Twitter Sentiment Analysis about ChatGPT

Group 28
Peiyu Yang, Hanyun Hu
Quanyin Liu, Yige Yang

Research Questions

- What is the most popular topic related to Chat GPT on Twitter?
- What is the overall comment or attitude when people are commenting on Chat GPT on Twitter?

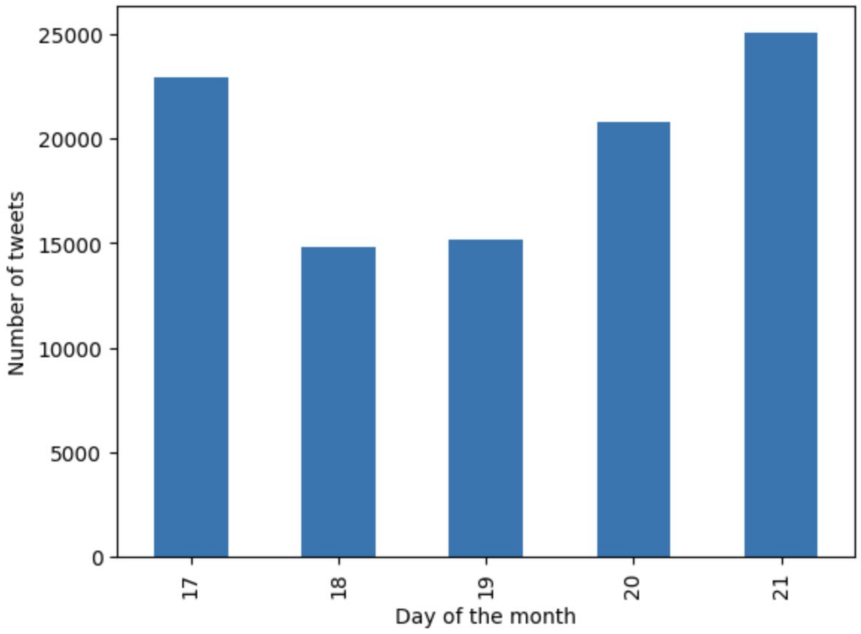
Data Preprocessing

Data Preprocessing

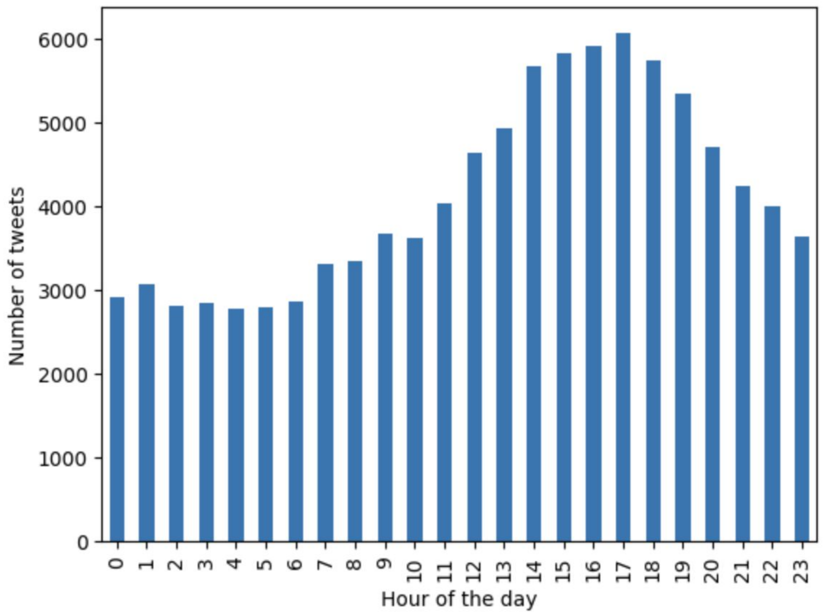
- Tokenization
    - Remove hashtags, retweet, punctuations, huperlinks, non-ASCII characters
    - Lowercase
- Remove stop words
- Built bigram and trigram
- Create Dictionary and Corpus
    - Dictionary: collection of unique words or terms that appear in the text data
    - Corpus: collection of documents or texts that are analyzed as a group
- Filter extreme words

Get Time Information

Distribution of tweets by day



Distribution of tweets by hour



Topic Modeling

Introduction for Topic Modeling

1. Model we used:

   Latent Dirichlet Allocation (LDA) Model

2. How it works?

   The aim of LDA is to find topics a document belongs to, based on the words in it. LDA classifies or categorizes the text into a document and the words per topic, these are modeled based on the Dirichlet distributions and processes.

