



UNDERSTANDING PUBLIC PERCEPTION: SENTIMENT ANALYSIS OF CHATGPT-RELATED TWEETS

Presented by:
Justine Tinio



TABLE OF CONTENTS

01

INTRODUCTION

02

DATA SOURCE AND EXPLORATION

03

DATA PREPARATION AND FEATURE ENGINEERING

04

MODEL EXPLORATION

05

MODEL RECOMMENDATION

06

CONCLUSION AND RECOMMENDATION

SENTIMENT ANALYSIS

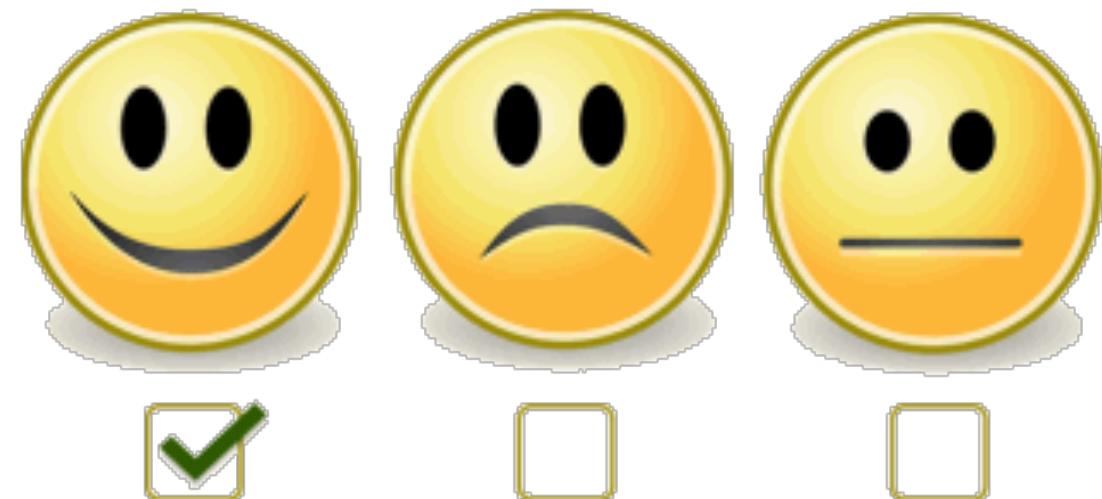
Definition: Analyzing text data to determine the sentiment or emotion it conveys.

Types of Sentiment:

- **Positive:** Expresses favorable opinions or emotions.
- **Negative:** Expresses unfavorable opinions or emotions.
- **Neutral:** Neither positive nor negative; often factual.

Applications: Product reviews, social media monitoring, feedback analysis, market research.

Challenges: Sarcasm, ambiguity, context-dependence, cultural differences.



INTRODUCTION TO X

Brief Overview: A microblogging and social networking platform launched in 2006.

User Base: Over 330 million monthly active users worldwide.

