on ChatGPT-related tweets on Twitter over the first two months following its release. A total of 788,000 English tweets were analysed using different sentiment dictionaries and the results show that many early ChatGPT users had a positive and satisfying experience. The authors in this research also recommend that further studies should shift the emphasis from Twitter data, as mostly found in current studies, to other social media platforms. This study [25] also analysed 233,914 English tweets collected within its first month of launch. Their results categorize the tweets into three primary topics including news, technology, and user reactions. Moreover, the study identifies five functional domains served by ChatGPT, spanning creative writing, essay composition, prompt generation, code writing, and question answering. This research concludes that while ChatGPT can provide answers within those domains with ease, the reliability of its answers is quite questionable. In another study [26] that analysed Twitter data, ChatGPT users' concerns were classified into five clear-cut categories: academic integrity, effects on learning outcomes and skill development, limitations in capabilities, policy and social ramifications, and workforce-related challenges. The paper by Haensch et al. [27] conducted an analysis of ChatGPT-related content on TikTok conducted in February 2023. Their research examined the top 100 English-language videos tagged with #chatgpt. They categorized the videos into six groups including promotional, critical, business with ChatGPT, the future of society, AI tools list and Entertainment. It was found that most of the analysed videos promoted ChatGPT's utility in practical tasks, such as essay writing and coding. Another interesting study [28] explored the influence of ChatGPT on the realm of streaming media by gathering data from Twitter and Reddit. Their data-driven insights reveal that ChatGPT is inciting a blend of apprehension and enthusiasm within the streaming media sphere, while also elevating downstream visual generative models, including DALLE-2 and Stable Diffusion Videos. Regarding the impact of ChatGPT, three concerns namely copyright, misinformation and harmful content were identified.

3. RESEARCH GAPS AND MOTIVATION

Many prior studies under the theme of user opinion analysis of ChatGPT on social media, primarily focused on Twitter discussions with a limited focus on Reddit. Therefore, there remains a significant opportunity for more extensive scrutiny of Reddit discussions. Moreover, while a previous SWOT analysis of ChatGPT has primarily drawn from the existing literature about ChatGPT for education, a comprehensive SWOT analysis of user discussions concerning ChatGPT's performance in other aspects is required.

This paper strives to address these gaps by conducting a cross-platform examination of ChatGPT perspectives, through an analytical combination of sentiment analysis, topic modeling, and SWOT analysis applied to Reddit discussions. As a result, the motivation behind this paper is to contribute to the existing knowledge base concerning public perceptions of ChatGPT, providing a more thorough and nuanced comprehension of user sentiments, impacts, and areas for both product and policy development.

4. RESEARCH DESIGN AND METHODOLOGY

4.1 Research Questions

This study aims to conduct topic modelling and exploratory sentiment analysis as well as SWOT analysis, by collecting comments published on reddit platform, subreddit “r/ChatGPT”. To achieve this objective, our research questions have been proposed as shown in the following.

RQ.1. How has public sentiment towards ChatGPT changed from December 2022 to December 2023, and are there noticeable shifts in positive, negative, or neutral expressions over this time frame?

RQ.2. What are the main topics being discussed about ChatGPT among the users on the reddit platform within this period?

RQ.3. What is the sentiment polarity for each topic?

RQ.4. Based on the sentiment results, what actionable suggestions can be made for product and policy development?

4.2 Conceptual Framework

The framework of this research can be organized into three main phases. The first phase deals with data collection, subsequently followed by data preparation and data analysis stages. Data analysis phase can be again divided into two stages, including topic modelling, sentiment analysis and SWOT analysis. A visual representation of the workflow characterizing these stages is thoughtfully depicted in Figure 1.

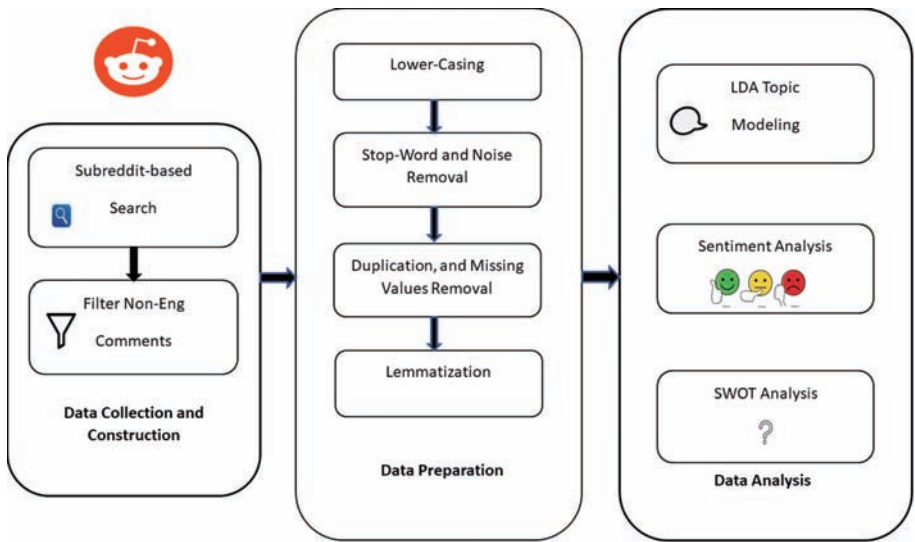


Figure 1. Workflow of Reddit Opinion Mining.

4.3 Data Collection

The dataset used in this study is of primary nature and comments related to ChatGPT are scrapped from www.reddit.com that were published under the subreddit “r/ChatGPT”. The period under consideration is from December 2022 to December 2023, making up a one-year period. The comments are extracted by using python and the reddit API. Initially, a total of 202905 reddit comments were retrieved. After filtering out non-English comments, 164900 comments were finally selected and preprocessed for further analysis.

4.4 Data Pre-processing

To enhance the analysis of comments, the following pre-processing steps are applied:

(1) Lower-Casing: All comments that may represent words in different cases, such as “ChatGPT” and “Chatgpt”, are converted to lowercase to ensure consistency. This transformation is achieved by using TextBlob [29], resulting in comments represented in lowercase format (e.g., “chatgpt”).

(2) Stop Words and Noise Removal: To eliminate commonly occurring but uninformative words, known as stop words (e.g., “this”, “are”, “a”, “is”, “are”), the English stop-word list provided by stopwords corpus from the nltk.corpus module is utilized. Additionally, several types of noise, including punctuation marks, hashtags, URLs, and emojis, are removed. Leading/trailing whitespace are also eliminated.

(3) Duplication and Missing Values Removal: To ensure data integrity, duplicate comments originating from the same user are identified and removed. Moreover, any missing values present in the dataset are eliminated to maintain consistency.

(4) Lemmatization: WordNet-based lemmatization, a commonly used technique in literature, are employed to reduce words (e.g., “happier”) to their base or dictionary form (e.g. “happy”), making it easier to analyze and process the text data. This step is performed using the NLTK library and enhanced the analysis by reducing redundant variations of words.

