EAI Endorsed Transactions
on Internet of Things
Research Article EAI.EU

A Multifaceted Approach at Discerning Redditors Feelings Towards ChatGPT

Shreyansh Padarha[1,*], S Vijaylakshmi[1]

[1] CHRIST (Deemed to be University) Pune Lavas, India

Abstract

Generative AI platforms like ChatGPT have leapfrogged in terms of technological advancements. Traditional methods of scrutiny are not enough for assessing their technological efficacy. Understanding public sentiment and feelings towards ChatGPT is crucial for pre-empting the technology's longevity and impact while also providing a silhouette of human psychology. Social media platforms have seen tremendous growth in recent years, resulting in a surge of user-generated content. Among these platforms, Reddit stands out as a forum for users to engage in discussions on various topics, including Generative Artificial Intelligence (GAI) and chatbots. Traditional pedagogy for social media sentiment analysis and opinion mining are time consuming and resource heavy, while lacking representation. This paper provides a novice multifrontal approach that utilises and integrates various techniques for better results. The data collection and preparation are done through the Reddit API in tandem with multi-stage weighted and stratified sampling. NLP (Natural Language processing) techniques encompassing LDA (Latent Dirichlet Allocation), Topic modelling, STM (Structured Topic Modelling), sentiment analysis and emotional analysis using RoBERTa are deployed for opinion mining. To verify, substantiate and scrutinise all variables in the dataset, multiple hypothesises are tested using ANOVA, T-tests, Kruskal–Wallis test, Chi-Square Test and Mann–Whitney U test. The study provides a novel contribution to the growing literature on social media sentiment analysis and has significant new implications for discerning user experience and engagement with AI chatbots like ChatGPT.

Keywords: Opinion Mining, Topic Modelling, Generative AI, Multi-Stage Sampling, Multiple Hypothesis Testing

Received on 02 11 2024, accepted on 30 05 2024, published on 28 06 2024

doi: 10.4108/eetiot.6447

1. Introduction

In recent years, the popularity of social media platforms has grown exponentially, and with it, the volume of user-generated content [1]. Among these platforms, Reddit stands out as a forum where users engage in discussions on a wide range of topics, including Artificial Intelligence (AI) and Natural Language Processing (NLP) [2]. One such AI-based application that has received a lot of attention on Reddit is ChatGPT. ChatGPT is an advanced chatbot that uses NLP to converse with users [3]. However, there is a need to better understand the feelings and attitudes of Reddit users (Redditors) towards ChatGPT. The outright prohibition of ChatGPT usage in Italy, an advanced economy highlights the need to differentiate between traditional AI models and Large Generative AI Models (LGAIMs), such as ChatGPT, due to their distinct technical capabilities and potential impact on society. There is a lack of representative scrutiny into people's opinion towards ChatGPT.

The purpose of this study is to introduce a comprehensive approach that integrates multi-stage sampling, statistical analysis, and natural language processing (NLP) techniques for analysing Redditors' feelings towards ChatGPT (Figure 2). To achieve this goal, the study makes use of Reddit's API to collect relevant posts and comments related to ChatGPT, followed by a

*Corresponding author. Email: shreyansh.padarha@outlook.com

comprehensive multi-stage sampling, involving weighted, stratified, and simple random sampling, to ensure a representative sample of the Reddit data is obtained. Sentiment Analysis, Emotion Analysis and LDA Structured Topic modelling are used in the study for opinion mining. Multiple hypothesis testing and statistical tests are used for reaffirmation of inferences. The study successfully obtained valuable insights into Reddit users' attitudes and sentiments towards ChatGPT, shedding light on factors that influence these attitudes. The study contributes to social media sentiment analysis and has important implications for enhancing user experience and engagement with AI chatbots.

2. Background

Artificial Intelligence (AI) is increasingly being adopted in various areas of personal and professional life alike [4,5]. But it's important to note that people's overall perceptions of AI can significantly impact their willingness to accept and use AI. A study on "The validation of the General Attitudes towards Artificial Intelligence Scale" was conducted by Astrid Schepman and Paul Rodway (2020), which showed that attitudes towards AI varies from traditional technology acceptance. Comfort level and proficiency in specific AI applications were assessed, revealing that AI for big data is perceived more positively than AI for complex human judgments. Ethical considerations were found to affect attitudes towards AI. The study highlighted the importance of considering attitudes towards AI when introducing it in different domains [6]. Generative AI models, such as ChatGPT or Stable Diffusion, are distinct from traditional AI models and are transforming the way humans communicate and create. Until recently, AI regulations had largely focused on traditional AI models, leaving a gap in regulating generative AI [7].

2.1. Web Scraping: Reliable Source of Data Collection

Since the 1970s, budgetary constraints have caused National Science Organisations to reduce manpower [8]. At the same time, the possible data being generated has exponentially risen, especially through the 2010s [9]. The basic understanding is that effectiveness and efficiency are at the core of any data collection process within research studies [8]. In the modern contemporary world, web scraping is an extensively used tool for exploiting the same. In recent years, the use of web scraping has increased significantly due to its ability to collect vast amounts of data quickly and inexpensively [10]. Scraping data from the internet for any study has its pros and cons, which must be weighed diligently. Some advantages of web-scraping and capturing data on reddit include, having a vast pool of data to collect samples from, bigger range, better accessibility and a more time and cost-efficient data acquisition methodology [11]. The limitations include technical blockages, lack of data quality assurance, and most importantly legal and ethical constraints [12]. In the case of Reddit, the platform offers an API and web scraping helpers that make it easier to collect data from the site. This can especially be useful for collecting large amounts of comments related to a specific topic, such as in the case of studying Redditors' feelings towards ChatGPT. Using the Reddit API or web scraping helpers can also allow for the collection of data in real time, which can be valuable for time-sensitive research [10,13].

2.2. Multistage Sampling Techniques

Social media data is said to suffer from polarisation, segregation, and algorithmic bias [14–16]. This presents challenges in transforming and selecting representative data for sampling from a social media platform [17]. This prompted the usage of Multi-Stage Sampling in this study. Multi-stage sampling is a sampling technique that involves selecting samples in stages, where each stage involves a different sampling method. This technique is particularly useful when the population is large and heterogeneous, making it impractical to select a simple random sample. Instead, multi-stage sampling can be used to retrieve a representative sample of the population by selecting samples at different levels of granularity [18–20].

Multi-Staged sampling techniques have been pivotal in improving efficiency of various scientific studies such as Monte-Carlo Computation Algorithms, Spatial Population Estimations and in medical surveys [21–23]. There have been studies conducted in the field of computer vision using multi-stage sampling, for video and image classification [24,25]. On the other hand, for NLP tasks, there are almost zero studies done using Multi-Stage Sampling, for text classification, sentiment analysis or topic modelling. The types of sampling being employed are, but not limited to, stratified, weighted, systematic, proportional, simple random sampling [26]. This research paper will make use of weighted, stratified, and simple random sampling for getting representative sentiments of people towards ChatGPT. Weighted sampling and Stratified sampling, have previously been used in individual NLP based studies like Relational Semantics classification and Cross Lingual Entity alignment, yielding successful results [27,28].

2.2.1. Weighted Sampling

Weighted sampling is a sampling technique that involves assigning weights to each member of a population based on some criterion [29,30]. Weighted Sampling is also called probability proportional to size (PPS) sampling or Weighted Random Sampling. It is a type of sampling in which each item in the population is assigned a weight proportional to its size, and then items are selected randomly with probabilities proportional to their weights [31].

$$P(i) = \frac{W_i}{\sum_{i=1}^{n} W_i} \quad (1)$$

In the above equation, $P(i)$ represents the probability of selecting the ith item, $W(i)$ is the weight assigned to the ith item, n is the total number of items, and the denominator $\sum_{i=1}^{n} W_i$ represents the sum of all the weights.

After calculating the probabilities $P(i)$, a random sample of size k is selected from the population of n items. This can be accomplished using any appropriate sampling technique, such as simple random sampling or stratified sampling.

2.2.2. Stratified Sampling

Stratified sampling is a probability sampling technique that involves dividing the population into subgroups, or strata, and selecting a sample from each stratum. Elements within each subgroup or stratum are similar in nature. The technique is used in tandem with simple random sampling frameworks [32].

For each stratum h with size N_h in the population of size N, a sample of size n_h is selected, where the sample size for stratum h is proportional to the size of stratum h in the population and n is the desired sample size [18]:

$$n_h = \frac{n \cdot N_h}{N} \quad (2)$$

The ideal stratum size has various formulas, but the most common one is:

$$stratum\_size = \frac{sample\_size \cdot length(stratum\_data)}{length(data)} \quad (3)$$

where $stratum\_data$ is the data in the current stratum, and $data$ is the entire dataset.