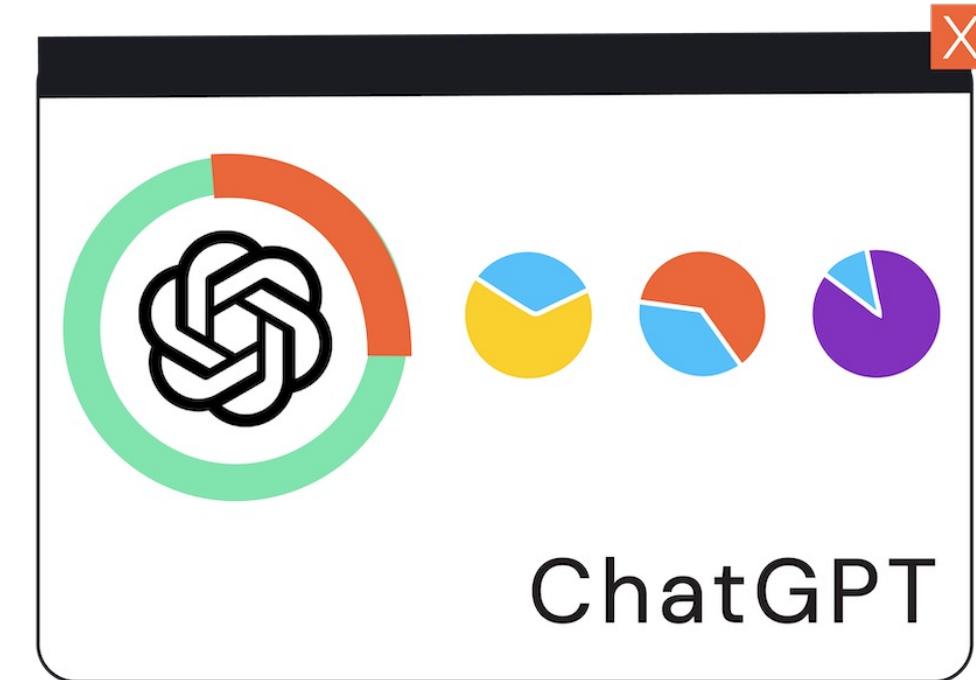
Tweet Characteristic: Limited to 280 characters, promoting concise communication.

Global Impact: A significant source for breaking news, trending topics, and global conversations.

Key Features: Retweets, hashtags, mentions, and direct messages.

ChatGPT - The Next Generation Conversational AI

- **Evolution of Chatbots:** ChatGPT, developed by OpenAI, represents the pinnacle of conversational AI, leveraging the power of GPT (Generative Pre-trained Transformer) models.
- **Deep Learning Power:** Built on state-of-the-art deep learning architectures, ChatGPT can understand context, generate human-like text, and engage in meaningful interactions.
- **Versatile Applications:** From customer service automation to virtual assistants and creative writing, ChatGPT finds usage across diverse domains.
- **Continuous Learning:** With each interaction and feedback, ChatGPT refines its responses, ensuring an ever-improving user experience.





EXECUTIVE SUMMARY

- **Who is the client?**

Companies, researchers, or developers interested in understanding public sentiment and feedback regarding ChatGPT and its applications.

- **What are their needs?**

To gauge public opinion on ChatGPT, understand user satisfaction, identify areas for improvement, and assess the overall market reception.

- **What is our solution?**

Deploy a comprehensive sentiment analysis model that analyzes ChatGPT-related tweets, categorizing them as positive, neutral, or negative.

Extract key themes and trends from the data to provide actionable insights.

- **Why is our solution the best solution?**

Leveraging advanced NLP techniques and models ensures high accuracy in sentiment prediction.



PROBLEM STATEMENT

"How can sentiment analysis algorithms effectively process and interpret the nuanced emotions expressed in short-form tweets to improve ChatGPT's response accuracy and enhance the user experience?"