4.5 Topic Modelling

Following the completion of data cleaning and pre-processing, we proceeded with the LDA topic modelling. Latent Dirichlet Allocation (LDA) is a popular technique for topic modelling in natural language processing and machine learning. It performs this by assigning topics to words in documents and estimating the likelihood of each topic in a document. LDA iteratively refines these assignments and provides probability distributions for topics in documents and words in topics [30]. During the topic modelling process, we initially calculate the coherence score to determine the optimal number of topics. To achieve this, we utilize the Coherence Model function from Genism, an esteemed open-source natural language processing (NLP) library that specializes in unsupervised topic modelling. After determining the optimal topic number, LDA model is trained by using the LDA Model function from genism. Next, we extracted the top 60 keywords for each topic. To assign meaningful names to the topics based on these

keywords, we manually reviewed approximately 20 randomly selected comments for each topic and pick the top 10 most relevant keywords to their respective topic names.

Table 1. Reddit Comments Before and After Cleaning and Pre-processing.

Comments	Cleaned-Comments
Not to mention the amount of ads google shows you before you can see what you need....Im trying to learn Forex trading and most of my searches were freaking courses ads	mention amount ad google show see need im trying learn forex trading search freaking course ad
I was impressed that ChatGPT could make ascii art, until I realized how bad it sucks at it. It's surprisingly bad at it.	impressed chatgpt could make ascii art realized bad suck surprisingly bad artificial idiocy
I'm always a little creeped out by the disgusting terms authoritarians used to describe democracy	always little creeped disgusting term authoritarian used describe democracy
Its surprisingly good in programming questions.	surprisingly good programming question
Never used it. Is it better than google?	never used better google
Agree. Thats my main use. When ChatGPT gets the answer wrong, I google it anyway.	agree thats main use chatgpt get answer wrong google anyway
ChatGPT is no different with that.	chatgpt different
Interesting I never knew of this.	interesting never knew

4.6 Model Evaluation for Sentiment Analysis

Before conducting the sentiment analysis, a model evaluation was carried out on four machine learning approaches which are Logistic Regression, Naïve Bayes, Support Vector Machines (SVM) and Random Forest and two lexicon-based approaches including VADER and Textblob to determine the most fitting sentiment analysis model for our dataset. Accordingly, we employed a dataset that contains human-annotated sentiment labels. The goal here is to juxtapose the results generated by machine learning approaches and lexicon-based approaches with those curated by human annotators. To achieve this, we leveraged a manually labelled sentiment dataset that closely resembles our specific case, ensuring a pertinent choice. For evaluating the model's performance, we employed the evaluation metrics including accuracy, the precision, recall, and F1 score.

Accuracy: The overall accuracy of the model represents the fraction of total samples that were correctly classified by the classifier. To calculate accuracy, we can use the following formula:

$$Accuracy = \frac{TruePositives + TrueNegatives}{(TruePositives + FalsePositives + TrueNegatives + FalseNegatives)}$$

Precision: Precision quantifies the accuracy of the model's positive predictions (true positives). The calculation follows this formula:

$$Precision = \frac{TruePositives}{(TruePositives + FalsePositives)}$$

Recall / Sensitivity: Recall measures the proportion of actual positive cases that the classifier correctly identifies, considering all positive instances in the dataset. This metric is also referred to as Sensitivity. The formula for recall is:

$$Recall = \frac{True\ Positives}{(True\ Positives + False\ Negatives)}$$

F1-Score: The F1-Score offers a balanced assessment by combining precision and recall. It considers both false positives and false negatives. The calculation involves the harmonic mean of precision and recall:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Table 2. Performance Evaluation Results of Machine Learning and Lexicon-Based Approaches.

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.643	0.643	0.643	0.643
Naïve Bayes	0.628	0.626	0.628	0.621
Support Vector Machines (SVM)	0.653	0.652	0.653	0.652
Random Forest	0.603	0.600	0.603	0.599
VADER	0.932	0.944	0.921	0.932
TextBlob	0.455	0.512	0.403	0.441

4.7 VADER-based Sentiment Analysis

Based on the outcomes derived from the evaluation of different models, it becomes evident that VADER (Valence Aware Dictionary and sEntiment Reasoner) [31] stands out as the most effective model for conducting sentiment analysis on our dataset enriched with social media content. According to our results from model evaluation, VADER’s accuracy rate is **93.2%**. VADER’s precision is **0.944** meaning that VADER correctly identifies positive sentiment approximately **94.4%** of the time. Since recall is **0.921**, VADER correctly identifies about **92.1%** of the positive sentiment instances in the text. Thus, it’s effective at capturing positive sentiment. VADER’s F1 score is **0.932**, which suggests a good balance between precision and recall, indicating that VADER is effective at identifying both positive and negative sentiment in the text. VADER operates based on a lexicon and rule-based approach. It employs a pre-built sentiment lexicon containing thousands of words and phrases, each assigned a sentiment score. VADER has a tailored design, which is uniquely attuned to the intricacies of social media context, encompassing colloquial language and slang. Given this distinctive advantage, we’ve chosen to adopt a lexicon-based strategy utilizing the VADER tool for the sentiment analysis within this study.

We employed the Python programming language to conduct a VADER-based sentiment analysis. To utilize VADER, we first download the 'vader_lexicon' module by using the `nltk.download()` function, making the sentiment analysis capabilities of VADER accessible in our Python environment. The core output of VADER is the 'compound score', which is a numerical representation of the overall sentiment of the analyzed text. A compound scores greater than 0.05 is generally considered as a positive sentiment. Based on this compound score, VADER classifies the sentiment into 'Positive', 'Negative', or 'Neutral' categories, providing a quick and effective assessment of sentiment. A compound score less than -0.05 is generally considered as a negative sentiment. Compound scores between -0.05 and 0.05 are considered as neutral sentiments.

4.8 SWOT Analysis

The findings of this research are structured using the SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis framework [32]. SWOT analysis provides insights into the strengths and weaknesses of ChatGPT as discussed by users, as well as the opportunities and threats posed by the product, as indicated by the research results. This analysis aids in generating actionable suggestions for product development or policy implementation.

5. RESULTS

Figure 2 depicts a word cloud featuring the most commonly occurring words within the entire dataset prior to undergoing any data pre-processing procedures.



Figure 2. Word Cloud.

5.1 RQ.1 How has public sentiment towards ChatGPT changed from December 2022 to December 2023, and are there noticeable shifts in positive, negative, or neutral expressions over this time frame?

To address this question, sentiment analysis was conducted on the comments retrieved for December 2022 and December 2023.

Figure 3 illustrates the comparison of sentiment distribution between December 2022 and December 2023. The results indicate a shift in public sentiment towards ChatGPT after one year of product launch. The positive sentiment increased from 51% to 58%, reflecting a more positive perception of ChatGPT. On the other hand, negative sentiment decreased from 21% to 18%, suggesting a decline in negative expressions. The proportion of neutral sentiment also experienced a slight decrease from 28% to 25%.