2.3. NLP (Natural Language Processing) Techniques in Opinion Analysis

2.3.1. Sentiment Analysis

Sentiment analysis is a popular NLP technique used to determine the sentiment of text, which can be useful for understanding the feelings and opinions of a large group of people [33]. Sentiment analysis is a tried and tested mechanism for classifying sentiments, into positive, negative, and neutral, which has been applied on "billions" of web pages [34]. Sentiment analysis has been successfully used in multiple studies across various social media platforms, including reddit, twitter and Facebook amongst many [35–37].

RoBERTa Model: For calculating the sentiment analysis scores, a model is needed for accurate predictions. RoBERTa (Robustly Optimised BERT Pre-training Approach) architecture is a better version of the Bert model, which is a transformer-based model [38]. Research has shown that training with larger batches and longer sequences can result in better generalisation of data [39]. RoBERTa is designed to better capture the contextual relationships between words, which is especially useful for sentiment analysis on social media platforms like Reddit where slang and informal language are prevalent [40].

VADER is another commonly used Sentiment Analysis tool for social media text, considered as an easier lexicon rule-based classifier [41]. Some studies also use VADER to predict labels and verify it through BERT and RoBERTa accuracy scores [42]. In case, a dataset exists that is suited for the problem at hand, RoBERTa can be used directly for better results.

2.3.2. Topic Modelling

Topic modelling techniques are highly effective methods extensively used in the field of natural language processing for uncovering topics and extracting semantic meaning from unordered documents [43]. The goal is to automatically group similar documents together into clusters or topics based on the shared characteristics of their content. The algorithm accomplishes this by analysing the frequency and co-occurrence of words across documents and then creating a set of topics that represents the most significant recurring patterns in the text [44]. Topic modelling studies do not provide a full meaning of the text, but they provide a holistic overview of the themes and categories, that would have been missed otherwise [45].

LDA (Latent Dirichlet Allocation) is an extensively used topic modelling technique. It has been used for opinion mining, text classification and emotion classification [46]. LDA is a generative probabilistic model, which is implemented on collections of discrete data [43] and is based on the exchangeability assumption for words and topics in a corpus, an application of De Finetti's exchangeability and representation theorem [47].

The LDA model assumes that the words of each document arise from a mixture of topics, each of which is a distribution over the vocabulary. A limitation of LDA is the inability to model topic correlation even though, for example, a document about nuclear radiation is more likely to also be about nuclear disasters than gamma radiation. This limitation comes from modelling variability of topic proportions through Dirichlet distribution. For this purpose, CTM (Correlated Topic Modelling) was created, which relies on logistic normal distribution [48]. This was further refined and tweaked into STM (Structured Topic Modelling), a tool developed especially for use in social sciences to incorporate corpus structure and document meta-data like political affiliation into the topic modelling pedagogy [49].

The model (Figure 1) integrates three existing models: CTM, DMR, and SAGE. It replaces the logistic normal prior with a logistic-normal linear model and employs a multinomial logit for word distribution. Regularizing priors for GLM coefficients are provided in the software [49].

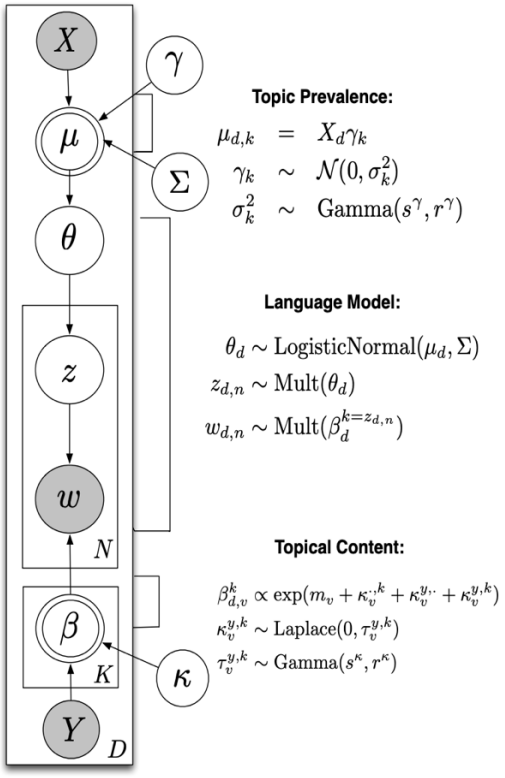


Figure 1. Plate Diagram for the Structural Topic Model

2.4. Statistical tests for NLP based Studies

Statistical significance has been deemed an extremely important metric to judge NLP based tasks [50]. There are many protocols on choosing the right statistical tests, F1 score, Pearson Correlation, or hypothesis tests, for re-assessing coincidental affirmations in NLP surveys [51]. Chi-square tests have been used within sentiment analysis for feature selection [52]. Similarly, tests have been deployed for NLP tasks. But there are no social media corpus-based studies that combine the effective nature of sentiment compound score, with other attributes prevalent within online comments or reviews datasets. There are many statistical tests that can be used for providing insights into the relationship between sentiment and other interactive features like an upvote button or dislike button. For example, ANOVA and Kruskal Wallis, tests can be implemented for their robust stochastic homogeneity [53].

2.4.1. Multiple Hypothesis Tests

Global hypothesis tests play a vital role in clinical trials, genetic studies, or meta-analyses by enabling researchers to assess whether none of the hypotheses being examined are false, rather than specifically testing individual hypotheses [54]. They are deemed status-quo, while conducting scientific studies. But all these cases listed have a binary end- result, which is substantiated by the rejection or acceptance of the null hypothesis. In wider, open-ended problems, multiple hypotheses can be used to gauge a broader statistical affirmation. *Ceteris paribus* multiple inferences or hypotheses involve simultaneous single hypothesis tests that are weighed subjectively or objectively, in accordance with the study, and a single or multiple outcomes can be attained [55].

3. Proposed Methodology

The research methodology proposed in this study aims at improving user representatives and mitigating high computational power limitations that arise with social media opinion mining. The proposed process involves collecting data in an intricate manner, multistage sampling, opinion mining using NLP techniques and validation through multiple hypothesis testing (Figure 2).

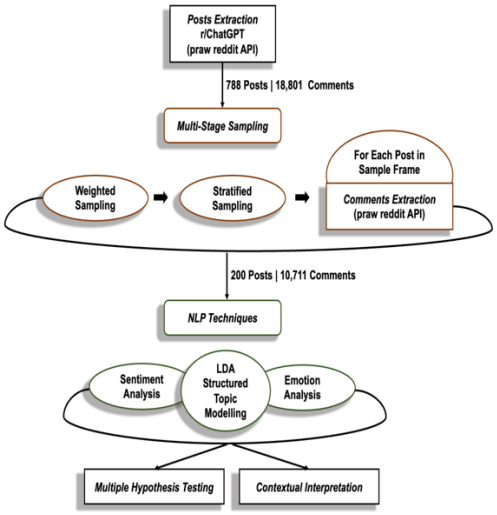


Figure 2. Study Flowchart

3.1. Data Collection Through Reddit API

The Praw Python library was used for calling the Reddit API for requesting the objects within the HTML pages. The r/ChatGPT Community was selected as the population frame for its potential representatives of the bulk of Redditors. As of June 2023, the community had 2.1 million active users and ranked within the top 1% of the Reddit communities, by size.

3.1.1. Posts Dataset

Initially, all posts posted within the r/ChatGPT community were scrapped, that enumerated the sample frame size to 788 posts. Table 1a shows a sample instance, features and values, from the posts' dataset. Information in line with the attributes are readily available in the reddit documentation [13]. The dataset collected was sparse and heterogeneous. Upon applying multi-stage sampling, the posts in consideration decreased from 787 to 200.

Table 1. Sample Instances from Reddit Datasets

(a) Posts Dataset

Feature	Value
Id	12tdpvb'
Title	"Chat GPT doesn't work no matter what I do. No VPN. Works on phone using the same connection. Any suggestions?"
Author	lockeslay'
Created	2023-04-21
URL	https://i.redd.it/ru2xkl78k3va1.png'
Score	0
Comments	1
Flair	Other'
CSS Class	None
Upvote Ratio	0.33
Over 18	FALSE
Tag	Other

(b) Comments Dataset

Feature	Value
post_id	12r69k8
post_title	"Microsoft readies its own AI chip to reduce Nvidia reliance. ChatGPT....."
post_num_cmts	241
post_upvote_ratio	0.98
post_tag	Other
comment_id	jguyph4
comment_body	"Yeh but that isn't a $50 million cheque. That's several smaller ones..."
comment_score	-1
is_comment	FALSE

3.1.2. Comments Dataset

Comments and replies within each of the 200 representative posts were retrieved, amassing to 10,711 comments. Scrapping 10,711 comments, involved considerable amount of computational time on High-performance CPU. Table 1b shows a sample instance from the comment's dataset. The comments dataset was a crucial study element that is used for further analysis, including sentiment analysis, hypothesis testing and topic modelling, within the study.