Customer feedback and experience

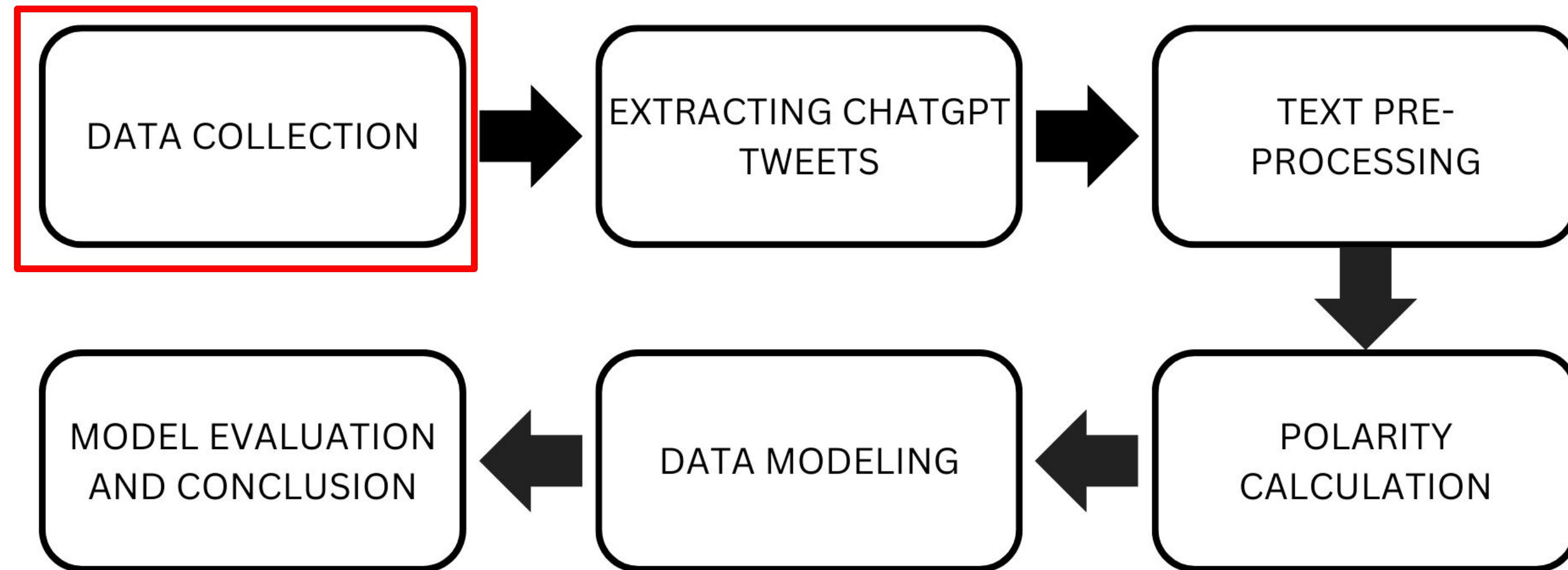


Market research and trend analysis for product development



Competitive Analysis

METHODOLOGY



DATA SOURCES



DATASET

"ChatGPT 1000 Daily Tweets" found in Kaggle offers a collection of tweets produced over time by ChatGPT.



ATTRIBUTES AND DATA RANGE

Tweets span a date range from April 2023, to May 10, 2023 containing 41,003 entries and 20 features (or attributes).



CONTENT

- Features capturing tweet content, user interactions, timestamps, and other relevant information.

Find the right access for you

Free

- For write-only use cases and testing the Twitter API
- Rate limited access to v2 tweet posting and media upload endpoints
 - 1,500 Tweets per month - posting limit at the app level
 - 1 app ID
 - Login with Twitter
 - Free

Get started

Basic

- For hobbyists or prototypes
- Rate limited access to suite of v2 endpoints
 - 3,000 Tweets per month - posting limit at the user level
 - 50,000 Tweets per month - posting limit at the app level
 - 10,000 Tweets per month - read-limit rate cap
 - 2 app IDs
 - Login with Twitter
 - \$100 per month

Subscribe now

Pro

- For startups scaling their business
- Rate-limited access to suite of v2 endpoints, including search and filtered stream
 - 1,000,000 Tweets per month - GET at the app level
 - 300,000 Tweets per month - posting limit at the app level
 - 3 app IDs
 - Login with Twitter
 - \$5,000 per month

Subscribe now

Enterprise

- For businesses and scaled commercial projects
- Commercial-level access that meets your and your customer's specific needs
 - Managed services by a dedicated account team
 - Complete streams: replay, engagement metrics, backfill, and more features
 - Monthly subscription tiers