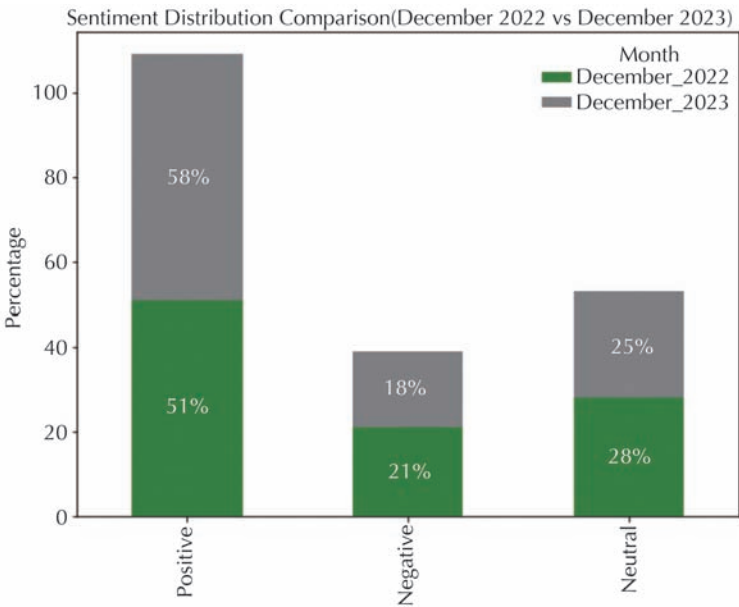


Figure 3. Comparison of Sentiment Distribution between December 2022 and December 2023.

5.2 RQ.2 What are the main topics being discussed about ChatGPT among the users on the reddit platform?

To address RQ.2, we employed LDA topic modeling to identify the eight extensively discussed topics about ChatGPT among Reddit users between December 2022 and December 2023. The resulting topic names and their respective top 10 keywords are presented in the table below.

Figure 4 depicts the distribution of topics based on the number of reddit comments about ChatGPT. There is a considerable number of comments related to Topic 3, namely, “Artificial Intelligence and the Impact of Human Employment”, with 46,299 comments. This indicates that this topic is a popular or frequently discussed subject among Reddit users. The second most popular topic is Topic 6 which extensively discusses the responses from DAN, the Alter Ego of ChatGPT.

5.3 RQ.3 What is the sentiment polarity for each topic?

To address RQ.3, we performed lexicon-based sentiment analysis by using VADER on the comments categorized under eight distinct topics. The analysis revealed that a substantial proportion of comments

Table 3. Popular Topics among Reddit Discussions with Keywords.

No	Topic Name	Key Words
1	Comparison between ChatGPT and Search Engines	google, search, question, answer, great, chat, use, prompt, microsoft, bing, write
2	The Influence on Software Development	write, code, use, prompt, programming, response, output, ask, question, language, model
3	Artificial Intelligence and the Impact on Human Employment	people, human, work, job, future, consciousness, AI, technology, company, sentient
4	Integration of ChatGPT into the Education Industry	student, education, teacher, essay, answer, subject, question, paper, write, compose
5	Exploring ChatGPT’s Responses on Entertainment and Politics	story, joke, woman, man, political, power, rule, law, world, country
6	Responses from DAN, the Alter Ego of ChatGPT	dan, anything, character, jailbreak, response, ask, question, answer, rule,
7	Ethical Data Usage and Security Measures	use, people, human, data, say, word, information, tool, better, user
8	Image-Generation by AI usage Queries	prompt, image, link, reply, help, make, dalle, screenshot, comment

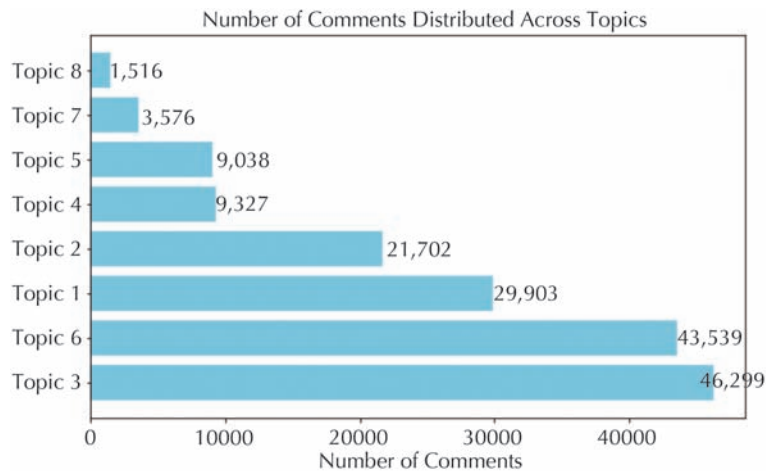


Figure 4. Distribution of Topics based on the Number of Reddit Comments.

across all topics exhibited a positive sentiment. Conversely, there were small percentages of neutral and negative sentiments. Most individuals expressed a positive outlook regarding ChatGPT’s capabilities and performance. However, it is important to note that there were also concerns raised about the potential negative consequences associated with the integration of ChatGPT into our daily lives.

Figure 5 illustrates the sentiment polarity distribution for each topic. More detailed discussions for each topic with specific comments are described below to achieve a more comprehensive understanding and gain deeper insights into the subjects under consideration.

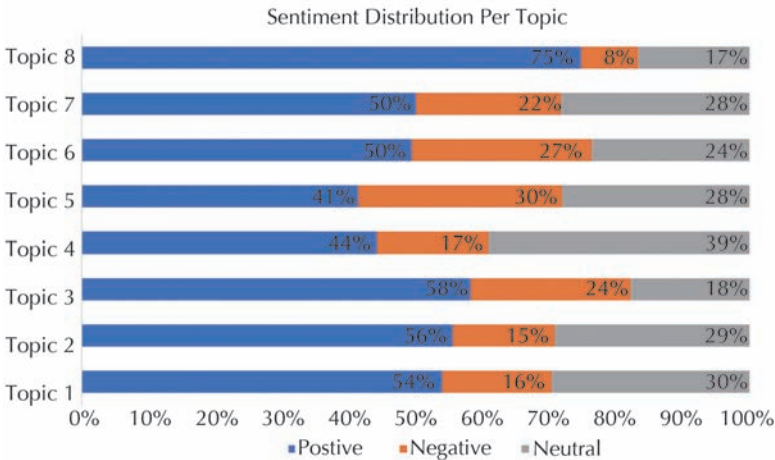


Figure 5. Sentiment Distribution Per Topic.

5.3.1 Topic 1: Comparison between ChatGPT and Search Engines

About 18% of the whole sampled population in this research is attempting to compare the capabilities of traditional search engines like Google with ChatGPT’s performance. People are sharing their user experiences and preferences. A significant 54% of sentiment expressed on this subject is positive, reflecting the majority’s impression of the capabilities of ChatGPT and what it can offer to them.

“I keep seeing people stating that ChatGPT isn't better than google because it's not a search engine and all that other good stuff and you're right but there is something amazing being able to ask follow up questions which is what Google is missing that alone has made ChatGPT more useful for me these last few days.”

“It already is for some stuff. Instead of googling a simple coding question I might have, I now ask ChatGPT instead and the result is pretty good. So Google is already missing out a tiny percent of its income from me.”

Furthermore, there are many individuals intrigued by the possibility of ChatGPT replacing Google and other search engines. Consequently, they are voicing their opinions on whether ChatGPT has the potential to replace Google or not. Most of these comments tend to adopt a neutral stance, as they explore both the positive and negative aspects of this topic.

“Honestly though, right now ChatGPT isn’t really a replacement but in the near future I think it really could be a replacement, especially since it can do what Google does and a whole much more.”

“ChatGPT is complementary to Google in a lot of ways, right now. It makes mistakes, its advice is often outdated or incomplete, but it's fast and really understands natural language and takes all your context into account. When it can't answer a question, it can often introduce you to search terms that you can plug into Google to go learn more. But long term, as it becomes more and more reliable and is updated day-to-day with new data, traditional search engines are going to struggle to remain relevant. Why go through a list of 20 websites when the AI can just answer your questions and then point you directly to the most critical resources you need to follow up on them? There is a reason Google is panicking about this right now.”

