

END-TO-END DATA SCIENCE PROJECT

Twitter (X) Sentiment Analysis: Public Perception of AI-Generated Arts

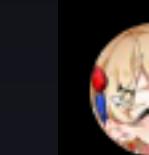
Analysis by Muhammad Zydan Priambada



Introduction

Generative AI has become a trend on the internet, including AI-generated images. The increasing image quality produced by AI brings a lot of controversies, especially in the artistic community. Some concerns includes the questionable ethics on how the AI models gathered images for training, the fear of artists losing their jobs to AI, how people impersonated themselves as an artists using AI art, and also real artists that are falsely accused of using AI.

 Dec 15 ⚡ ...
this portion illustrates everything. non artists love **ai** because they don't view **art** as **art**, they view **art** as a consumer good. content. something that "fails" unless sufficiently consumed

 · Jan 18, 2023 ⚡ ...
It seems more apparent by the day that there is an increase in **AI art** requests through both Pixiv and Skeb.

Please be aware of this when you are browsing for artists as there are many bad faith actors trying to get a good buck out of you.

Analysis By,
Muhammad Zydan Priambada

TWITTER DATA SCRAPING

Around 978 tweets related to AI-generated images are scraped

All tweets are scraped using twitterapi.io, fast and cost-efficient alternative API for scraping public social data.

The API will output a JSON response, which then later converted into pandas dataframe and saved to CSV.

Access Jupyter Notebook:
https://github.com/dan9111/DataSciencePortfolioProject/blob/main/Twitter%20Sentiment%20Analysis/sentiment_analysis_for_ai_art_twitter.ipynb

SCRAPER CODE SNIPPET

```
import time
import json
import requests

# Define the query
query = "AI art"
encoded_query = query.replace(" ", "%20")

# query up to 50 page with cursor pagination
page = 50
cursor = ""
response_list = [] # Initialize response_list to store all responses
for i in range(page):
    time.sleep(5)
    if cursor == "":
        url = f"https://api.twitterapi.io/twitter/tweet/advanced_search?queryType=Top&query={encoded_query}"
    else:
        url = f"https://api.twitterapi.io/twitter/tweet/advanced_search?queryType=Top&cursor={cursor}&query={encoded_query}"
    headers = {"X-API-Key": TWITTERAPI_KEY}
    response = requests.get(url, headers=headers)
    response_data = response.json()
    response_list.append(response_data) # Append each response data to the list
    # if no cursor left break the loop
    if not response_data["has_next_page"]:
        break
    # get cursor
    cursor = response_data["next_cursor"]
```

LANGUAGE

PYTHON

DATA CLEANING

Only 516 out of 978 tweets are left from the cleaning process

The scraped tweets have a lot of unwanted tweets that can be cleaned, that includes:

Duplicate Tweets

Twitter is prone of bot accounts with automated tweets, which can mess up the analysis.

Noises for Engagement farming

Some tweets are considered spam which will be irrelevant to the main topic. That includes: NFT/Crypto spam, giveaways, follow 4 follow, etc.

Semantic Relevance

Lastly, tweets must resembles the opinions and discussion related to AI generated art and it's tools. This can be resolved using semantic filtering.

Access Jupyter Notebook:

https://github.com/dan9111/DataSciencePortfolioProject/blob/main/Twitter%20Sentiment%20Analysis/sentiment_analysis_for_ai_art_twitter.ipynb

DATA CLEANING CODE SNIPPET

```
# DATA CLEANING
from sentence_transformers import SentenceTransformer, util

tweets_clean = tweets_df
# removing duplicates
tweets_clean = tweets_clean.drop_duplicates(subset=['text'])
# keyword filtering
noise_keywords = ['nft', 'mint', 'crypto', 'giveaway', 'wallet', 'opensea']
pattern = '|'.join(noise_keywords)
tweets_clean =
tweets_clean[~tweets_clean['text'].str.contains(pattern, case=False,
na=False)]

# semantic filtering
model = SentenceTransformer('all-MiniLM-L6-v2')
anchor_text = "Opinions, ethics, and discussions about AI generated
art and tools like Midjourney, DALL-E, Sora"

# Encode the anchor and the tweets
anchor_embedding = model.encode(anchor_text,
convert_to_tensor=True)
tweet_embeddings = model.encode(tweets_clean['text'].tolist(),
convert_to_tensor=True)

# Compute cosine similarity between the anchor and the tweets
cosine_scores = util.cos_sim(anchor_embedding,
tweet_embeddings)[0]

# add score
tweets_clean['relevance_score'] = cosine_scores.cpu().numpy()

final_tweets = tweets_clean[tweets_clean['relevance_score'] > 0.3]
```