3. Two key assumptions:

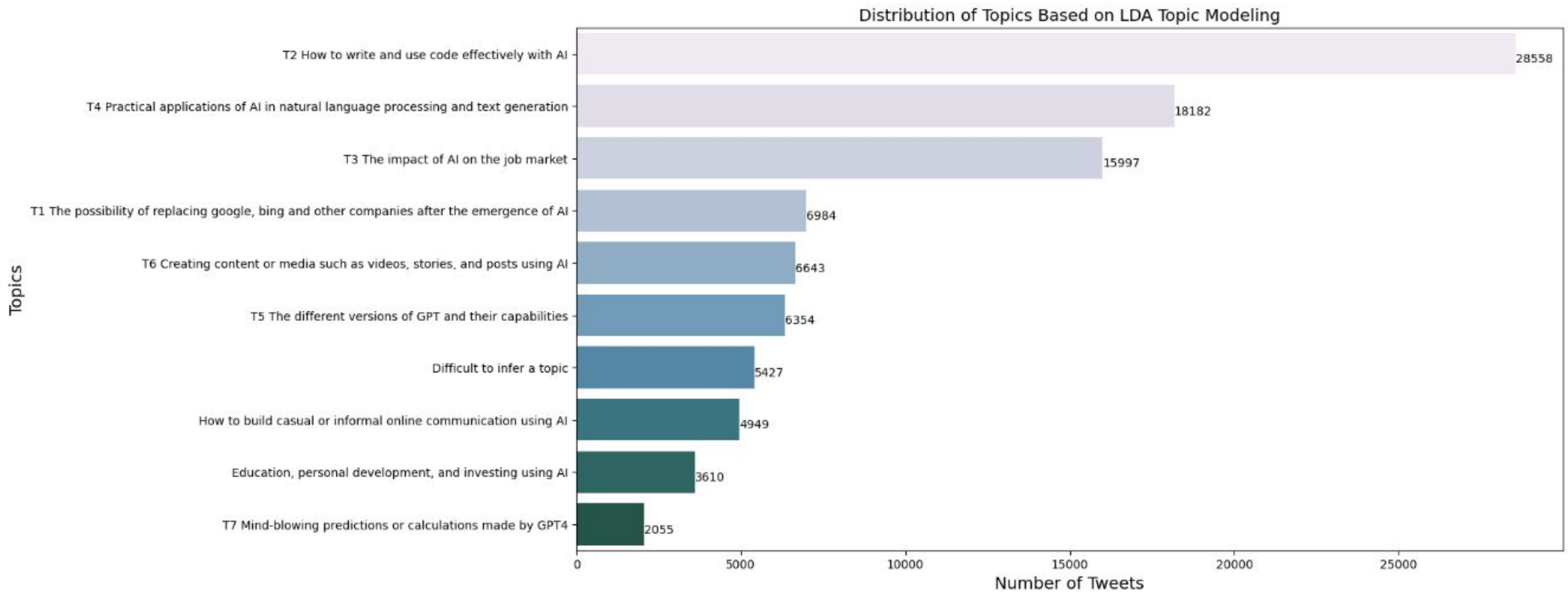
- Documents are a mixture of topics, and
- Topics are a mixture of tokens (or words)

