

CIA – III

Text and Social Media Analytics

Topic

Analysing the sentiments of people about ChatGPT in the first month of its launch

Under the Guidance of

Dr Tripti Mahara



CHRIST (DEEMED TO BE UNIVERSITY), BANGALORE

January 2023

Submitted by:

MANAS SHRIVASTAV

2127821

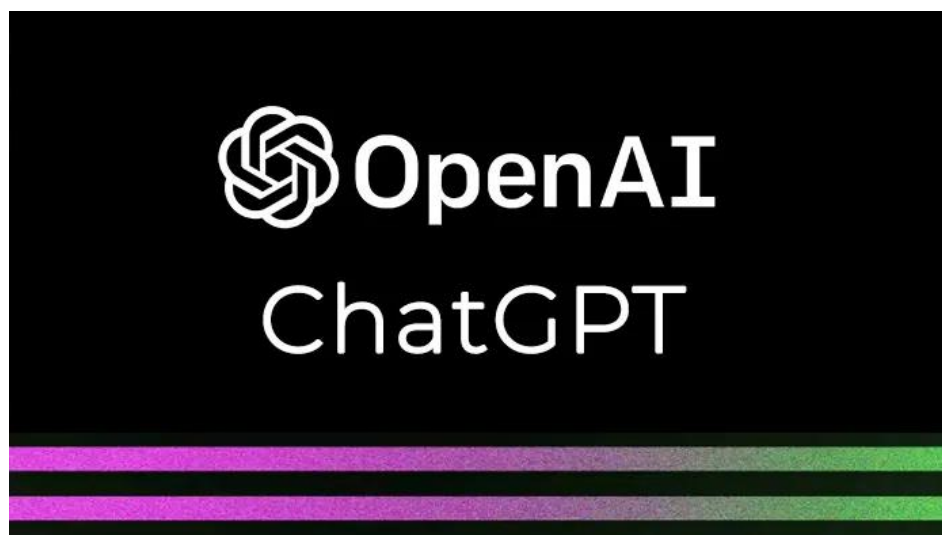
BA-2

INTRODUCTION

The field of Artificial Intelligence has seen significant advancements in recent years, and one of the most notable developments has been in the area of language generation models. One such model is ChatGPT, developed by OpenAI, which has received a lot of attention and interest in the AI community. ChatGPT is a large language generation model that is trained on a dataset of conversational exchanges and is able to generate human-like responses to text-based prompts. This makes it a powerful tool for tasks such as chatbot development and language translation.

The launch of ChatGPT in the market has sparked a lot of curiosity and interest among people. In this mini project, we aim to analyse the sentiments of people about ChatGPT in the first month of its launch. The goal of this project is to understand the initial reactions of people towards ChatGPT and to identify any major trends or patterns in the sentiment towards the model. The results of the sentiment analysis will be presented in the form of visualizations and charts. This will make it easier to identify any major trends or patterns in the sentiment towards ChatGPT. The visualizations will show the overall sentiment of the collected data.

In conclusion, this mini project aims to provide a valuable insight into the public perception of ChatGPT in the first month of its launch. By collecting and analyzing data from social media platforms, we will be able to understand the initial reactions of people towards ChatGPT and identify any major trends or patterns in the sentiment towards the model. The results of this project will provide valuable information to the AI community and will help in understanding how language generation models are perceived by the general public.



METHODOLOGY

1. **Data Collection**: To collect data for the project, we will be using social media platform Twitter. This platform has become a major source of public opinion and provide a valuable insight into the public perception of a particular topic. By using advanced web scraping technique like '*snsrape*', we will collect a large number of English language tweets, which have “#ChatGPT” in it and are from 1st December 2022 till 31st December 2022. Through these parameters, we were able to collect 60000 tweets for our analysis.
2. **Data pre-processing**: The collected data will be pre-processed to remove any irrelevant information like removing @mentions, hastags and links, to format it in a way that is suitable for analysis. This will be done using the '*regular expressions*' package.
3. **Sentiment Analysis**: Once the data has been pre-processed, we will perform sentiment analysis on the data. Sentiment analysis is a process of determining the emotional tone of a piece of text. It will be performed using VADER sentiment analysis which is a tool that is specifically designed to analyse social media text. It is particularly useful for analysing text that contains a lot of slang, emoticons, and other informal language.

Apart from VADER, we will be using NRC Lexicon which is a list of words and their associated emotional associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive).