3.2. Multi-Stage Sampling of Posts

Multi-stage sampling was used for selecting a representative subset of posts and comments from the r/ChatGPT community on reddit. Weighted sampling and subsequent stratified sampling were used for the study.

3.2.1. Weighted Sampling

The study considered the number of comments and upvote ratio as the weighted variables to select subset of posts from r/ChatGPT community on Reddit. The considered variables ensured user engagement, community feedback and diversity of content in the corpus. Algorithm 1 shows the process for weighted sampling used. Both the variables are normalised, for standardising the scales. Laplacian smoothing is applied to both the variables to prevent data imbalance, avoid zero probabilities and to generate better generalisation [14].

$$P(x) = \frac{Count(x) + k}{N + k \cdot V} \quad (4)$$

The formula above incorporates $P(x)$ representing the smoothed probability of event x, $Count(x)$ is the count of observations. N represents the total number of observations, while V corresponds to the total number of unique events or outcomes (2 Variables). The smoothing parameter (k) is set to 1, for stabilising the smoothing effect.

Post smoothing of the variables, the adjusted number of comments and upvote ratios are multiplied to get one final weight. Min – Max Feature scaling is applied to the final weight, to scale the weights between 0 and 1, deriving a resultant probability of selection. The probabilities are multiplied by random numbers from 0 to 1 and the top 400 products (post units) are selected.

Algorithm 1 Weighted Sampling Algorithm

Require: upvote ratio (upv_r), number of comments (cmt_n), number of posts ($numPosts$), smoothing factor (k)

Ensure: Selection based on upvote ratio and number of comments of a post.

1: weights = [] // Defining an Empty Array

2: **for** i **in** 1 to $numPosts$ **do**

// Normalize Features

3: $cmt_n[i] \leftarrow \frac{cmt_n[i]}{\max(cmt_n)}$, $upv_r[i] \leftarrow \frac{upv_r[i]}{\max(upv_r)}$

// Laplacian Smoothing

4: $cmt_n[i] \leftarrow \frac{cmt_n[i] + k}{length(cmt_n) + k \cdot \sum(cmt_n)}$

5: $upv_r[i] \leftarrow \frac{upv_r[i] + k}{length(upv_r) + k \cdot \sum(upv_r)}$

// Weights Multiplication

6: $weights.insert(cmt_n[i] \cdot upv_r[i])$

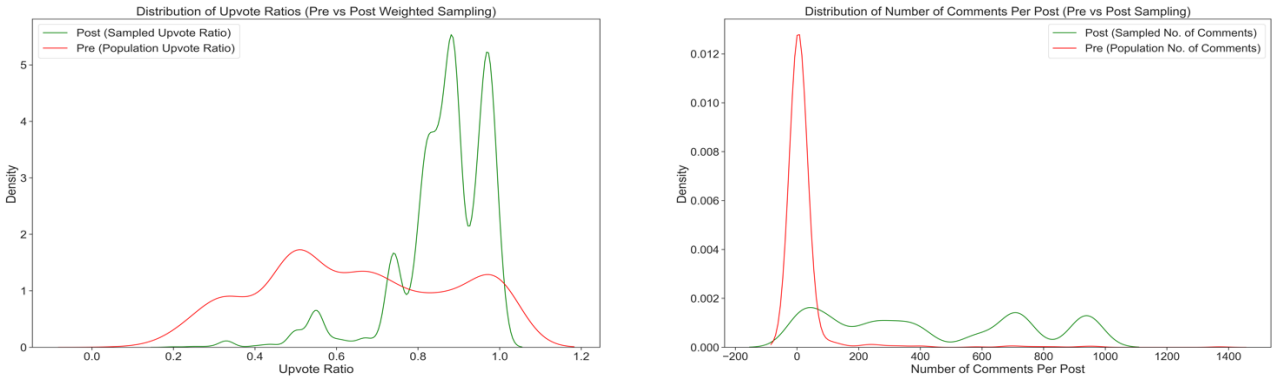
7: **End for**

// Min Max Scaling weights (probabilities)

8: $\frac{wgh_i - min(weights)}{max(weights) - min(weights)} \quad \forall \backslash wgh_i \in \backslash weights$

9: **Multiply** each probability in probabilities by a random number r ∈ [0, 1]

10: **Select** the top 400 products (post units) based on the computed probabilities P. =0



(a) Upvote Ratio **(b)** Number of Comments

Figure 3. Pre vs Post Weighted Sampling Comparison of Feature Based Distribution

Comparing the distribution of posts before and after weighted sampling (Figure 3) elicits two things. First, the posts sampled are more highly concentrated amongst the popular posts with an upvote ratio mostly between 0.7 and 1.0. Second, the posts sampled have a significantly lower empty comment units, with most posts having upwards of 100 comments each.

3.2.2. Stratified Sampling

After weighted sampling, the study involves the use of stratified sampling for selecting a subset of posts from the previous stage. As deduced from Figure 4, there is an imbalanced distribution of posts tags in the dataset. Posts are predominantly under the tags "Funny", "Other", "Serious replies only: closed-ai", "Gone Wild" and use cases. To get more representativeness from underrepresented tags, strata were formed based on shared subreddit tag attribute. A sample size of 200 posts was decided upon. The stratum size or sample size for each tag was calculated based on its proportion to the population and in line with the final sample size. Post determining the stratum size, the samples were selected randomly from the stratums.

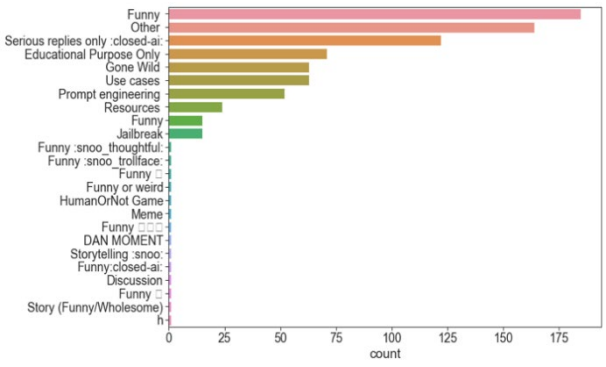


Figure 4. Number of Posts per Tag Chart

Post multi-stage sampling, the total representative comments from the 200 posts were 10,711. To better optimise resource allocation, sample units (comments and replies) were retrieved and called from the API only after stratified sampling of posts. This brought down the workload by 49%.

3.3. Applying Natural Language Processing Techniques

The filtered dataset from the multistage sampling is used through NLP techniques. The study extracted emotions and sentiments to understand Redditors' feelings towards ChatGPT, followed by structural topic modelling (STM) using LDA (Linear discriminant analysis).

3.3.1. Sentiment and Emotion Analysis

The RoBERTa base model is trained on 58 million twitter tweets, following the TweetEval benchmark, a framework developed by Cardiff University Computer Science graduates [56]. TweetEval's emotion and sentiment models were used for the study. HuggingFace's transformers are used for tokenizing and initialising the models [57]. The data is pre-processed for punctuations, stopwords, white spaces using NLTK and regular expressions. The sentiment analysis model used in the study generated separate positive, negative, and neutral scores between 0 and 1 for each comment. These scores were then combined using a compound scoring mechanism to generate an overall sentiment score for each comment. The compound score involved, adding weights, scaling the values, and normalising the final value between -1 and 1. The final sentiment awarded to each comment was based on threshold values, that segregated the sentiment compound score (cs) into positive ($cs > 0.1$), negative ($cs < -0.1$) and neutral ($-0.1 < cs > 0.1$) comments.

The emotion analysis model gave separate scores between 0 and 1 for joy, optimism, sadness, and anger. The highest scoring emotion was adjourned as the emotion for that comment. Table 2 shows a sample comment for each emotion and each sentiment, post-analysis.

Table 2. Sample Opinion-wise Comments

(a) Sentiment

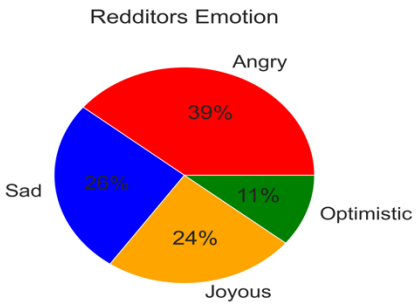
Sentiment	Comment
Positive	It's very good in spanish, almost as good as in english but only for the Spanish/Mexican variant of spanish. When I ask for Argentina spanish it only adds exaggerated idioms but words, conjugation, articles, etc. don't change
Negative	This leaves absolutely no room for students who have learning disabilities or who have exceptionalities
Neutral	It would be hilarious if it actually is developed with a decent epistemology and debunks all of his idiotic claims and beliefs.