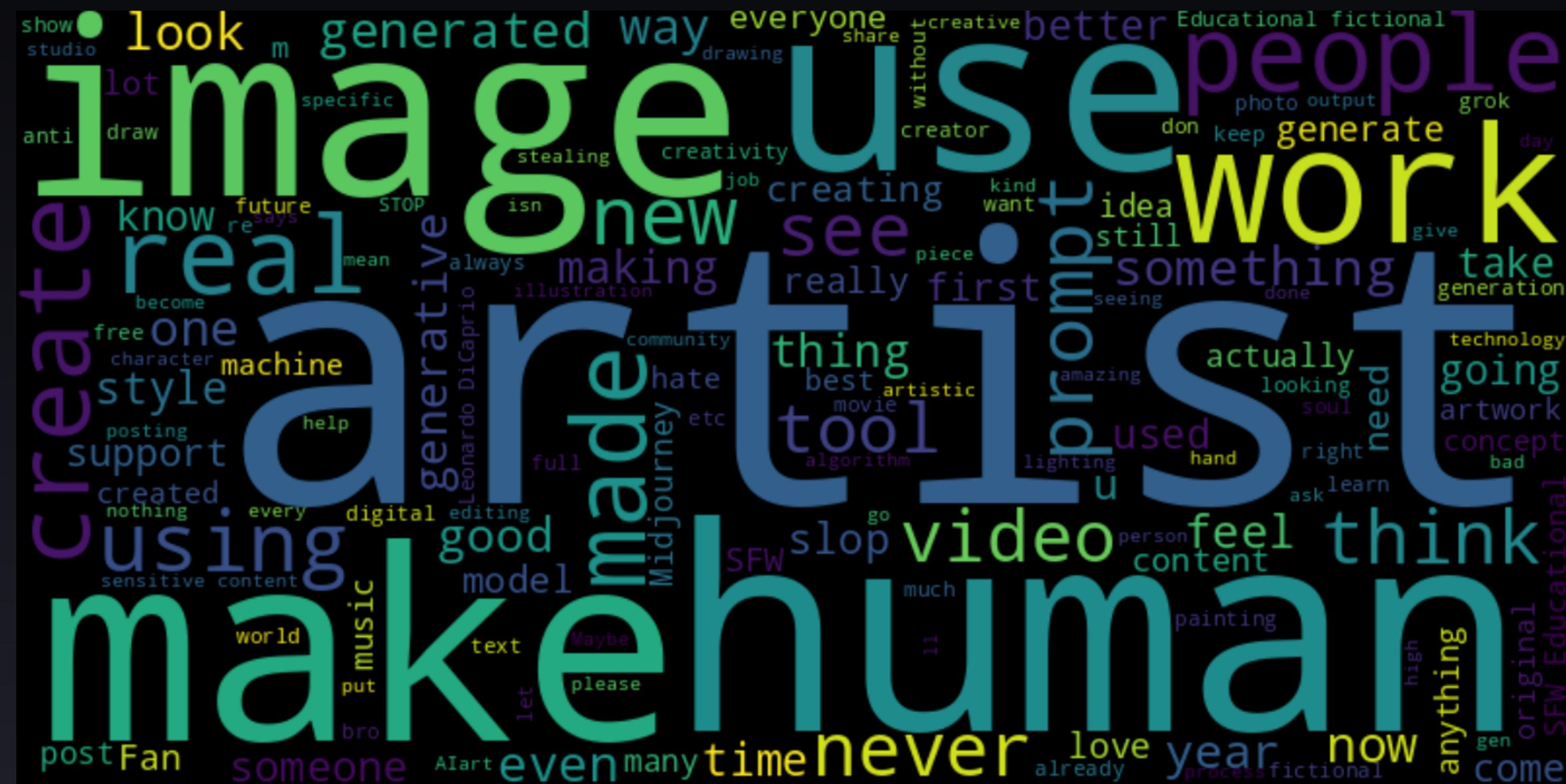
LANGUAGE

PYTHON

WORD CLOUD

Most Frequent Words Shown in AI Generated Image Related Tweets

Excluding the word: (AI, Art). With the word “artist” as the most frequent word, shows that tweets regarding AI generated images is mostly regarding to artistic professions



Access Jupyter Notebook:

https://github.com/dan9111/DataSciencePortfolioProject/blob/main/Twitter%20Sentiment%20Analysis/sentiment_analysis_for_ai_art_twitter.ipynb