Results

- Coherence Score: 0.46

- Percentage of tweets for each topic

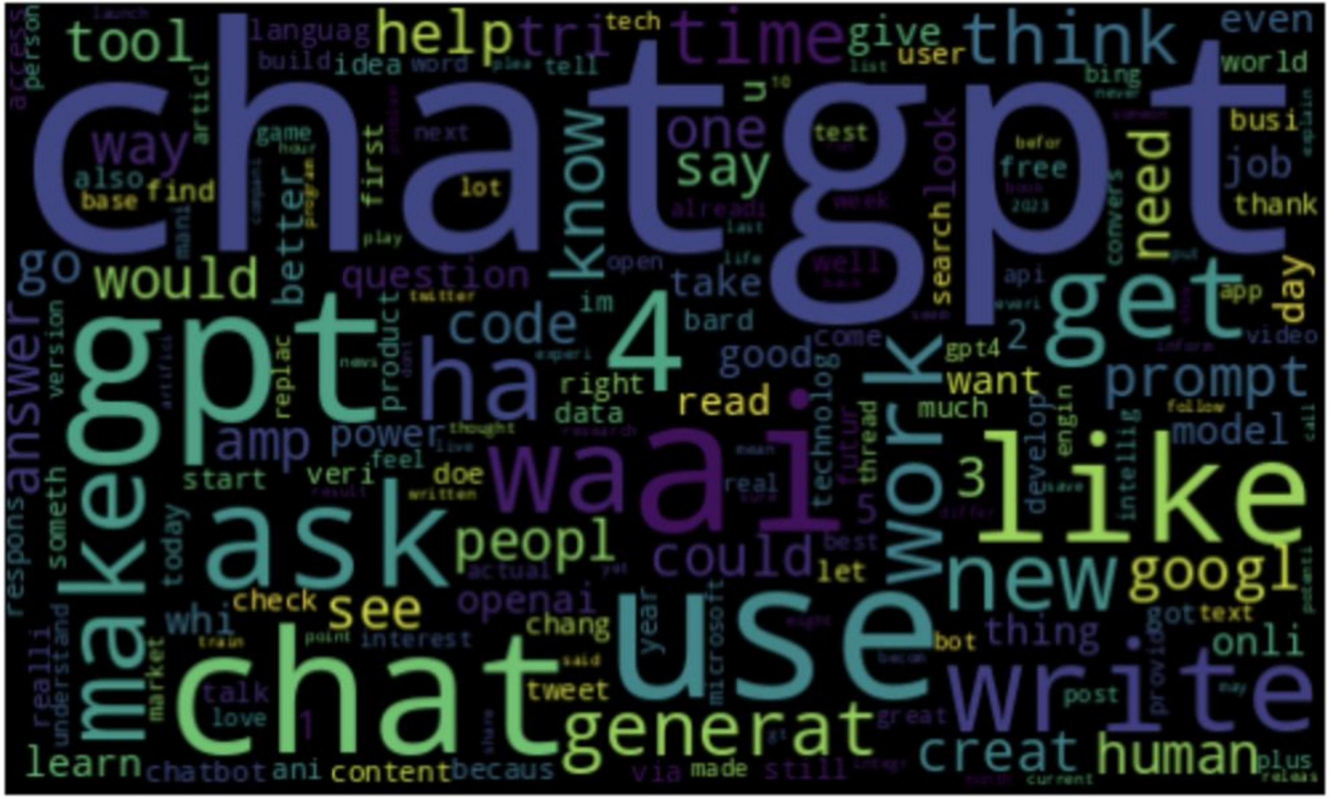
	Dominant_Topic	Total_tweet	Percentage%
0	2	28558	28.92
1	4	18182	18.41
2	3	15997	16.20
3	1	6984	7.07
4	6	6643	6.73
5	5	6354	6.43
6	10	5427	5.50
7	9	4949	5.01
8	8	3610	3.66
9	7	2055	2.08



T2 is the highest

Word Cloud

Topic 2 > Topic 4 > Topic 4 > Topic 1 > Topic 6 > Topic 5 > Topic 10 > Topic 9 > Topic 8 > Topic 7