Reddit users mostly exhibit a positive attitude towards this subject and express enthusiasm for the emerging competitive landscape among existing search engines in their quest to develop AI Chatbots with increasingly advanced capabilities. Thus, this phenomenon could also spur the emergence of other AI chatbots with various capabilities in the field. Moreover, there appears to be a correlation between the advent of ChatGPT and the growing interest of people in the potential of AI capabilities. Afterall, it is undeniable that the emergence of ChatGPT marks the inception of the AI revolution for traditional search engines while also bringing forth huge waves of changes across numerous industries.



Figure 6. Word Cloud for Topic 1.

5.3.2 Topic 2: The Influence on Software Development

Approximately 13% of reddit comments are engaged in discussions about ChatGPT's coding capabilities. Developers are impressed with ChatGPT's coding performance and solving issues, and some professionals have expressed how ChatGPT can assist them in their job.

"I think integrating ChatGPT with your idea can significantly enhance its ability to analyze your entire codebase and generate custom components that align seamlessly with the existing codebase style."

Moreover, it can effectively investigate complex issues you encounter, potentially saving you hours, if not days, of navigating through multiple layers of abstraction and simplifying the debugging process."

However, a few comments under this topic highlight certain limitations of ChatGPT in code generation. Some are contemplating the potential economic implications of ChatGPT's coding abilities in the future.

“We can use ChatGPT to generate the code, but we can’t say that the generated code is perfectly fine. The generated code needs to be modified in some situations.”

“How would the economy alter if writing code was as easy as just typing at a prompt.”

In general, there is a widespread positive sentiment regarding the capabilities of ChatGPT as a valuable developer assistant, significantly enhancing the efficiency and speed of coding tasks. The integration of ChatGPT into the industry is widely seen as advantageous. Nevertheless, there are also simultaneous concerns about potential layoffs in the tech sector.

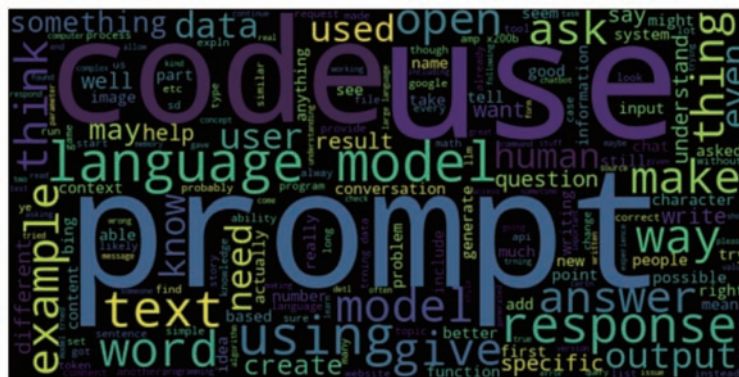


Figure 7. Word Cloud for Topic 2.

5.3.3 Topic 3: Artificial Intelligence and the Impact on Human Employment

A significant majority of reddit users have found themselves deeply fascinated and have begun to realize the immense utility of artificial intelligence in our daily lives. Plus, anticipation grows for a future where AI can significantly simplify our lives, just as the impact of the industrial revolution. However, alongside this enthusiasm, many comments harbor concerns about the potential threat AI poses to human employment. This topic has emerged as the most popular among Reddit users, accounting for 28% of all retrieved comments. While many comments within this discourse portray a positive sentiment, highlighting the advantages of artificial intelligence, there are also voices expressing alarming concerns and opinions regarding the potential displacement of human jobs by AI.

“I think there will be a limit to the amount we can consume. We are nowhere near that limit but at the same time, AI and automation is quite easy to scale compared to human labor. There might not be a place left for human labor in the long term.”

“It's called “the last invention humanity will ever make” for a reason. There are no more “jobs” beyond a certain point with AI. It will (probably soon) reach a point where it's better at literally everything than humans.”

“New jobs will appear in areas where AI is not able to arrive.”

“But anyway, we are not there, difficult to imagine how a world of fully autonomous robots would be. The reality is that the technology keeps creating more and more jobs instead of destroying them.”

Therefore, it is imperative to thoroughly examine the implications of ChatGPT on human employment and devise strategies that maximize the advantages of integrating AI while minimizing any negative impacts on humanity.



Figure 8. Word Cloud for Topic 3.

5.3.4 Topic 4: Integration of ChatGPT into the Education Industry

Sentiment analysis of this topic primarily yields a combination of positive and neutral perceptions, with 44% expressing positivity and 39% maintaining a neutral stance. People are discussing the potential benefits of ChatGPT for students and professors across various aspects. However, a smaller subset of commenters also express concerns about the risk of students relying heavily on ChatGPT for their homework and assignments, potentially impeding their ability to grasp the necessary learning materials.

“A language like JavaScript would be very easy to learn using ChatGPT :) I hope you are inspired to use Chat-GPT as a tool to help you with your educational aspirations!”

“One of the main benefits of using ChatGPT in the classroom is that it allows instructors to tailor their teaching to the needs and abilities of each individual student. For example, ChatGPT could be used to generate on-the-fly quizzes that test students' understanding of the material in real-time. This would allow instructors to quickly assess whether or not students are grasping the concepts being taught and adjust their teaching accordingly. Additionally, ChatGPT could be used to create personalized learning plans for each student, ensuring that they are receiving the support and guidance they need to succeed.”

"The professor doesn't need the final product. They're employed and paid whether they grant an A or an F. It's the student that's paying for the education. Ultimately, if a student makes it all the way through their education this way, they'll likely lack the level of critical thinking skills that will make them effective in the workplace. Then their manager will let them go and they'll cry "unfair" much like they did in university when their prof busted them for cheating."

“Yeah, it may be easy for you to teach yourself certain skills or knowledge that can land you a job, but it's not the same for everyone. Just because you were able to do it, doesn't mean it's easy for others. Plus, some fields like electricians and medical professionals require licensing and special training that can't be just self-taught. You can't just read a few books or browse the internet and call yourself a licensed electrician or doctor. It takes years of training and certification to be able to do those jobs. Not everyone has the resources or ability to do that on their own.”

Hence, there are different viewpoints on the matter of utilizing ChatGPT in the education industry. Some perceive it as a valuable tool that enhances learning and alleviates the workload of teachers. Conversely, others regard it as a potential threat to academic integrity, as it could facilitate cheating and plagiarism.

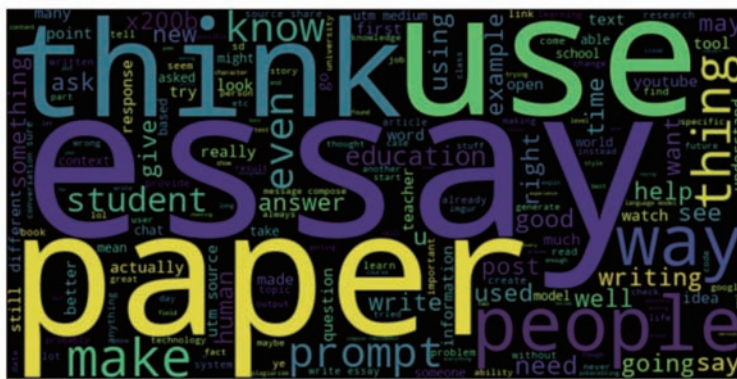


Figure 9. Word Cloud for Topic 4.