SENTIMENT ANALYSIS

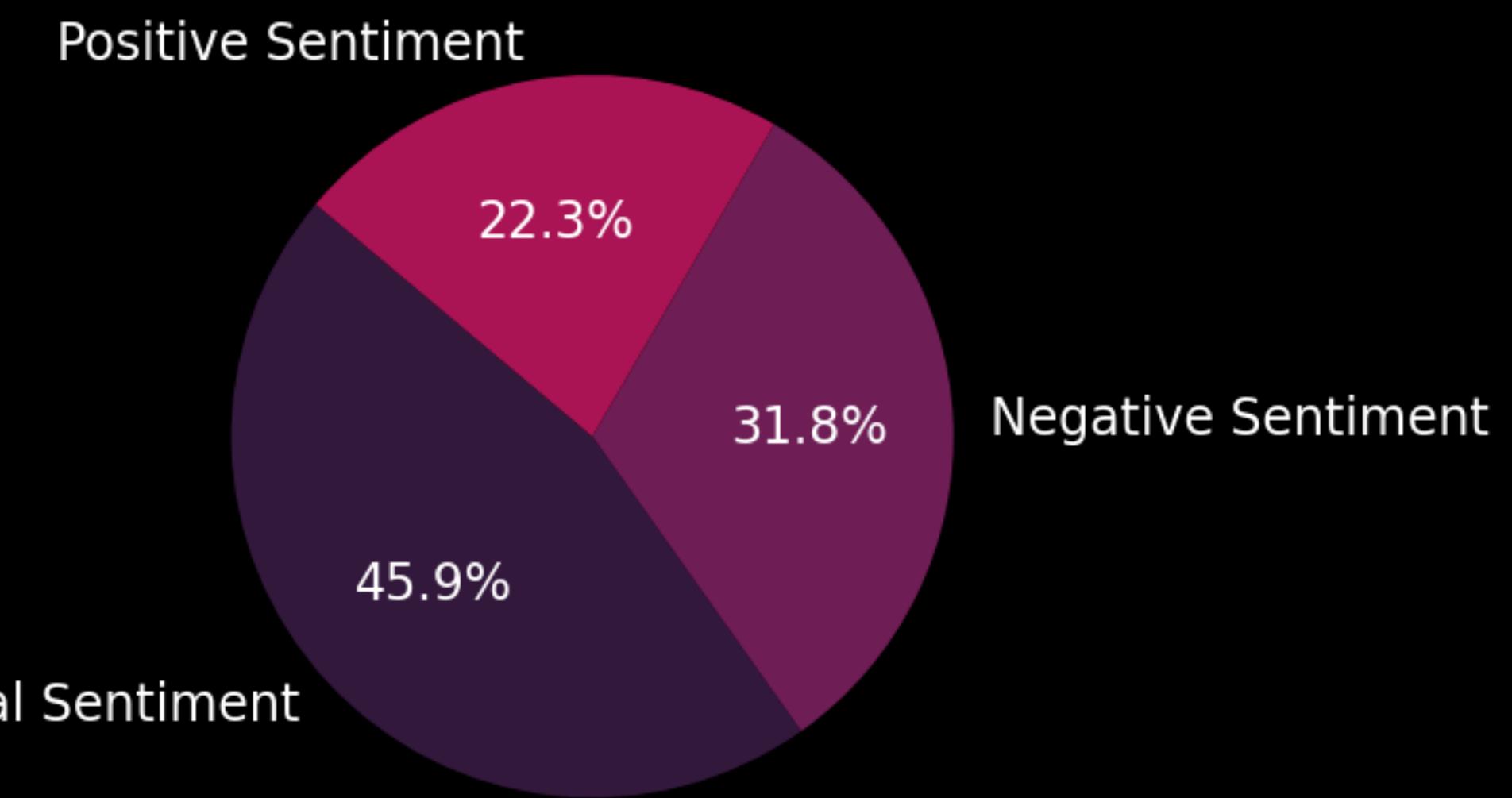
Labeling each tweets

Examples on how the text will be scored.
The output will be a label and the confidence score

Tweet	Label	Confidence
I love this new AI feature! I generated this photo-realistic drawing in 2 minutes.	Positive	0.9879729747772217
What AI does is just producing slops that only get 10 likes	Negative	0.7386000156402588
Presence of generative AI sparks challenges to the whole creative industries	Neutral	0.5361032485961914

Access Jupyter Notebook:

https://github.com/dan9111/DataSciencePortfolioProject/blob/main/Twitter%20Sentiment%20Analysis/sentiment_analysis_for_ai_art_twitter.ipynb



SENTIMENT DISTRIBUTION

**Only 22.3% Tweets
about AI-Generated
Images are Positive**

The sentiment analysis uses roBERTa model, which is trained from around 58M+ tweets. Out of all 516 tweet samples, 22.3% tweets are a positive sentiment, with 31.8% tweets strongly has negative sentiment, and the majority 45.9% tweets have a neutral sentiment.

Access Jupyter Notebook:

https://github.com/dan9111/DataSciencePortfolioProject/blob/main/Twitter%20Sentiment%20Analysis/sentiment_analysis_for_ai_art_twitter.ipynb