Get words with the most frequencies

Sentiment Analysis

Model1 TextBlob

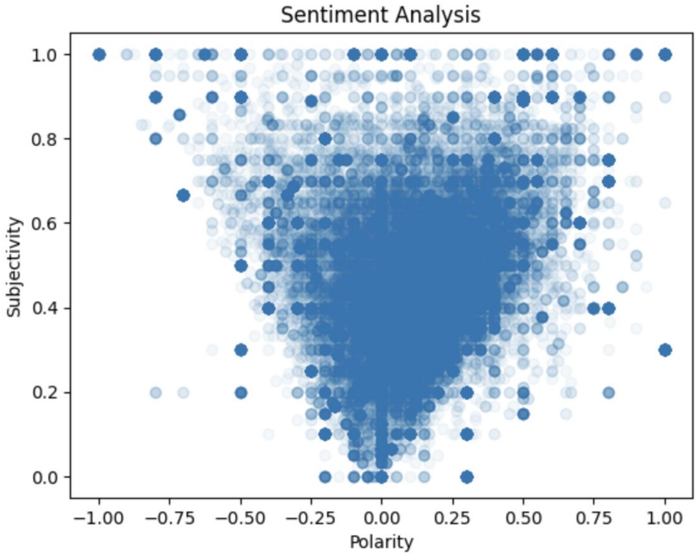
- TextBlob:
    - Rule-based sentiment analyzers
    - Built on Natural Language Toolkit (NLTK) and pattern libraries.
    - Output: Subjective (0 to 1) & Polarity (-1 to 1)
        - Polarity: smaller than 0 -> Negative; Equal to 0 -> Neutral; larger than 0 -> Positive

```
df['SentimentResult1'] = df['Polarity'].apply(lambda x: 'Positive' if x > 0.0 else ('Negative' if x < 0.0 else 'Neutral'))
df.head()
```

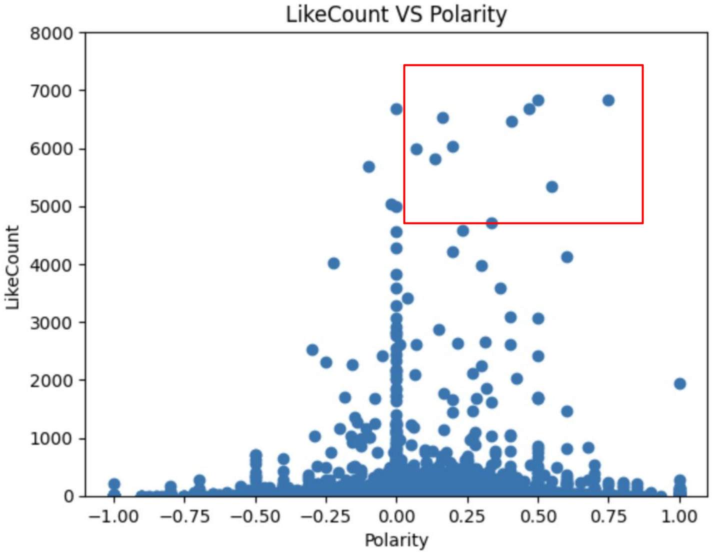
	ID	Date	Username	Tweet	ReplyCount	RetweetCount	LikeCount	QuoteCount	cleanTweet	year	month	day	hour	Subjectivity	Polarity	SentimentResult1
0	1638329623946878976	2023-03-21 23:59:55+00:00	lqgds36373	ChatGPT is another woke machine.	4	4	32	0	chatgpt is anoth woke machin	2023	3	21	23	0.0000	0.0000	Neutral
1	1638329621581275136	2023-03-21 23:59:55+00:00	yxwec12342	of the Atlantic, or only near the Atla # 推特账号 m...	0	0	0	0	of the atlant or onli near the atla more to me...	2023	3	21	23	0.4000	0.1333	Positive
2	1638329600471171074	2023-03-21 23:59:50+00:00	cwsea23772	This thread is saved to your Notion database....	0	0	0	0	this thread is save to your notion databas tag...	2023	3	21	23	0.0000	0.0000	Neutral
3	1638329587133194240	2023-03-21 23:59:46+00:00	jerje51666	Prompt AI – ChatGPT #0018	1	0	0	0	prompt ai chatgpt	2023	3	21	23	0.0000	0.0000	Neutral
4	1638329567759802368	2023-03-21 23:59:42+00:00	wwxly15746	Just had some interesting conversations with G...	1	0	0	0	just had some interest convers with googl s ba...	2023	3	21	23	0.5000	0.5000	Positive