(b) emotion

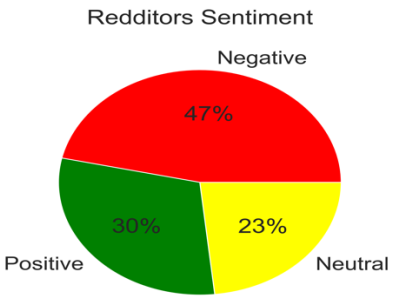
Emotion	Comment
Joyous	Wait this is nuts. You actually used the bot to *study* that's crazy
Optimistic	I can tell that most of them aren't written by ChatGPT, because most of ChatGPT's responses have good grammar and make logical sense.
Angry	Is it $20 one time or monthly? Because damn if I'd pay $20 monthly for something like a bot. :x
Sad	My god...' 'we live in a simulation????!

Figure 5a depicts the distribution of emotions, including anger, optimism, joy, and sadness, within the comments, displaying the respective percentages. The dataset has 39% angry comments, 26% sad comments, 24% Joyous comments and 11% optimistic comments.

Figure 5b provides percentages corresponding to positive, negative, and neutral sentiments constituting the comments. A majority (47%) of comments were Negative, followed by 30% Positive and 23% Neutral comments.



(a) Emotion

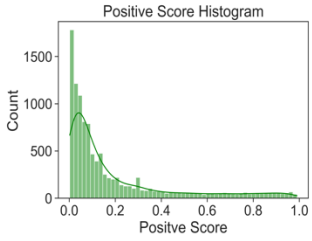


(b) Sentiment

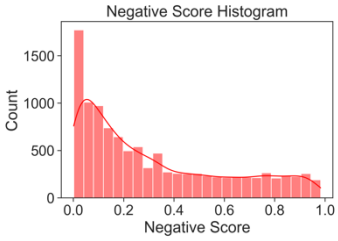
Figure 5. Overall Comments Opinion Pie Chart

The sentiment histogram (Figure 6) shows the extremity of sentiment for each comment, visualising the positivity, negativity and neutrality and overall sentiment prevalence.

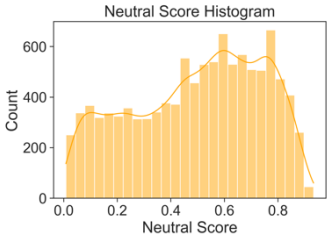
Positive score for comments is highly concentrated in the 0 to 0.25 range. The compound sentiment amongst comments is focused from -0.75 to 0.3, indicating the majority of comments have negative-neutrality as a sentiment.



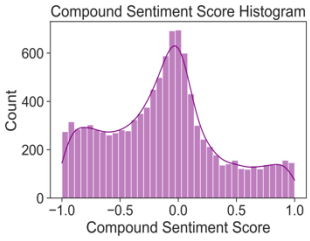
(a) Positive



(b) Negative

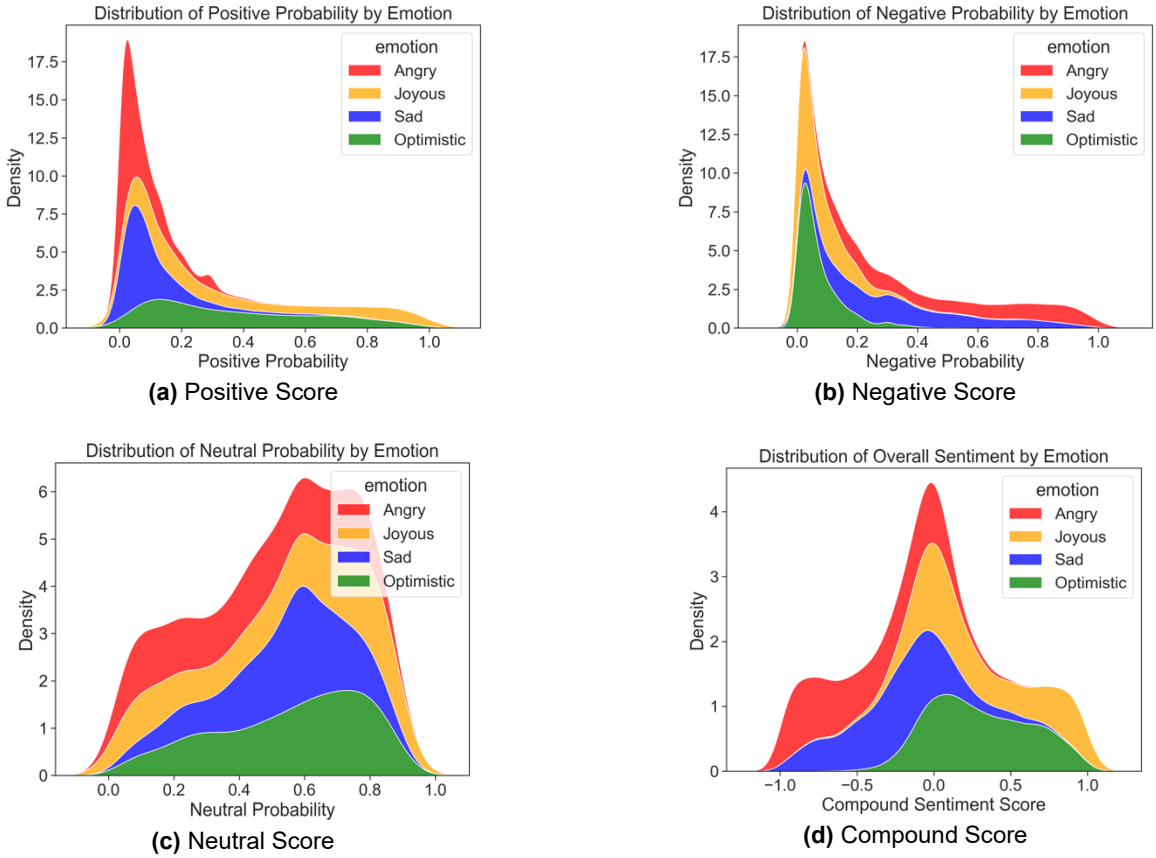


(c) Neutral



(d) Overall

Figure 6. Individual Comments Sentiment Histogram



(a) Positive Score **(b)** Negative Score

(c) Neutral Score **(d)** Compound Score

Figure 7. Sentiment Probability Kernel Distribution Estimation by Emotion

Figure 7 represents the Kernel Distribution Estimation of each sentiment by emotion. Comments with a positive probability of above 0.3 constitute optimistic and joyous emotions. Comments with a negative probability of above 0.4 comprise of sad and angry comments predominantly. Neutral probability is a mixture of all emotions. The same is persistent in the overall compound sentiment score.

3.3.2. Topic Extraction Through STM (Structured Topic Modelling)

Structured Topic Modelling (STM) was applied to the filtered dataset obtained through multi-stage sampling. The goal of STM was to identify and analyse latent topics present in the comment's dataset from the r/ChatGPT community on Reddit. The explains the approach taken for STM and the rationale behind selecting the optimal number of topics (k=5).

The dataset of comments was pre-processed using the tm and topicmodels packages in R. Pre-processing included tokenization, stop words removal and stemming. Subsequently, the documents, vocabulary, and metadata were prepared for the STM analysis. The metadata included information such as sentiment scores, compound scores, and emotion scores associated with each comment.

To determine the optimal number of topics (k) to be extracted, a search was conducted using the searchK() function in the R stm package. The search was performed for values of k ranging from 3 to 10. The prevalence of sentiment,

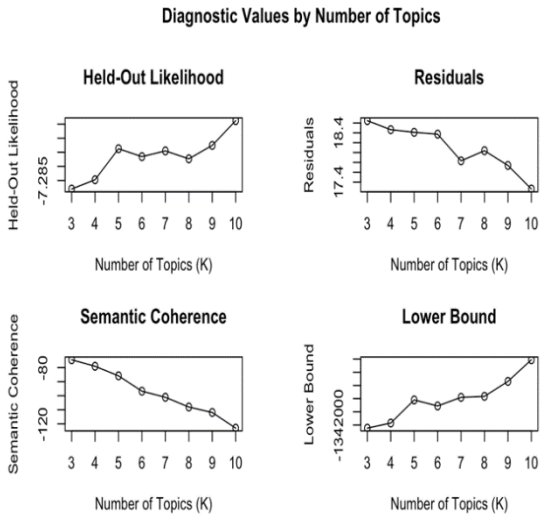


Figure 8. Diagnostic Plot for K

emotion, and compound score variables were included in the model.

The searchK() function calculated various metrics such as exclusivity, semantic coherence, held-out likelihood, residual, and bounds to evaluate the quality of the topics generated for each value of k.