4. **Data visualization & interpretation**: The results of the sentiment analysis will be presented in the form of visualizations and charts. This will make it easier to identify any major trends or patterns in the sentiment towards ChatGPT, and we will interpret the results of the analysis and draw conclusions about the public perception of ChatGPT in the first month of its launch. We will also identify any major trends or patterns in the sentiment towards the model

ANALYSIS OF SENTIMENT ANALYSIS TECHNIQUES

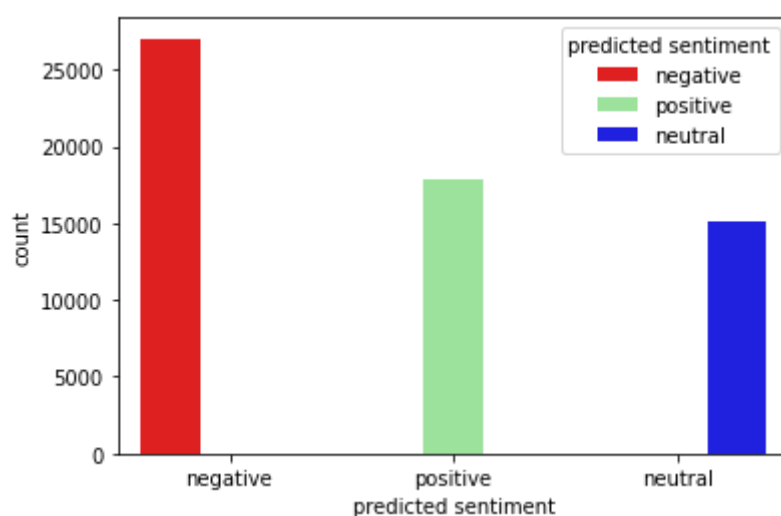
1. VADER:

The first step in using VADER is to clean the data by removing any irrelevant information such as links, hashtags, and tags. The regular expression package can be used to clean the data and remove any unwanted elements.

Once the data is cleaned, the next step is to perform sentiment analysis using the VADER algorithm. The VADER algorithm calculates a number of different sentiment scores for each piece of text. These include a negative score, a positive score, a neutral score, and a compound score. The negative score represents the proportion of negative words in the text, the positive score represents the proportion of positive words in the text, and the neutral score represents the proportion of neutral words in the text.

The compound score is a composite score that takes into account all the different sentiment scores. It ranges from -1 to 1, with a score of -1 indicating a strongly negative sentiment, a score of 0 indicating a neutral sentiment, and a score of 1 indicating a strongly positive sentiment.

Based on the compound score, the tweets are classified as positive, negative or neutral. A tweet with a compound score greater than 0.5 is classified as having a positive sentiment, a score between 0 and 0.5 is classified as neutral sentiment and a score less than 0 is classified as a negative sentiment.



Sentiment	Count
negative	27028
neutral	15106
positive	17867

The results of the sentiment analysis performed on the collected data, as represented in the visualizations and tables, indicate that 45% of the tweets have a negative sentiment towards ChatGPT. This suggests that a significant proportion of individuals expressing their opinions on the social media platform have been surprised or shocked by the capabilities of the tool. Example of a negative tweet:

“ChatGPT is scaring hell out of me. Both LBJ & Hoover admitted & all FBI analysts unanimously agreed someone impersonated Oswald on the phone in Mexico City in autumn 1963? WTH! This is SO dangerous for brainwashing. #chatGPT #JFKAssassination <https://t.co/LVahwInTic>”

Additionally, 25% of the tweets were found to have a neutral sentiment, indicating that a proportion of individuals do not have a concrete reaction towards the AI tool and are staying neutral in their expressions. In contrast, 29% of the tweets were found to have a positive sentiment, indicating that a proportion of individuals are excited about the tool and express appreciation for its capabilities.

Example of a positive tweet:

“I love the #ChatGPT. It is limited but I SO enjoy chatting with it, and it is able to help in research. Brilliant! @elonmusk Could you get Tesla's assistant understand like this, please!”

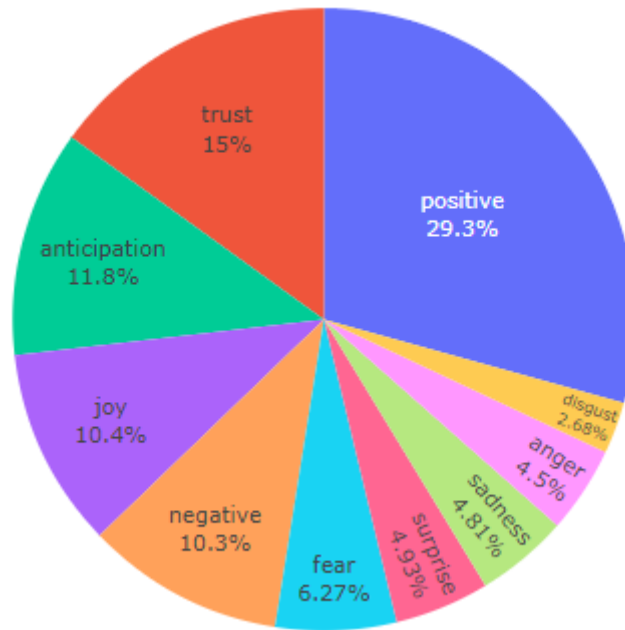
It is worth noting that these results should be interpreted with caution, as sentiment analysis is not an exact science and the subjectivity of language can lead to errors in the analysis. However, the results provide a general understanding of the public's perception of ChatGPT during the first month of its launch and can be used to inform future research and developments in the field of AI.