TextBlob Visualization

Compute Polarity and Subjectivity based on TextBlob



Polarity vs Subjectivity



Polarity vs LikeCount

Method 2 roBERTa-base Model

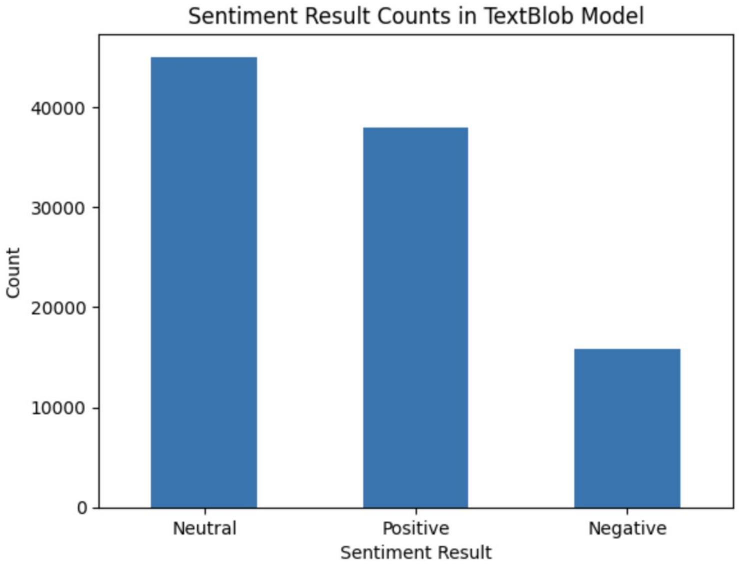
```
def get_roberta_sentiment_score(text):
    tokens = tokenizer.encode(text, add_special_tokens=True)
    output = roberta_model(torch.tensor([tokens]))[0]
    predicted_label = torch.argmax(output).item()

    if predicted_label == 0:
      return 'Negative'
    elif predicted_label == 1:
      return 'Neutral'
    else: #2
      return 'Positive'
```

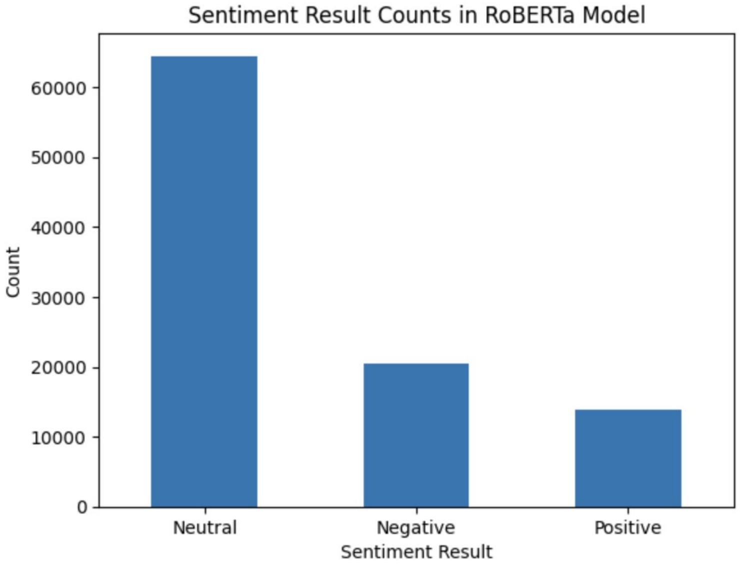
- Twitter-roBERTa-base model:
    - Trained on around 58M tweets
    - Finetune for sentiment analysis with the TweetEval benchmark
    - Output: a dictionary {labels: probabilities}
        - Labels: 0 -> Negative; 1 -> Neutral; 2 -> Positive
        - We take the label with the greatest probability as our sentence's sentiment result