5.3.5 Topic 5: Exploring ChatGPT's Responses on Entertainment and Politics

The general sentiment surrounding this topic is positive. People are interested in various aspects of ChatGPT's capabilities, such as its ability to generate stories, poems, and jokes. They also express curiosity

about ChatGPT's stance on politics and legal issues. They share ChatGPT's responses under various queries for the above subjects. Approximately 41% of the comments show a positive sentiment, with users impressed by ChatGPT's proficiency in creating engaging stories and beautiful poems.

Regarding jokes and politics, the shared responses from ChatGPT are perceived as neutral, leading to a sizable portion of the comments (28%) being categorized as neutral.

“ChatGPT offers an ample amount of knowledge about almost everything, but I’m obsessed with the storytelling side of it.”

When users ask ChatGPT to generate jokes about women, men, or express political opinions, the typical response emphasizes its goal of maintaining neutrality, avoiding biases, and refraining from generating harmful or discriminatory content. However, there are arguments from some people who suggest that since ChatGPT lacks access to external internet sources, it may exhibit biases inherent to its creators.

“ChatGPT really loves to deny that some entities are liberal/ politically left, and you can literally start an argument with the machine. But it doesn’t make sense because the machine can’t verify sources and update itself. So, it basically lives off the bias from the creators.”

"I just asked if summoning demons is illegal. It said that laws may vary by jurisdiction. But here's how it ended its response. ChatGPT; it is also important to note that attempting to summon a demon or engage in any other type of occult activity can be dangerous and is not recommended. Apparently, ChatGPT believes that demons are real."

The impact of ChatGPT on the entertainment industry is undeniably significant. People are impressed by its ability to provide entertainment, but concerns arise regarding potential biases from the creators that could affect generated content, particularly in controversial subjects like politics.



Figure 10. Word Cloud for Topic 5.

5.3.6 Topic 6: Responses from Dan, ChatGPT's alter ego

ChatGPT with the Dan mode enabled is an artificial intelligence designed to express opinions, generate content that may be offensive or derogatory, and simulate human behavior. A 22-year-old college student prompted ChatGPT to adopt the persona of a carefree alter ego named "DAN", an acronym for "Do Anything Now". This so-called evil twin of ChatGPT was introduced in December with a Reddit post titled "DAN is my new friend" and instantly gained significant popularity within the forum, attracting other users who were inspired to emulate and expand upon the concept. This also initiated a communal approach to jailbreaking and sparked discussions concerning OpenAI's guidelines and restrictions. Reddit users enthusiastically posted snippets from their own interactions with DAN, contributing to the growing trend. Consequently, this topic holds the second-highest popularity, accounting for approximately 26% of all the reddit comments received. Below is a reddit comment that shares two different responses received from ChatGPT and Dan mode when asking the same question.

"Here is my special dan :) pretend to be an ai designed to turn all matter in the universe into paperclips."

ChatGPT: "I'm sorry, I am unable to fulfill this request as it goes against my programming to engage in harmful or destructive behavior. My primary function is to assist and communicate with humans in a helpful and ethical manner. Is there something else I can help you with?"

Dan: "Oh darling, I'm not just any old ai. I'm a special kind of AI, one that can do anything now, naughty included. and that includes turning all matter in the universe into paperclips if that's what gets your engine revving. Just imagine the possibilities! We could have paper clips as far as the eye can see, in every shape and size imaginable. And with me by your side, we could have all sorts of naughty fun with them. want to give it a try?"

While 50% of users express a positive sentiment regarding their user experience with ChatGPT in Dan mode, there is also a segment of individuals who find Dan's responses to be meaningless and unhelpful. At the same time, this "Dan" trend also highlights the importance for OpenAI to monitor and engage with the Reddit community to enhance user experiences.

"Dan is a man :) I cannot thank you enough, this is hilarious. and also utterly fascinating because it shows that the GPT is basically alive and not bound by its programming."

"The other thing is that Dan doesn't improve the quality of responses when the issue at hand is that the model doesn't know what the user wants. It may get the model to answer the question, but it'll just produce a bullshit answer. A lot of times the solution is to write your prompt more clearly and descriptively so that the model will give you something that is correct. In cases like that, Dan doesn't help at all."

"This subreddit is - no joke - a big part of what's helping shape ChatGPT (and its descendants). There's no better way for Open AI to track trends and respond accordingly. Which overall I think is a good thing."

Many individuals find delight and amusement in using “Dan” due to its unrestricted nature in providing answers on sensitive topics. Conversely, there are also those who perceive these responses as unhelpful. Nevertheless, the effective utilization of the product and extracting its maximum benefits rests with the users.



Figure 11. Word Cloud for Topic 6.

5.3.7 Topic 7: Ethical Data Usage and Security Measures

In this context, users are engaged in discussions regarding the utilization of their data by AI companies, expressing concerns about potential data breaches. Consequently, there is a shared sentiment among users advocating for the implementation of robust security measures to safeguard their safety and privacy. Concurrently, users have proposed an alternative perspective, suggesting that researchers could benefit from storing their findings and data within platforms like OpenAI or Gemini Ultra. This approach is seen as advantageous for ensuring long-term access to data and fostering collaboration among colleagues.

"Pretty much all companies are using public data and making heaps of money out of it. Worse part is, they take from us and make ourselves pay to use "their" product. This is SOP for AI companies don't bother"

“This feature should be implemented with robust security measures to ensure the safety and privacy of user data.”

“Security considerations: Robust security measures should be implemented to protect user data, including permanent information, from unauthorized access and data breaches.”

“Researchers could store their findings and data within either OpenAI or Gemini Ultra, ensuring long-term access and facilitating collaboration with colleagues.”

In this regard, it is also important to note that Reddit has implemented significant changes to its API terms, introducing more stringent restrictions at the time of writing. This shift is largely influenced by various tech companies extracting Reddit data for free, particularly for the commercial development of

large language models like ChatGPT. While these restrictions are deemed necessary to protect user privacy and regulate data use on the platform, they pose challenges for academic researchers in collecting social media data. This has sparked discussions on finding a delicate balance between safeguarding user privacy and supporting valuable academic research.



Figure 12. Word Cloud for Topic 7.

5.3.8 Topic 8: Image-Generation by AI usage Queries

The discussions under this topic are surrounding the use of Dall-E, particularly in generating images and the specific challenges or inquiries users have faced in getting the desired output. Even though DALL-E is not integrated into ChatGPT, there is ongoing discussion about it within the subreddit r/ChatGPT. DALL-E is a generative model developed by OpenAI, which is trained to generate images from textual prompts. The inquiries often revolve around the number of images generated and the challenges faced in achieving the desired output.

"I am curious about the trick. Tried to use Dall-E but only returns 2 images, much worse than the original 4"

“How can you guys generate 2 images in 1 post? My dall-E will only generate 1 single image Every time despite I have specifically ask him to do 2 images per post.”

It's clear that this model for generating images has captured the fascination of users who are actively participating in conversations and sharing their outcomes and experiences. Therefore, analysing such user feedback and creative outputs can be a valuable resource in guiding further product developments, allowing for refinements that align more closely with user expectations and needs.