Table 3. Performance Metrics of Different Values of K

K	Exclusivity	Semantic Coherence	Held-Out Likelihood	Residual	Bound	Lower Bound
3	8.279299	-74.4402	-7.288007	18.64755	-1342478	-1342476
4	9.032705	-79.04821	-7.284718	18.46417	-1341704	-1341701
5	9.067013	-85.85897	-7.273733	18.41073	-1338203	-1338199
6	9.388168	-96.7834	-7.276533	18.37297	-1339116	-1339109
7	9.487478	-101.1194	-7.274502	17.83161	-1337811	-1337803
8	9.597624	-108.0328	-7.277301	18.03736	-1337661	-1337651
9	9.579945	-111.9525	-7.272544	17.73781	-1335392	-1335379
10	9.636439	-123.0716	-7.263759	17.26053	-1332118	-1332103

Based on the results (Table 3, Figure 8) of searchK, it was observed that the model with topics (k) as 5 had a high value of exclusivity, indicating that the topics were distinct and not overlapping significantly. The semantic coherence was also high, suggesting that the topics were coherent and meaningful. The held-out likelihood and residual values were within acceptable ranges. The bounds indicated the quality of the model fit, with lower values indicating better fit. Overall, the model with k=5 demonstrated optimal performance based on these metrics.

After selecting k = 5, a final STM model was fitted using the selectModel() function. The model incorporated the sentiment, emotion, and compound score variables as predictors of the topics. The initialization of the model was set to LDA (Latent Dirichlet Allocation) and the maximum number of expectant-maximization iterations (epochs) was set to 100. The modelled topics are further analysed using various R functions for assessing topic quality, estimated effects, multiSTM() and cloud charts.

Table 4 and Figure 9 depict the topics discovered with top word occurrences, highly unique and frequent topics (frex), strongly associated words within a topic (Lift) and overall prominent topics (score). Topic 1 encompassed discussions revolving around Elon Musk's influence and controversies related to GPT, characterized by the frequent use of words such as "musk," "truth," and "openai". Topic 2 focused on job scams and interview experiences, with discussions highlighting the prevalence of words like "scam," "interview," and "bird". The usage of ChatGPT in student academics emerged as Topic 3, where conversations revolved around tasks like writing essays and exams, evident from the presence of words like "write", "student", and "essay". Topic 4 cantered on ChatGPT as a bot and prompt generator, with discussions emphasizing words such as "bot", "prompt" and "comment". Lastly, Topic 5 delved into speculations regarding the future of AI, consciousness, and advanced technologies, with frequent mentions of words like "think", "know" and "human".

Table 4. Word Insights from Topic Modelling

Topic	Top Words	FREX	Lift	Score
Topic 1: Elon Musk's Influence and Controversies	can, say, want, right, gpt, tell, make	musk, truth, openai, news, fox, upload, legal	affiliation, ballpark, bleach, catturd, childish, condescend	musk, elon, say, want, can, truth, twitter
Topic 2: Job Scams and Interview Experiences	like, just, get, people, good, realli, bird	bird, pay, scam, interview, manag, recruit, hire	-year-old, "manag, soft, aaa, adam	bird, like, job, just, get, peopl, good
Topic 3 : ChatGPT in Student Academics	use, can, chatgpt, write, learn, work, student	write, student, essay, exam, paper, written, cheat	-minut, "read", "right", ambigu, approxim, april, artifact	write, student, learn, essay, can, exam, work
Topic 4: ChatGPT as a Bot and Prompt Generator	bot, prompt, post, generat, comment, pleas, thank	bot, post, comment, pleas, thank, automat, perform	deiti, hey, psa, pun, ritual, discord, mod	bot, chatgpt-rel,supportopenaicom, ushttpsdiscordgg,
Topic 5: Speculations on AI and the Future	will, think, know, one, make, human, thing	conscious, futur, agi, technolog, bodi, intellig, quantum	acronym, aerospac, air, aisl, album, ambit, androgyn	think, know, human, one, dont, school, thing

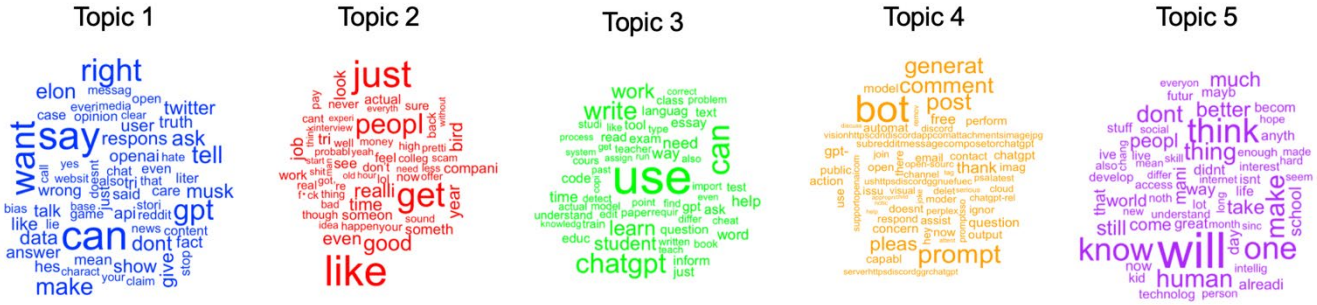


Figure 9. Topic Word Cloud

3.4. Multiple Hypothesis Testing

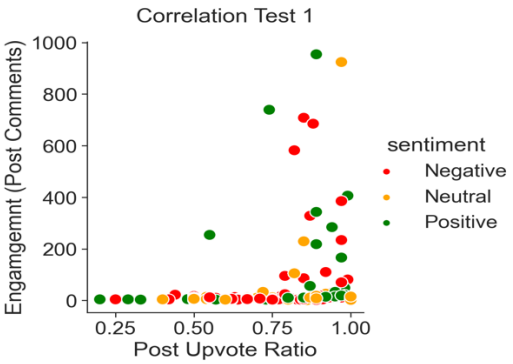
To find the significance (impact) of reddit nomenclature variables on sentiment and emotions, multiple hypothesis testing was used in the study. These tests investigated various relationships and associations between different variables.

3.4.1. Correlation tests

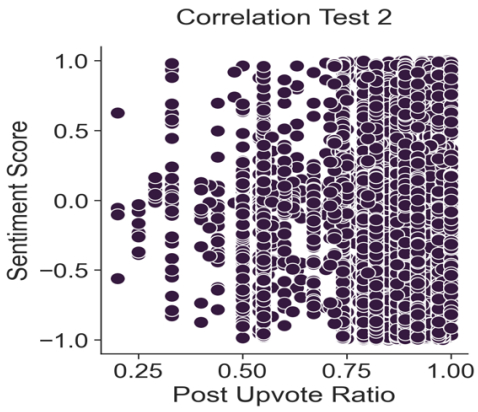
- **Test 1:** To determine whether a correlation exists between the post upvote ratio and the number of comments on a post.
- **Test 2:** To explore the relationship between the post upvote ratio and the compound score of comments.

The correlation coefficient, denoted by r, was calculated using the Pearson correlation formula.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (5)$$



(a) Post Upvote Ratio vs Redditors Engagement



(b) Post Upvote Ratio vs Sentiment Score

Figure 10. Correlation Tests Visualisation

The correlation test 1 revealed a significant positive correlation between the post upvote ratio and the number of comments on a post, with a Pearson correlation coefficient of 0.207 ($p < 0.001$), as evident in Figure 10a. The correlation test 2 indicated a weak positive correlation between the post upvote ratio and the compound score of comments, with a Pearson correlation coefficient of 0.043 ($p < 0.001$) (Figure 10b).

3.4.2. Mann-Whitney U Test

Test: To compare the compound scores of comments for posts with high and low upvote ratios, to establish difference in the compound scores between these two groups.

The Mann-Whitney U test was used for comparing two independent samples to determine if they came from the same population.

$$U = \min(U_1, U_2) \tag{6}$$

where U_1 and U_2 are the sums of ranks for the two samples.

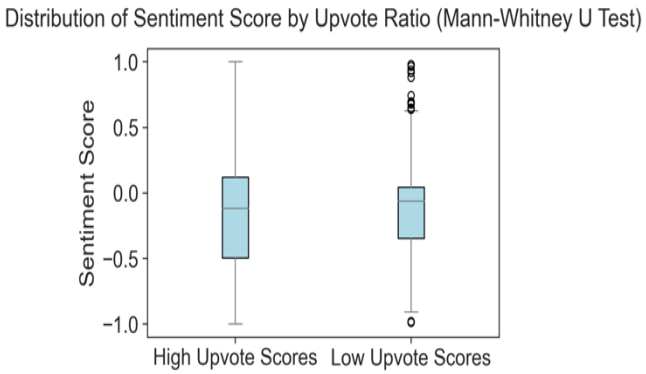


Figure 11. Mann-Whitney U test Variables Visualisation

Similar to Correlation test 2, the Mann-Whitney U test indicated no significant difference in the compound scores of comments between posts with high and low upvote ratios, with a U-statistic of 1246239 and a p-value of 0.288 (Figure 11).

3.4.3. Chi-Square Test of Independence

Test: To determine if there is an association between post tags and comment sentiment.

The Chi-Square test of independence was used to determine if there is a significant association between two categorical variables.

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{7}$$

where O_{ij} represents the observed frequencies and E_{ij} represents the expected frequencies.