ID	Date	Username	Tweet	ReplyCount	RetweetCount	LikeCount	QuoteCount	cleanTweet	year	month	day	hour	Subjectivity	Polarity	SentimentResult1	SentimentResult2
23946878976	2023-03-21 23:59:55+00:00	lqgds36373	ChatGPT is another woke machine.	4	4	32	0	chatgpt is anoth woke machin	2023	3	21	23	0.0000	0.0000	Neutra	Neutral
21581275136	2023-03-21 23:59:55+00:00	yxwec12342	of the Atlantic, or only near the Atla #æ ¨ç ¹...	0	0	0	0	of the atlant or onli near the atla more to me...	2023	3	21	23	0.4000	0.1333	Positive	Neutral
00471171074	2023-03-21 23:59:50+00:00	cwsea23772	This thread is saved to your Notion database....	0	0	0	0	this thread is save to your notion databas tag...	2023	3	21	23	0.0000	0.0000	Neutra	Neutral
87133194240	2023-03-21 23:59:46+00:00	jerje51666	Prompt AI â€ ChatGPTÂ #0018	1	0	0	0	prompt ai chatgpt	2023	3	21	23	0.0000	0.0000	Neutra	Neutral
67759802368	2023-03-21 23:59:42+00:00	wwxly15746	Just had some interesting conversations with G...	1	0	0	0	just had some interest convers with googl s ha	2023	3	21	23	0.5000	0.5000	Positive	Positive

Compare sentiment results of two methods

TextBlob vs roBERTa

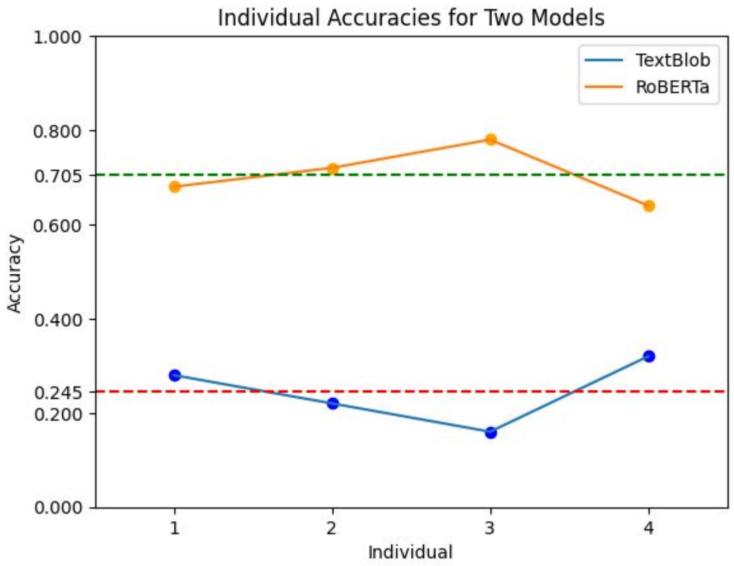




Further Investigation

- Want to know which model predict more accurately
- Problem: Lack of test dataset
- How to solve:
    - We extracted 50 random samples from the tweets that have different sentiment results between the two models.
    - To minimize subjectivity, each of us manually annotated the sentiments of the samples and compared the human annotations to the sentiment results of the two models to calculate accuracies. The mean of the four accuracies was then computed.

Evaluation & Finding



Although base on our test data, the roBERTa-base model turns out to be more accurate than TextBlob, we still cannot say the roBERTa is the best model, since our test dataset is so small. Also, we only extract 50 sentences from thousands of lines.

But we still find some interesting things during evaluation...

For example, “This is reassuring. ChatGPT almost feels cozy now. It is a fool like the rest of us.” three of us label “negative” and one of us label “positive” .

Models that are trained on annotated data may be biased towards the perspectives of the human annotators, who can bring their own biases and cultural backgrounds to the task of labeling the data. As a result, a model's performance may vary across different populations or contexts, depending on the inherent biases in the training data.

Thank You!