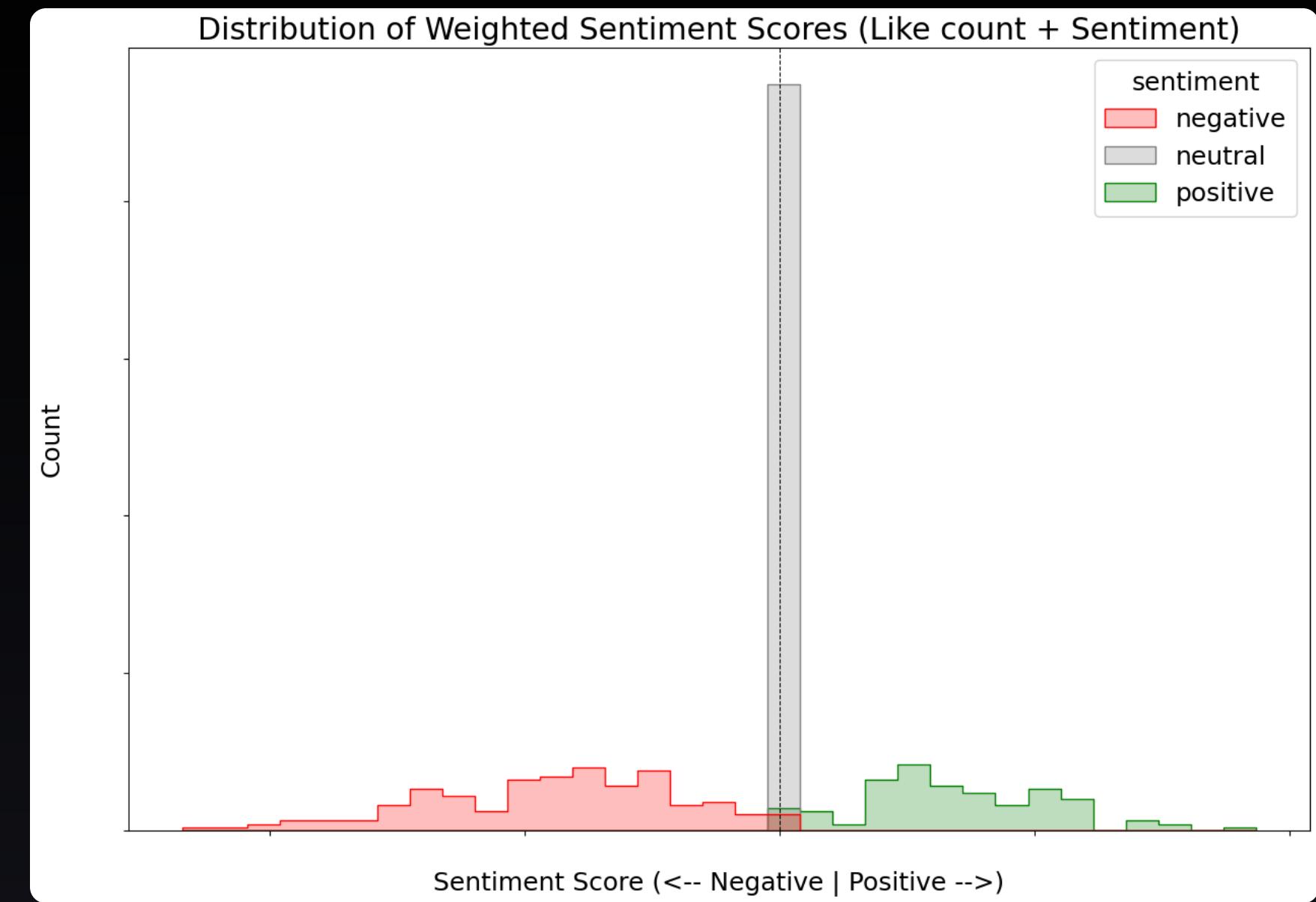


WEIGHTED SENTIMENT SCORES BASED ON LIKES

What if the amount of likes on each tweet determines the sentiment's impact?

Skewness

The like engagements are identical to both positive and negative sentiments, with the negative sentiments having broader sentiment scores.



Less engaging posts became closer to neutral

Statements that doesn't have enough engagements are considered to be neutral enough.



Access Jupyter Notebook:

https://github.com/dan9111/DataSciencePortfolioProject/blob/main/Twitter%20Sentiment%20Analysis/sentiment_analysis_for_ai_art_twitter.ipynb



CONCLUSION

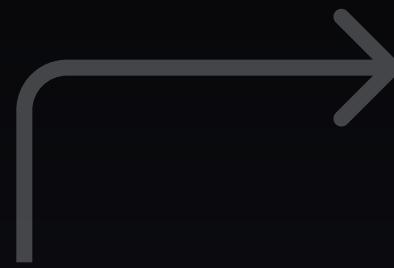
What can be concluded in this sentiment analysis?

The word “artist” is the most frequent words found in AI art related tweet. This shows that most discourse regarding AI-generated images are about artistic profession and creative industry.

Approximately around 22 out of 100 tweets has a good sentiment in the discussion of AI-generated arts which placed them in the minority, while the rest are feeling neutral and negative about it. This would raise a question whether AI images should be used or not, since it will inevitably attracts a lot of negative responds.

Tweets that has negative sentiments on AI-generated art topics provides more like engagements than the positive or neutral ones, making the negative sentiments to be more impactful on twitter.

What could be improved from this analysis



1

More Data

This analysis only contains 516 tweet samples. Having more data may improve the reliability of the analysis

2

Use Latest Model

The roBERTa model has a new sentiment analysis model which is trained from 128M+ tweets. This implementation of new model needs to be explored more

3

Add Parameters

Parameters such as retweet counts, author's status, are helpful for determining the sentiment weight of each tweet.



Thank You

REFERENCE USED FOR SENTIMENT ANALYSIS MODEL

Barbieri, F., Camacho-Collados, J., Espinosa Anke, L., & Neves, L. (2020, November). TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification. Findings of the Association for Computational Linguistics: EMNLP 2020, 1644–1650. doi:10.18653/v1/2020.findings-emnlp.148