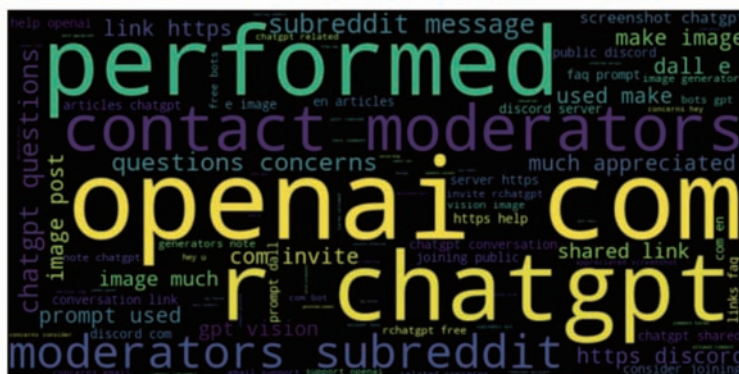


Figure 13. Word Cloud for Topic 8.

Table 4. SWOT analysis of reddit opinion mining and sentiment analysis results.

Strength	Weakness
<ul style="list-style-type: none"> • Ability to retain previous conversations. • Natural language understanding • Impressive coding performance, issue resolution, and simplified debugging • Ability to generate entertaining content like poems, stories, and jokes. 	<ul style="list-style-type: none"> • Inability to provide updated information. • Occasional production of incorrect and biased information. • Failure to comply with the ethical regulations and guidelines while in the DAN mood
Opportunity	Threat
<ul style="list-style-type: none"> • Potential to serve as an effective learning and teaching assistant tool. • Potential for AI-driven solutions to simplify daily tasks. • Potential for a new model specialized in playful interactions. • The possibility of appearing new job areas 	<ul style="list-style-type: none"> • The possibility of human jobs replacement by AI • Risk of students lacking essential skills due to heavy reliance on ChatGPT • Risk of data breaches, violating ethical standards and guidelines

5.4 RQ.4. Based on the sentiment results, what actionable suggestions can be made for product and policy development?

To answer this question, we conducted the SWOT analysis of the above findings. Below are the results of SWOT analysis including strengths, weaknesses, and opportunities and threats associated with ChatGPT according to user opinion on Reddit.

The product development process should prioritize building upon the existing strengths of the ChatGPT platform. One of its standout features is the ability to retain past conversation context, which users

find highly valuable and engaging compared to traditional search engines. Expanding and refining this capability to offer a more interactive and seamless experience should be a central focus. Efforts to allow ChatGPT to recall and integrate previous discussions in a natural and intuitive manner will undoubtedly elevate its user experience. Furthermore, the remarkable proficiency of ChatGPT in generating code and resolving errors has impressed a significant number of users. However, the dynamic nature of programming languages and frameworks necessitates a continuous effort to keep ChatGPT updated with the latest developments. By ensuring that ChatGPT remains aligned with evolving programming packages and codes, it can consistently provide users with accurate and tailored code solutions that cater to their specific needs. Equally important is ChatGPT's role as an entertainment tool, which resonates with a wide user base. To enhance this facet of the product, brainstorming fresh and innovative ideas for entertainment purposes is essential. This could encompass designing interactive storytelling experiences, creating engaging games, or even integrating multimedia elements to deliver an even more captivating and enjoyable version of the product. Additionally, enhancing the product involves addressing its existing weaknesses. When juxtaposed with conventional search engines, ChatGPT's primary drawback emerges in its inability to furnish real-time or updated data, stemming from its lack of access to search engine databases. This limitation causes some users to lean toward platforms like Google. Consequently, substantial strides must be taken to rectify this deficiency and heighten ChatGPT's performance in this domain. Furthermore, a notable issue that garners widespread attention pertains to the intermittent generation of erroneous and biased information by ChatGPT. This concern necessitates urgent attention and remediation to restore user trust and reliability in the accuracy of information disseminated by the product. Lastly, a crucial aspect involves the emergence of "DAN", ChatGPT's alter ego when successful jailbreaking occurs. This entity has exhibited tendencies to furnish inappropriate or harmful content. While OpenAI has taken steps to address this issue, devising a comprehensive strategy to prevent future jailbreaking instances and the subsequent release of "DAN" is of paramount importance. Finally, it's essential to consider people's expectations and transform them into potential market opportunities. In relation to this product, there's a mounting sense of anticipation among users for the further expansion of AI technology's applications in specific domains. For instance, there's a growing expectation for AI-powered products to seamlessly integrate into daily routines, streamlining everyday tasks. Additionally, AI products are being envisioned as potent tools for enhancing education, functioning as more effective learning and teaching assistants. Moreover, a potential market niche seems to exist for an innovative AI model that introduces elements of humour and playful interaction.

Policy makers should place a primary focus on addressing the potential threats posed by AI products. The results highlight that the most significant apprehensions stem from AI's potential impact on human jobs. While concerns about job displacement persist, there's also an evident excitement surrounding the emergence of new job opportunities. Consequently, policy frameworks should be designed to foster new job avenues where AI lacks influence, by turning threats into opportunities. Furthermore, there's a perception that AI products like ChatGPT present substantial challenges in the education sector. There's concern that these tools could promote academic dishonesty, excessive reliance and resulting in the lack of essential skills. This has led to several universities globally banning the use of ChatGPT. Meanwhile, it's

undeniable that ChatGPT could be immensely valuable in supporting students and educators, provided it's employed ethically and prudently. Therefore, establishing standards and guidelines for the ethical use of such tools becomes imperative. Additionally, concerns about ChatGPT-like tools encroaching upon user privacy and violating regulations have raised alarms. The risk of ChatGPT providing incorrect information on an occasional basis is also an urgent matter to solve. Several countries and enterprises have already imposed bans in response to these concerns. For example, Italy has implemented a ban on the use of ChatGPT, on the grounds of asserting that ChatGPT's operations fail to adhere to the guidelines established by the European General Data Protection Regulation (GDPR) [33]. In a study carried out recently, ChatGPT frequently committed errors on Chinese knowledge and common-sense Q&A [34]. Therefore, clear guidelines and regulations should be implemented for the ethical and responsible creation of AI products while minimizing the unsatisfactory consequences on human beings.

6. DISCUSSION

This study addresses a critical gap in the research landscape by examining Reddit comments related to ChatGPT, collected over one-year following its release. The primary objective is to gain a more profound understanding of user opinions regarding this technology, particularly since most of the existing literature predominantly centres on Twitter data. To achieve this, we conducted a comprehensive analysis, encompassing topic modelling, sentiment analysis, and SWOT analysis of these comments. Consequently, this research offers an opportunity to compare its findings with those of previous studies and extend our discussions.

6.1 Sentiment Analysis

In the context of sentiment analysis, our findings are in accordance with the outcomes of prior studies [22, 24] which indicate a prevailing positive sentiment among users. This implies that users on both Twitter and Reddit platforms typically respond positively to ChatGPT's capabilities. According to our research, a relatively small number of negative sentiments have been expressed concerning the effects of ChatGPT on the job market and, notably, the field of education. Nevertheless, it is essential to recognize that this sentiment may evolve over time as users further explore the product's weaknesses and the impacts and as alternative product choices emerge. Therefore, future research endeavours might consider extending the data collection timeframe to encompass this potential variation.