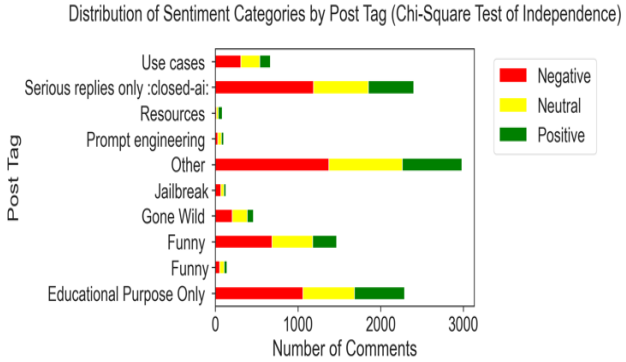


Figure 12. Chi-Square Test Variables Visualisation

The Chi-Square test of independence revealed a strong association between post tags and comment sentiment, with a chi-square statistic of 142.890 and a p-value of 1.77×10^{-21} (Figure 12).

3.4.4. Kruskal-Wallis H Test

Test: To determine the impact of tags on overall sentiment score.

The Kruskal-Wallis H test is a non-parametric test used for comparing three or more independent samples.

$$H = \frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(N+1) \tag{8}$$

where R_i is the sum of ranks for the ith sample, n_i is the size of the ith sample, N is the total sample size, and k is the number of samples.

The Kruskal-Wallis H test demonstrated that tags have a significant impact on the overall sentiment score, with a test statistic of 85.579 and a p-value of 1.25×10^{-14} (Figure 13).

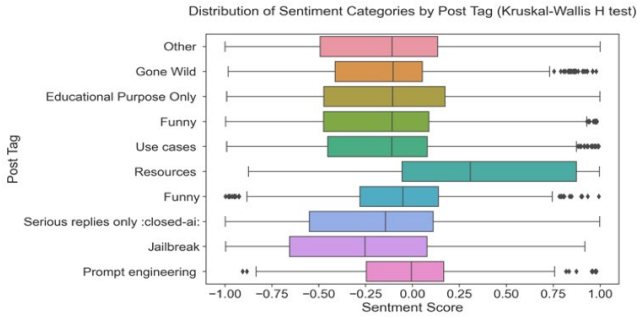


Figure 13. Kruskal-Wallis H Test Groups Box Plot Chart

3.4.5. t-tests (independent samples)

- **Test 1:** To determine whether there is a significant difference in the sentiment scores between the comments and replies in a post.

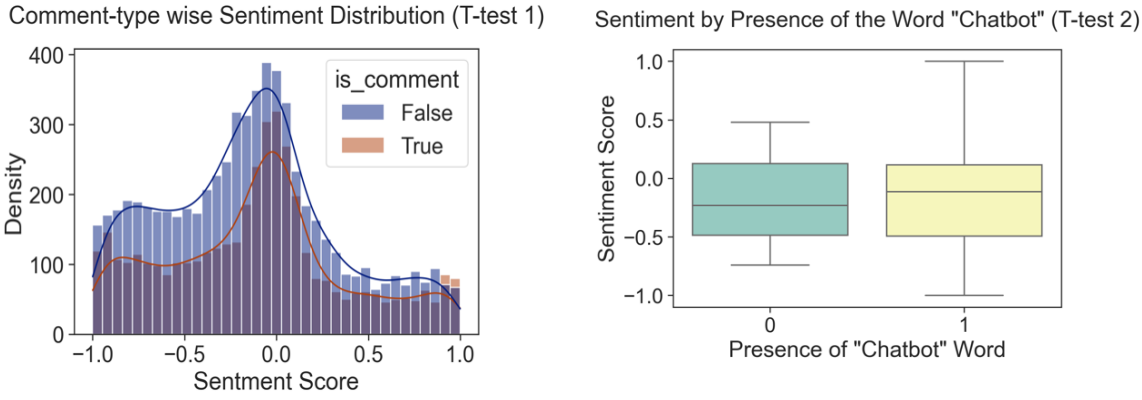
- **Test 2:** To determine the significance of the presence of the word "chatbot" in the comments on sentiment score.

The t-test is used for comparing the means of two independent samples.

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (9)$$

where X_1 and X_2 are the sample means, s_1^2 and s_2^2 are the sample variances, and n_1 and n_2 are the sample sizes for the two groups.

The independent samples t-test 1 indicated a significant difference in sentiment scores between the comments and replies in a post, with a t-statistic of 3.463 and a p-value of approximately 0.0005 (5.36×10^{-4}) (Figure 14a). The independent samples t-test 2 revealed no significant difference in sentiment scores based on the absence or presence of the word "chatbot" in the comments, with a t-statistic of -0.110 and a p-value of 0.913 (Figure 14b).



(a) Post Upvote Ratio vs Redditors Engagement

(b) Post Upvote Ratio vs Sentiment Score

Figure 14. t-tests Visualisation

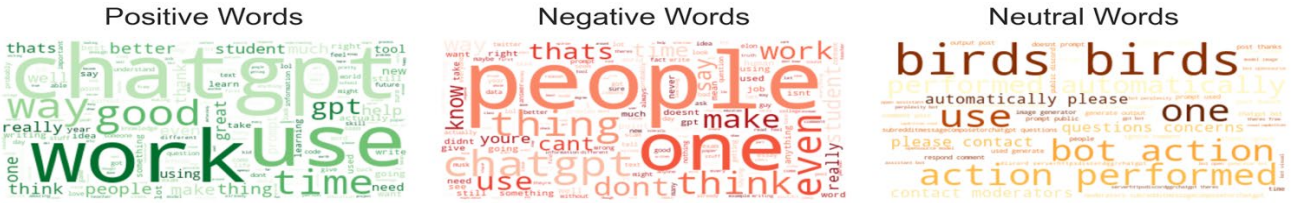


Figure 15. Sentiment-wise Word Chart

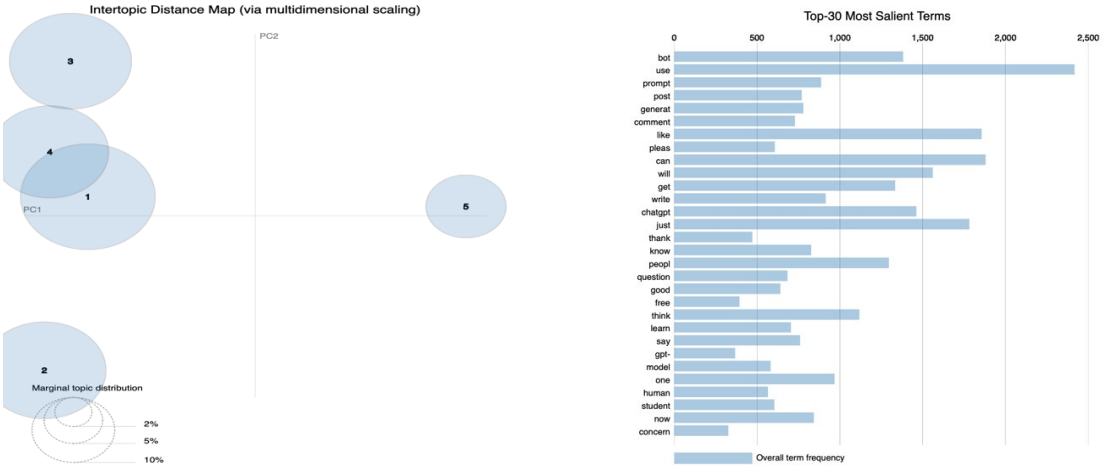
4. Result

This section presents a summary of the methodical analysis conducted in the previous section. Sentiment analysis revealed a majority of negative (47%) comments, followed by positive (30%) and neutral (23%) (Figure 5b). Emotion analysis showed that anger (39%), sadness (26%), joy (24%), and optimism (11%) were among the prominent emotions expressed in the comments (Figure 5a). Compound sentiment indicated a prevailing negative-neutrality. KDE Analysis (Figure 7) demonstrates that positive sentiment (probability > 0.3) aligns with optimism and joy, while negative sentiment (probability > 0.4) aligns with sadness and anger. Neutral sentiment includes mixed emotions, indicating psychological alignment. Figure 15 shows highest occurring words in each sentiment. The word "ChatGPT" is extensively used in both positive and negative comments. Positive sentiment comments include words like "work", "use", "good", "time", “student", while negative sentiment comments include "people", "make", "one", "don’t" and neutral comments include "bot", "action", "birds", "use".