2. NRC LEXICON:

In order to gain a deeper understanding of the emotions expressed in the tweets, we used the NRC Emotion Lexicon (NRC Lexicon) to analyse the collected data. The NRC Lexicon is a list of words and their associated emotional associations, which can be used to quickly determine the emotional tone of a piece of text or identify specific emotions present in a text. By using this lexicon, we were able to identify the frequencies of the different emotions present in the data and determine the top emotion.

The analysis revealed that the top emotion present in the data was positive. This suggests that the majority of individuals expressing their opinions on the social media platform had positive emotions towards ChatGPT. This is inconsistent with the results of the VADER sentiment analysis, which indicated a negative sentiment towards the tool.

	Sentiment	Count
0	positive	63537
1	anticipation	25536
2	joy	22461
3	trust	32612
4	sadness	10422
5	fear	13592
6	negative	22297
7	disgust	5816
8	anger	9748
9	surprise	10679



The analysis of the collected data using the NRC Emotion Lexicon revealed that in addition to positive emotions, individuals expressing their opinions on the social media platform also displayed negative, fear, sadness and surprise emotions towards ChatGPT. This suggests that a proportion of individuals have a sceptical attitude towards the AI tool and its capabilities. It also implies that they have concerns about the future of AI and its impact on humans.

It is important to note that emotions are complex and multi-dimensional and can be expressed in different ways. The presence of negative, fear, sadness and surprise emotions does not necessarily indicate a completely negative sentiment towards the tool, but rather a more nuanced and complex attitude towards it. Therefore, it is important to take into account the frequencies of these emotions in relation to the overall sentiment and emotions expressed in the data.

RESULTS

Both VADER and the NRC Emotion Lexicon (NRC Lexicon) are powerful tools for performing sentiment analysis on Twitter data as seen in our project. They both have their own advantages and disadvantages and the choice between them depends on the specific requirements of the analysis.

VADER is specifically designed to analyse social media text and is particularly useful for analysing text that contains a lot of slang, emoticons, and other informal language. On the other hand, the NRC Lexicon is a list of words and their associated emotional associations that can be used to quickly determine the emotional tone of a piece of text or identify specific emotions present in a text. It provides more detailed information about the emotions present in a text. It can give us the frequencies of the different emotions present in the data and determine the top emotion.

In this study, we will be taking *NRC lexicon's results* as it gives importance of taking into consideration a multitude of emotions. By analysing the frequencies of the different emotions present in the data, we were able to determine the top emotions and understand the complexity of the sentiment towards ChatGPT. The results of the analysis revealed that people have a complex attitude towards ChatGPT and its capabilities, with a majority of individuals expressing positive emotions, but also some expressing negative and fear emotions.

The most frequent emotion present in the data is positive, with a frequency of 63,537. This suggests that a majority of individuals expressing their opinions on the social media platform had positive emotions towards ChatGPT. The second most frequent emotion is anticipation, with a frequency of 25,536, indicating that a proportion of individuals have an expectation towards the AI tool and its capabilities.

The third most frequent emotion is joy, with a frequency of 22,461, indicating that a proportion of individuals are happy and satisfied with the tool and its capabilities. Trust is also present with a frequency of 32,612, indicating that a proportion of individuals trust the tool and its capabilities.

However, the data also reveals the presence of negative emotions such as sadness, fear, negative and disgust with frequencies of 10,422, 13,592, 22,297 and 5,816 respectively. This suggests that a proportion of individuals have a sceptical attitude towards the AI tool and its capabilities and are concerned about the future of AI and its impact on humans.

CONCLUSION

In conclusion, this mini project aimed to analyse the sentiments of people about ChatGPT in the first month of its launch. Data was collected from Twitter and pre-processed to remove any irrelevant information. Sentiment analysis was performed using both VADER and NRC Emotion Lexicon (NRC Lexicon) to determine the public's perception of ChatGPT.

The results of VADER sentiment analysis indicate that 45% of the tweets have a negative sentiment towards ChatGPT, 25% of tweets were found to have a neutral sentiment, and 29% of tweets were found to have a positive sentiment. The NRC Lexicon analysis revealed that people are showing negative, fear, sadness, and surprise emotions in their tweets towards ChatGPT, indicating that they have a complex attitude towards the AI tool and its capabilities. The results of this analysis provide a deeper understanding of the emotions behind the tweets and can be used to inform future research and developments in the field of natural language processing and sentiment analysis.

The results of this project provide a general understanding of the public's perception of ChatGPT during the first month of its launch and can be used to inform future research and developments in the field of AI. The VADER algorithm provides a simple and quick way to classify tweets as positive, negative or neutral and the NRC Lexicon provides a deeper understanding of the emotions behind the tweets. The use of both methods together can provide a more comprehensive analysis of the sentiments towards ChatGPT.

Overall, this mini project demonstrates the importance of considering different emotions and not only the overall sentiment while analysing social media data and the results of this project can be used as a reference for future research in analysing the sentiments and emotions towards new technologies, products and services.