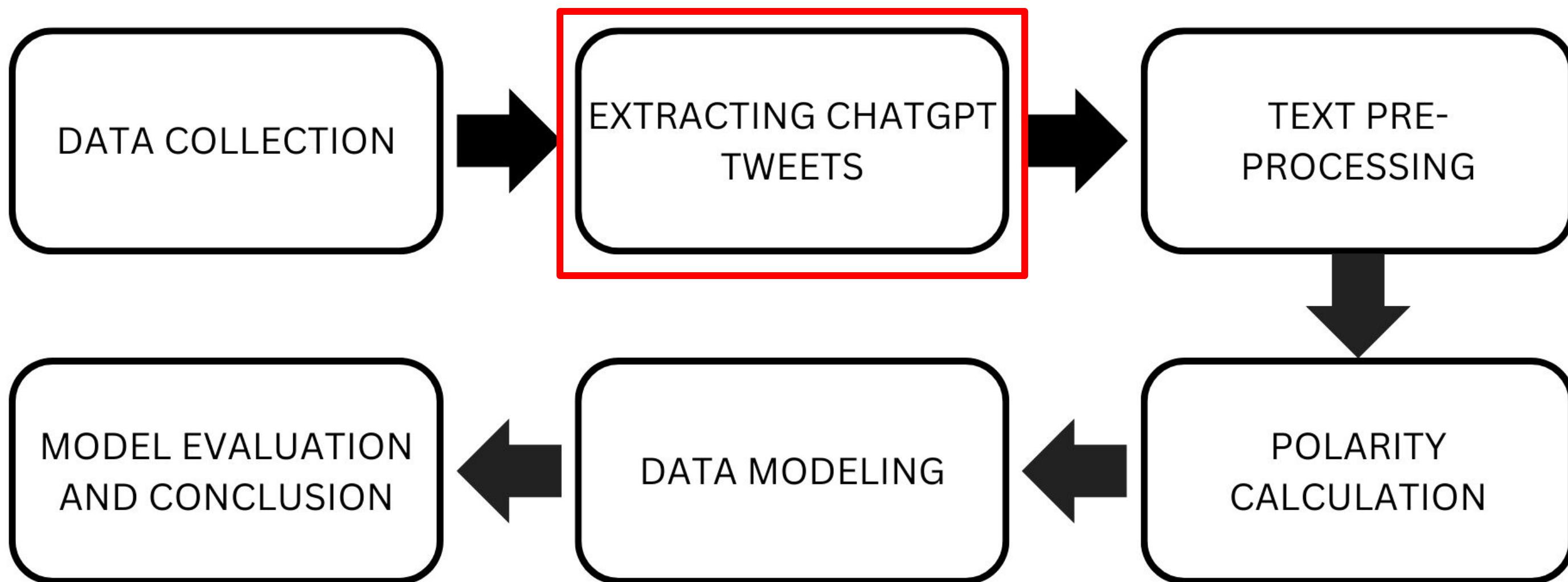
Apply now

Updated API pricing

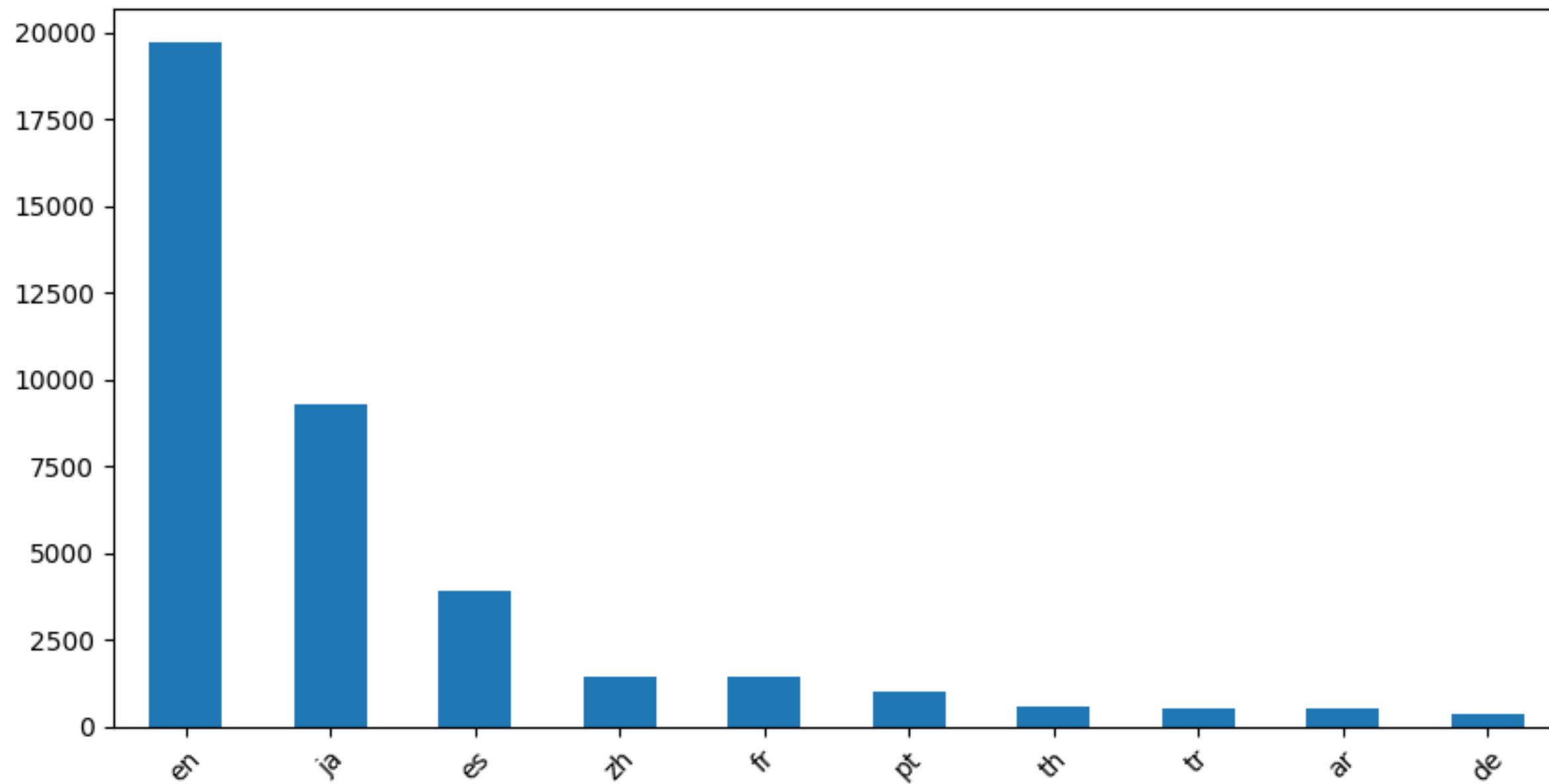
DATA DICTIONARY

Attribute	Description
tweet_id	Unique identifier for a tweet.
tweet_created	Date and time when the tweet was created.
tweet_extracted	Date and time when the tweet was extracted or retrieved.
text	The actual text content of the tweet.
lang	Language in which the tweet is written (ex. 'en' as English tweets).
user_id	Unique identifier for the user who posted the tweet.
user_name	Name of the user who posted the tweet.
user_username	Username or handle of the user who posted the tweet.
user_location	Location mentioned in the user's profile.
user_description	Description or bio provided by the user in their profile.
user_created	Date and time when the user's account was created.
user_followers_count	Number of followers the user has.
user_following_count	Number of accounts the user is following.
user_tweet_count	Total number of tweets posted by the user.
user_verified	Indicates whether the user's account is verified (True/False).
source	The source or platform from which the tweet was posted.
retweet_count	Number of times the tweet has been retweeted.
like_count	Number of times the tweet has been liked.
reply_count	Number of replies received by the tweet.
impression_count	Number of times the tweet has been seen or displayed.