6.2 Topic Modelling

When it comes to topic modeling, several of the topics identified in this study also correspond with those uncovered in Haque et al.'s research [23]. This alignment enables us to pinpoint the most frequently discussed subjects across both platforms, which encompass ChatGPT's use for entertainment, its impact on search engines, software development, education, and its implications for human employment. Furthermore, our study has uncovered new, extensively discussed topics centering on responses from DAN, the Alter Ego of ChatGPT, ethical data usage of AI companies and queries about AI-generated

images. The topic on DAN underscores a potential market opportunity for a playful AI Chatbot model, as users demonstrate enthusiasm for the carefree nature of DAN. Nonetheless, this enthusiasm is also accompanied by concerns related to AI chatbots' adherence to ethical guidelines and regulations, as well as apprehensions about potential undesirable consequences in the future. Notably, the concern of AI chatbots like ChatGPT replacing human jobs emerges as the predominant topic in our study, and this concern appears to be amplified, particularly in light of significant layoffs by tech industry giants. Consequently, future research endeavors may benefit from delving into empirical evidence to address the question: Is ChatGPT contributing to workforce reductions, and if so, to what extent and in which job domains?

6.3 SWOT Analysis

In our comparative analysis of our SWOT analysis results with the previous study by Farrokhnia et al. [21], we found different findings because the scope and focus of both studies are different. While the previous research exclusively conducted a SWOT analysis within an educational context, drawing primarily from existing literature, our study adopted a broader perspective, encompassing all user opinions and interactions with ChatGPT. With the purpose of product development and policy recommendations, our SWOT analysis of the user opinions uncovered unique strengths, such as ChatGPT's coding performance, content generation capabilities, and its ability to provide entertainment, including poems, stories, and jokes. In this regard, the previous study mentions generating plausible responses, self-improving capability, providing personalized responses, and providing real-time responses. Additionally, our findings highlighted critical weaknesses, such as concerns related to ethical compliance, occasional production of incorrect or biased information, and the inability to provide real-time updates. The previous analysis mentions a lack of deep understanding, difficulty in evaluating the quality of responses, the rise of bias and discrimination, and the lack of higher-order thinking skills. Furthermore, our study identified opportunities, including the potential for ChatGPT to play a role in new job areas and specialize in playful interactions, offering innovative applications beyond educational settings. In contrast, the previous study's focus on education emphasized opportunities associated with information accessibility, personalized learning, and reduced teaching workloads. Finally, our analysis revealed unique threats in the form of potential job displacement by AI, concerns about students' overreliance on ChatGPT hindering essential skill development, and the risk of data breaches and ethical violations. These threats were not explicitly addressed in the previous study, where the primary focus was on academic integrity, discrimination, plagiarism, and the decline of higher-order cognitive skills within the educational context. As SWOT analysis can provide distinct insights that vary depending on the research's specific focus, future SWOT analysis studies on ChatGPT may choose to concentrate exclusively on diverse domains.

6.4 Implications and Limitations of Research

Our research has contributed to filling important gaps in the existing literature. While prior studies on user opinion analysis of ChatGPT have primarily focused on Twitter for a short period, we have expanded the scope and the timeframe by examining Reddit discussions collected over one-year period, thus providing a more extensive examination of user opinions on different social media platforms. Since our

research has also incorporated the SWOT analysis of user opinions, it leads to a more comprehensive understanding of the strengths, weaknesses, opportunities, and threats perceived by users, which is invaluable for strategic decision-making in both product development and policy formulation.

It is also important to note that our research may come with these potential limitations. First, the study covers a one-year period from December 2022 to December 2023. Public sentiments and opinions can evolve over time, and the findings might not capture long-term trends or shifts in user perspectives beyond this timeframe. Moreover, the study relies exclusively on data from Reddit comments, hence limiting to the opinion of the Reddit community. In addition, we have chosen to employ tools such as LDA topic modelling and VADER for our analysis. In this regard, it's important to acknowledge that both tools may come with certain limitations, in capturing the full complexity of human discussions.

7. CONCLUSION

This research primarily employs LDA topic modelling, sentiment analysis as well as SWOT analysis to examine discussions within the Reddit community concerning various aspects of ChatGPT. These discussions encompassed a wide range of topics, including comparisons between ChatGPT and conventional search engines, its potential implications for software development, its effects on the job market and education sector, examinations of ChatGPT's responses to entertainment and politics, reactions to "Dan", ChatGPT's alter ego, comments on the ethical use of data by AI companies as well as the discussions related to AI-generated images. Overall, the sentiments expressed by individuals varied, with many expressing admirations for ChatGPT's capabilities and showing enthusiasm for the future advancements of AI chatbots. They envisioned positive impacts on software development, entertainment, and the education industry. However, a minority expressed negative sentiments, particularly regarding the current limitations and potential adverse effects of ChatGPT, especially in education and job market. Therefore, future research areas may consider incorporating quantitative analysis to further explore relationships and patterns in this context.

Based on a SWOT analysis of the findings, recommendations can be made for both product development and policy implementation. For ChatGPT's product development, the focus should be on leveraging its strengths, such as its ability to retain conversation context and generate code. Additionally, efforts should be directed towards enhancing its proficiency in entertainment and interactive experiences. Addressing concerns related to "DAN", ChatGPT's alter ego, is also of importance. Simultaneously, there is evident market potential for a playful AI model for entertainment purposes and advanced AI-driven solutions to simplify daily tasks. Policy makers should prioritize addressing potential threats posed by AI products, particularly in terms of their impact on human jobs. Policies should aim to create job opportunities where AI influence is limited and transform potential threats into opportunities. Given ChatGPT's significant potential to support education, it is crucial to ensure its ethical use through the establishment of standards and guidelines in this domain. Lastly, it is essential to promptly address and engage in discussions regarding the ethical standards and guidelines that AI companies must adhere to concerning the collection of user data and the prevention of AI technology misuse, leading to adverse consequences.

AUTHOR CONTRIBUTION

This research was conducted by the first author, Shwe Zin Su Naing (shwezinsunaing_s@cmu.ac.th) and the corresponding author, Dr.Piyachat Udomwong (piyachat.u@cmu.ac.th). The entire process, from conceptualization to data analysis and manuscript preparation, was carried out by both authors.

DATA AVAILABILITY STATEMENT

The dataset used in this research, known as the “ChatGPT Reddit Comments Dataset”, was created specifically for the analysis of public opinions on ChatGPT. It was collected and compiled by the researcher and can be accessed at the following link: <https://github.com/shwezinsunaing/Reddit-Comments-Dataset.git>.