The corpus was subjected to topic extraction using structured topic modelling (STM) in the r/ChatGPT Reddit community (Table 3, Figure 8). Five distinct topics were identified and labelled based on the top words associated with each topic (Table 4, Figure 9). Topic 1, titled "Elon Musk’s Influence and Controversies" focused on discussions related to Elon Musk and his statements about GPT, as evidenced by words like "musk," "truth," and "openai." Topic 2, named "Job Scams and Interview Experiences" revolved around conversations about fraudulent job opportunities and interview related challenges, with prominent words including "bird", "scam" and "interview". Topic 3, labelled "ChatGPT in Student Academics" centred on the usage of ChatGPT by

students for academic purposes, with prevalent words such as "write", "student" and "essay". Topic 4, "ChatGPT as a Bot and Prompt Generator" encompassed discussions about ChatGPT functioning as a bot and generating prompts and comments, reflected by words like "bot", "prompt" and "comment." Lastly, Topic 5 named "Speculations on AI and the Future" involved broader discussions about the future of AI, consciousness, and advanced technologies, characterized by words such as "think", "know", and "human".

Figure 16 visualisations were generated through LDAvis software, for the topics extracted [58]. Figure 16a shows the inter-topic distance on a 2 principal component axis, where topics 1 and 4 share the most amount of words. Topics 5 and 2 are the most distinct and unique topics. Figure 16b shows the top 30 most salient and relevant terms across all the topics. "bot", "use", "prompt", "post", "concern", "learn", "human", "student" are some notable terms in the topics.



(a) Inter Topic Distance Map

(b) Top 30 Salient & Relevant Terms

Figure 16. STM Extracted Topics Analysis

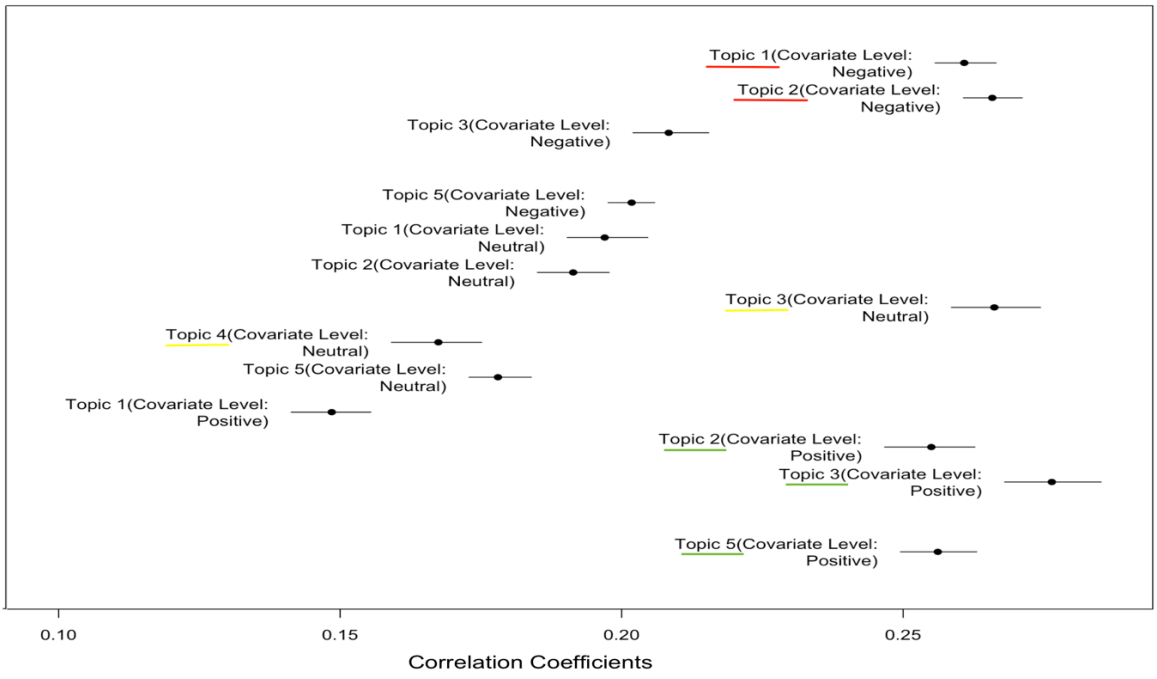


Figure 17. Topic-wise Sentiment Correlation Coefficient

The Figure 17 visualises the sentiment regression prevalence within topics, through STM. Comments related to topic 1 (Elon Musk's Influence and Controversies) are the most negative, while topic 3 comments (ChatGPT in Student

Academics) are the most positive, followed by topic 5 comments. Topic 2 comments have strong sentiment polarity. Topic 4 comments have a high level of neutrality.

Multiple hypothesis testing was conducted to assess the significance of Reddit nomenclature variables on sentiment and emotions (Table 5). Correlation tests indicated a positive correlation between upvote ratio and number of comments and a weak correlation between the upvote ratio and the sentiment associated. The Mann-Whitney U test gave similar results, backing the correlation test 2. The Chi-Square test revealed an association between post tags and comment sentiment. The Kruskal-Wallis H test demonstrated tags have an impact on overall sentiment. Independent samples t-tests showed a significant difference in sentiment scores between comments and replies, but no impact on sentiment from the word "chatbot" in comments.

Table 5. Multiple Hypothesis Testing Results

Test Name	Objective	Statistic	Statistic Value	P-value
Correlation Test 1	To see if there is any correlation between the post_upvote_ratio and the number of comments on a post.	Pearson correlation coefficient	0.207	7.97×10^{-104}
Correlation Test 2	To determine if there is a relationship between post upvote ratio and the compound score of comments	Pearson correlation coefficient	0.043	6.92×10^{-6}
Mann-Whitney U Test	To compare the compound scores of comments for posts with high and low upvote ratios.	U-Statistic	1246239	0.288
Chi-Square Test of Independence	To determine if there is an association between post tags and comment sentiment	Chi-Square Statistic	142.890	1.77×10^{-21}
Kruskal-Wallis H test	To determine the impact of tags on overall sentiment score	-	85.579	1.25×10^{-14}
t-test (independent samples) 1	To determine whether there is a significant difference in the sentiment scores between the comments and replies in a post	T-Statistic	3.463	5.36×10^{-4}
t-test (independent samples) 2	To determine the significance of the absence/presence of the word "chatbot" in the comments	T-Statistic	-0.110	0.913

5. Discussion & Conclusion

The multifaceted approach employed in this study successfully revealed valuable insights into Redditors' sentiments towards ChatGPT. The sentiment analysis uncovered majority were negative comments (47%), followed by positive (30%) and neutral (23%) sentiments. Emotion analysis further illuminated the prevalent emotions within the comments, with anger (39%), sadness (26%), joy (24%), and optimism (11%) being prominent. Sentiment analysis and opinion analysis overlapped in psychological opinion expressions. Through Structured Topic Modelling (STM) analysis, 5 distinct topics and their associated sentiments were identified. Topics such as "Elon Musk's Influence and Controversies" exhibited negative sentiments, while discussions about ChatGPT in student academics conveyed pre-dominantly positive sentiments. Comments within the topic "Speculations and future of AI" were of a progressive, yet sceptical nature, indicating hesitance to adopt the new terminology. The sentiment correlation coefficient within each topic validated these findings. The study also demonstrated the divergent nature of Redditors through the discussion of distinct and unique terms while addressing a topic. However, the study lacks the examination of opinions over time series data, thus missing the shifts and fluctuations in opinions over time. Overall, the study's findings highlight the diverse range of sentiments expressed by Redditors towards ChatGPT, contributing to a deeper understanding of user experiences and engagement with AI chatbots. The effectiveness of the proposed framework is evident in its ability to derive meaningful insights from user-generated content.

References

[1] O'Keeffe GS, Clarke-Pearson K. The Impact of Social Media on Children, Adolescents, and Families. Pediatrics 2011;127:800–4. https://doi.org/10.1542/peds.2011-0054.

[2] Thukral S, Meisheri H, Kataria T, Agarwal A, Verma I, Chatterjee A, et al. Analyzing Behavioral Trends in Community Driven Discussion Platforms Like Reddit. 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE; 2018. https://doi.org/10.1109/asonam.2018.8508687.

[3] Wang F-Y, Miao Q, Li X, Wang X, Lin Y. What Does ChatGPT Say: The DAO from Algorithmic Intelligence to Linguistic Intelligence. IEEE/CAA Journal of Automatica Sinica 2023;10:575–9. https://doi.org/10.1109/jas.2023.123486.

[4] Izak M, Mansell S, Fuller T. Introduction: Between no future and business-as-usual: Exploring futures of capitalism. Futures 2015;68:1–4. https://doi.org/10.1016/j.futures.2015.03.006.

[5] Olhede SC, Wolfe PJ. The growing ubiquity of algorithms in society: implications, impacts and innovations. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 2018;376:20170364. https://doi.org/10.1098/rsta.2017.0364.

[6] Schepman A, Rodway P. Initial validation of the general attitudes towards Artificial Intelligence Scale. Computers in Human Behavior Reports 2020;1:100014. https://doi.org/10.1016/j.chbr.2020.100014.

[7] Hacker P, Engel A, Mauer M. Regulating ChatGPT and other Large Generative AI Models 2023. https://doi.org/10.48550/ARXIV.2302.02337.