METHODOLOGY

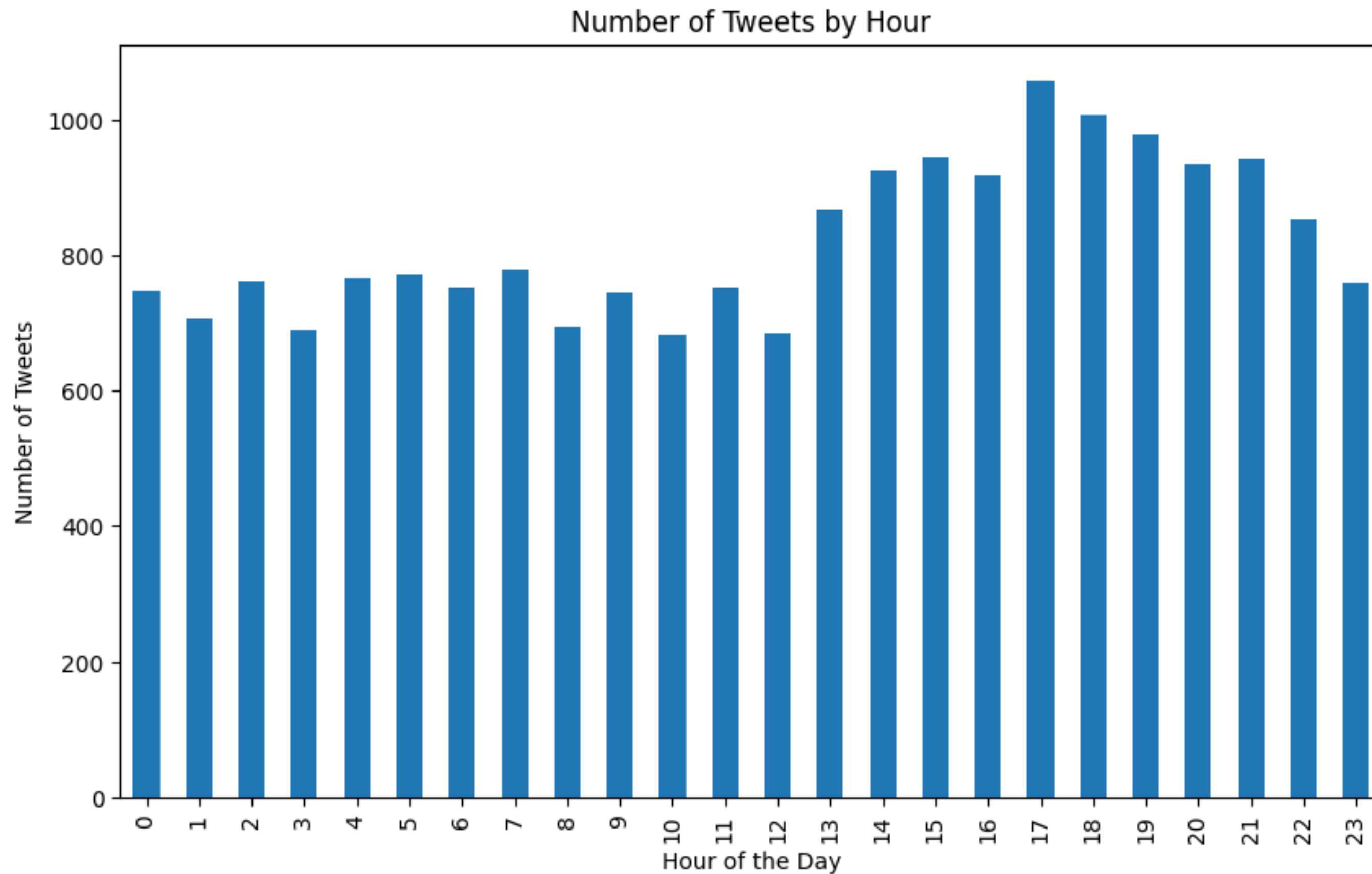


DATA EXPLORATION - LANGUAGE



en - English - 19,711
ja - Japanese
es - Espanol/Spanish
zh - Chinese
fr - French
pt - Portugese
th - Thai
tr - Turkish
ar - Arabic
de - Deutsch

DATA EXPLORATION - TWEETS



DATA ASSUMPTIONS AND LIMITATIONS

Assumptions:

Representative Sample: The tweets collected provide a representative sample of the overall sentiment towards ChatGPT. This assumes that active X users and their sentiments mirror the broader user base or public opinion.