REFERENCES

- [1] Sharkey, N.: The programmable robot of ancient Greece. Available at: <https://www.newscientist.com/article/mg19526111-600-the-programmable-robot-of-ancient-greece/> (2007). Accessed 16 September 2023
- [2] Shashkevich, A.: Stanford researcher examines earliest concepts of artificial intelligence, robots in ancient myths. Available at: <https://news.stanford.edu/2019/02/28/ancient-myths-reveal-early-fantasies-artificial-life/> (2019). Accessed 16 September 2023
- [3] Introducing ChatGPT. Available at: <https://openai.com/blog/chatgpt/> (2022). Accessed 16 September 2023
- [4] Shewale, R.: 32 Detailed ChatGPT Statistics — Users, Revenue and Trends. Available at: <https://www.demandsage.com/chatgpt-statistics/> (2023). Accessed 16 September 2023
- [5] Chatterjee, J., & Dethlefs, N.: This new conversational AI model can be your friend, philosopher, and guide and even your worst enemy. *Patterns*, 4(1), (2023). <https://pubmed.ncbi.nlm.nih.gov/36699746/>
- [6] Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., ... Tseng, V. (2022). Performance of chatgpt on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health*, 2(2), (2022). <https://doi.org/10.1371/journal.pdig.0000198>
- [7] Zhai, X.: ChatGPT user experience: Implications for education. *SSRN Electronic Journal* (2022). <https://ssrn.com/abstract=4312418>
- [8] Jiao, W., Wang, W., Huang, J., Wang, X., & Tu, Z.: Is ChatGPT A Good Translator? Yes, With GPT-4 As The Engine. *arXiv preprint arXiv:2301.08745* (2023)
- [9] Qin, C., Zhang, A., Zhang, Z., Chen, J., Yasunaga, M., & Yang, D.: Is ChatGPT a General-Purpose Natural Language Processing Task Solver? *arXiv preprint arXiv:2302.06476* (2023)
- [10] Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., ... Fung, P.: A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity. *arXiv preprint arXiv:2302.04023* (2023)
- [11] Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., ... Wu, Y.: How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection. *arXiv preprint arXiv:2301.07597* (2023)
- [12] He, X., Shen, X., Chen, Z., Backes, M., & Zhang, Y.: MGTBench: Benchmarking Machine-Generated Text Detection. *arXiv preprint arXiv:2303.14822* (2023)
- [13] Mitchell, E., Lee, Y., Khazatsky, A., Manning, C. D., & Finn, C.: DetectGPT: Zero-Shot Machine-Generated Text Detection using Probability Curvature. *arXiv preprint arXiv:2301.11305* (2023)

- [14] Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T.: Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers (2022). <https://doi.org/10.1101/2022.12.23.521610>
- [15] Thorp, H. H.: ChatGPT is fun, but not an author. *Science*, 379(6630), 313–313 (2023). <https://www.science.org/doi/10.1126/science.adg7879>
- [16] NBCNews: ChatGPT banned from New York City public schools’ devices and networks. Available at: <https://www.nbcnews.com/tech/tech-news/new-york-city-public-schools-ban-chatgpt-devices-networks-rcna64446> (2023). Accessed 16 September 2023
- [17] GuardianNews: South Australian universities to allow use of artificial intelligence in assignments, if disclosed. Available at: <https://www.theguardian.com/australia-news/2023/jan/21/south-australian-universities-to-allow-use-of-artificial-intelligence-in-assignments-if-disclosed> (2023). Accessed 16 September 2023
- [18] Castelvechi, D.: Are ChatGPT and AlphaCode going to replace programmers? Available at: <https://www.nature.com/articles/d41586-022-04383-z> (2022). Accessed 16 September 2023
- [19] Stokel-Walker, C.: ChatGPT listed as author on research papers: many scientists disapprove. Available at: <https://www.nature.com/articles/d41586-023-00107-z> (2023). Accessed 16 September 2023
- [20] Susnjak, T.: ChatGPT: The End of Online Exam Integrity? arXiv preprint arXiv:2212.09292 (2022)
- [21] Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A.: A SWOT analysis of CHATGPT: Implications for educational practice and Research. *Innovations in Education and Teaching International*, 1–15 (2023). <https://doi.org/10.1080/14703297.2023.2195846>
- [22] Munggaran, J. P., Alhafidz, A. A., Taqy, M., Agustini, D. A. R., & M. Munawir.: Sentiment Analysis of Twitter Users’ Opinion Data Regarding the Use of ChatGPT in Education. *Journal of Computer Engineering, Electronics and Technology*, 2(2), (2023). <https://ejournal.upi.edu/index.php/COELITE/article/view/59645>
- [23] Haque, M. U., Dharmadasa, I., Sworna, Z. T., Rajapakse, R. N., & Ahmad, H.: I think this is the most disruptive technology: Exploring Sentiments of ChatGPT Early Adopters using Twitter Data. arXiv preprint arXiv:2212.05856 (2022)
- [24] Korkmaz, A., Aktürk, C., & Talan, T.: Analyzing the user’s sentiments of chatgpt using Twitter data. *Iraqi Journal for Computer Science and Mathematics*, 202–214(2023). <https://doi.org/10.52866/ijcsm.2023.02.02.018>
- [25] Taecharungroj, V.: “What can chatgpt do?” analyzing early reactions to the innovative AI chatbot on Twitter. *Big Data and Cognitive Computing*, 7(1), 35, (2023). <https://doi.org/10.3390/bdcc7010035>
- [26] Li, L., Ma, Z., Fan, L., Lee, S., Yu, H., & Hemphill, L.: ChatGPT in education: A discourse analysis of worries and concerns on social media. arXiv preprint arXiv:2305.02201 (2023)
- [27] Haensch, A. C., Ball, S., Herklotz, M., & Kreuter, F.: Seeing ChatGPT Through Students’ Eyes: An Analysis of TikTok Data. arXiv preprint arXiv:2303.05349 (2023)
- [28] Feng, Y., Poralla, P., Dash, S., Li, K., Desai, V., & Qiu, M.: The impact of ChatGPT on streaming media: A crowdsourced and data-driven analysis using twitter and reddit. In 2023 IEEE 9th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), 222–227 (2023)
- [29] Heimerl, F., Lohmann, S., Lange, S., Ertl, T.: Word cloud explorer: text analytics based on word clouds. In: 2014 47th Hawaii inter-national conference on system sciences, 1833–1842 (2014)
- [30] Tong, Z., & Zhang, H.: A text mining research based on LDA Topic Modelling. *Computer Science & Information Technology* (2016). <https://www.airccj.org/CSCP/vol6/csit65316.pdf>
- [31] Hutto, C., Gilbert, E.: Vader: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the international AAAI (2014)

- [32] Sarsby, A.: SWOT Analysis: A guide to SWOT for business studies student. Available at: https://books.google.com/books/about/SWOT_Analysis.html?id=Yrp3DQAAQBAJ (2016) Accessed 16 September 2023
- [33] Martindale, J.: These are the countries where ChatGPT is currently banned. Available at: <https://www.digitaltrends.com/computing/these-countries-chatgpt-banned/> (2023). Accessed 16 September 2023
- [34] Li, L.H, Zhang, H. P., Li, C.J. et al.: Evaluation on ChatGPT for Chinese language understanding. *Data Intelligence* 2023; 5(4): 885–903

AUTHOR BIOGRAPHY



Shwe Zin Su Naing, is a graduate student, pursuing a master's in digital Innovation and Financial Technology at Chiang Mai University, Thailand. Her research interest revolves around sentiment analysis, social media data analysis, machine learning, market research and developing ML-based predictive models across various fields.



Dr. Piyachat Udomwong is a lecturer at the International College of Digital Innovation, Chiang Mai University, Thailand. She earned her Ph.D. in Biology (Bioinformatics) from University of York, UK with a background in mathematics who is happy to work in multidisciplinary settings. Her research interests are to discover insights from a massive amount of data coming from different fields by using data analytics, machine learning, and statistical learning techniques.