[8] Berthelot J-M, Latouche M. Improving the Efficiency of Data Collection: A Generic Respondent Follow-up Strategy for Economic Surveys. Journal of Business &amp Economic Statistics 1993;11:417. https://doi.org/10.2307/1391632.

[9] Borgi T, Zoghlami N, Abed M, Naceur MS. Big Data for Operational Efficiency of Transport and Logistics: A Review. 2017 6th IEEE International Conference on Advanced Logistics and Transport (ICALT), IEEE; 2017. https://doi.org/10.1109/icadlt.2017.8547029.

[10] Dewi LC, Meiliana, Chandra A. Social Media Web Scraping using Social Media Developers API and Regex. Procedia Comput Sci 2019;157:444–9. https://doi.org/10.1016/j.procs.2019.08.237.

[11] Web Scraping: Applications and Scraping Tools. International Journal of Advanced Trends in Computer Science and Engineering 2020;9:8202–6. https://doi.org/10.30534/ijatcse/2020/185952020.

[12] Krotov V, Silva L. Legality and Ethics of Web Scraping, 2018.

[13] reddit inc. Reddit API Documentation Overview n.d.

[14] Baeza-Yates R. Bias on the web. Commun ACM 2018;61:54–61. https://doi.org/10.1145/3209581.

[15] Colleoni E, Rozza A, Arvidsson A. Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data. Journal of Communication 2014;64:317–32. https://doi.org/10.1111/jcom.12084.

[16] Lawrence E, Sides J, Farrell H. Self-Segregation or Deliberation? Blog Readership, Participation, and Polarization in American Politics. Perspectives on Politics 2010;8:141–57. https://doi.org/10.1017/s1537592709992714.

[17] Couper MP. The Future of Modes of Data Collection. Public Opin Q 2011;75:889–908. https://doi.org/10.1093/poq/nfr046.

[18] Cochran WG. Sampling Techniques: 3d Ed. Wiley; 1977.

[19] Fuller WA. Sampling Statistics. New York: Wiley; 2011.

[20] Sedgwick P. Multistage sampling. BMJ 2015:h4155. https://doi.org/10.1136/bmj.h4155.

[21] Marshall AW. The use of multi-stage sampling schemes in Monte Carlo computations. 1954.

[22] Kuno E. Multi-stage sampling for population estimation. Popul Ecol 1976;18:39–56. https://doi.org/10.1007/bf02754081.

[23] Wang J, Ge G, Fan Y, Chen L, Liu S, Jin Y, et al. The estimation of sample size in multi-stage sampling and its application in medical survey. Appl Math Comput 2006;178:239–49. https://doi.org/10.1016/j.amc.2005.11.043.

[24] Xia W, Ma C, Liu J, Liu S, Chen F, Yang Z, et al. High-Resolution Remote Sensing Imagery Classification of Imbalanced Data Using Multistage Sampling Method and Deep Neural Networks. Remote Sens (Basel) 2019;11:2523. https://doi.org/10.3390/rs11212523.

[25] Gualdi G, Prati A, Cucchiara R. Multi-stage Sampling with Boosting Cascades for Pedestrian Detection in Images and Videos. Computer Vision ECCV 2010, Springer Berlin Heidelberg; 2010, p. 196–209. https://doi.org/10.1007/978-3-642-15567-3_15.

[26] Hankin DG, Mohr MS, Newman KB. Multi-stage sampling. Sampling Theory, Oxford University PressOxford; 2019, p. 173–99. https://doi.org/10.1093/oso/9780198815792.003.0009.

[27] Qian L, Zhou G, Kong F, Zhu Q. Semi-supervised learning for semantic relation classification using stratified sampling strategy. Proceedings of the 2009 conference on empirical methods in natural language processing, 2009, p. 1437–45.

[28] Shi X, Xiao Y. Modeling multi-mapping relations for precise cross-lingual entity alignment. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, p. 813–22.

[29] Efraimidis P, Spirakis P. Weighted Random Sampling. Encyclopedia of Algorithms, Springer US; 2008, p. 1024–7. https://doi.org/10.1007/978-0-387-30162-4_478.

[30] WINSHIP C, RADBILL L. Sampling Weights and Regression Analysis. Sociological Methods &amp Research 1994;23:230–57. https://doi.org/10.1177/0049124194023002004.

[31] Skinner CJ. Probability Proportional to Size (scpPPS/scp) Sampling 2016:1–5. https://doi.org/10.1002/9781118445112.stat03346.pub2.

[32] Parsons VL. Stratified Sampling 2017:1–11. https://doi.org/10.1002/9781118445112.stat05999.pub2.
[33] Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal 2014;5:1093–113. https://doi.org/10.1016/j.asej.2014.04.011.
[34] Nasukawa T, Yi J. Sentiment analysis. Proceedings of the 2nd international conference on Knowledge capture, ACM; 2003. https://doi.org/10.1145/945645.945658.
[35] Melton CA, Olusanya OA, Ammar N, Shaban-Nejad A. Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence. J Infect Public Health 2021;14:1505–12. https://doi.org/10.1016/j.jiph.2021.08.010.
[36] Chong WY, Selvaretnam B, Soon L-K. Natural Language Processing for Sentiment Analysis: An Exploratory Analysis on Tweets. 2014 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology, IEEE; 2014. https://doi.org/10.1109/icaiet.2014.43.
[37] Troussas C, Virvou M, Espinosa KJ, Llaguno K, Caro J. Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning. IISA 2013, IEEE; 2013. https://doi.org/10.1109/iisa.2013.6623713.
[38] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding 2018. https://doi.org/10.48550/ARXIV.1810.04805.
[39] Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach 2019. https://doi.org/10.48550/ARXIV.1907.11692.
[40] Tarunesh I, Aditya S, Choudhury M. Trusting RoBERTa over BERT: Insights from CheckListing the Natural Language Inference Task 2021. https://doi.org/10.48550/ARXIV.2107.07229.
[41] Hutto C, Gilbert E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Proceedings of the International AAAI Conference on Web and Social Media 2014;8:216–25. https://doi.org/10.1609/icwsm.v8i1.14550.
[42] Shah SMA, Singh S. Hate Speech and Offensive Language Detection in Twitter Data Using Machine Learning Classifiers. Innovations in Computer Science and Engineering, Springer Nature Singapore; 2023, p. 221–37. https://doi.org/10.1007/978-981-19-7455-7_17.
[43] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of Machine Learning Research 2003;3:993–1022.
[44] Blei DM. Probabilistic topic models. Commun ACM 2012;55:77–84.
[45] Jockers ML, Mimno D. Significant themes in 19th-century literature. Poetics 2013;41:750–69. https://doi.org/10.1016/j.poetic.2013.08.005.
[46] Jelodar H, Wang Y, Yuan C, Feng X, Jiang X, Li Y, et al. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. Multimed Tools Appl 2018;78:15169–211. https://doi.org/10.1007/s11042-018-6894-4.
[47] de Finetti. Theory of Probability. vol. 1–2. Chichester: John Wiley & Sons Ltd.; 1990.
[48] Blei D, Lafferty J. Correlated topic models. Adv Neural Inf Process Syst 2006;18:147.
[49] Roberts ME, Stewart BM, Tingley D, Airoldi EM, others. The structural topic model and applied social science. Advances in neural information processing systems workshop on topic models: computation, application, and evaluation, vol. 4, 2013, p. 1–20.
[50] Berg-Kirkpatrick T, Burkett D, Klein D. An empirical investigation of statistical significance in NLP. Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 2012, p. 995–1005.
[51] Dror R, Baumer G, Shlomov S, Reichart R. The hitchhiker's guide to testing statistical significance in natural language processing. Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: Long papers), 2018, p. 1383–92.
[52] Hussein M, Özyurt F. A new technique for sentiment analysis system based on deep learning using Chi-Square feature selection methods. Balkan Journal of Electrical and Computer Engineering 2021;9:320–6.
[53] Vargha A, Delaney HD. The Kruskal-Wallis Test and Stochastic Homogeneity. Journal of Educational and Behavioral Statistics 1998;23:170–92. https://doi.org/10.3102/10769986023002170.
[54] Futschik A, Taus T, Zehetmayer S. An omnibus test for the global null hypothesis. Stat Methods Med Res 2019;28:2292–304.
[55] Shaffer JP. Multiple Hypothesis Testing. Annu Rev Psychol 1995;46:561–84. https://doi.org/10.1146/annurev.ps.46.020195.003021.
[56] Barbieri F, Camacho-Collados J, Neves L, Espinosa-Anke L. Tweeteval: Unified benchmark and comparative evaluation for tweet classification. ArXiv Preprint ArXiv:201012421 2020.
[57] Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. Huggingface's transformers: State-of-the-art natural language processing. ArXiv Preprint ArXiv:191003771 2019.
[58] Sievert C, Shirley K. LDAvis: A method for visualizing and interpreting topics. Proceedings of the workshop on interactive language learning, visualization, and interfaces, 2014, p. 63–70.