Authenticity: Tweets are genuine and not influenced by bots, promotional campaigns, or other artificial means of manipulation.

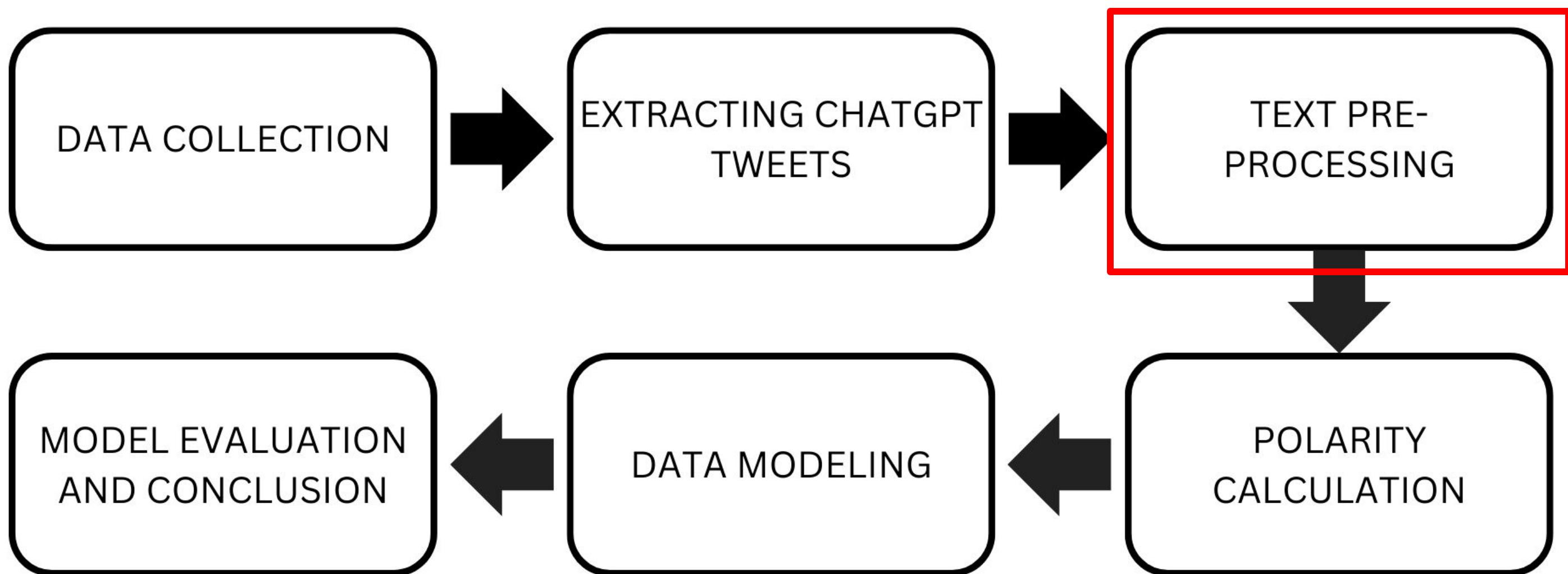
Clear Intent: Tweets related to ChatGPT are mainly about the model itself and not other external factors or events.

Limitations:

Character Limit: X's character limit might lead to oversimplified sentiments or opinions that lack nuance.

Limited Context: Due to the brevity and unavailability of extracting huge number of tweets, the context might be missing, leading to potential misclassification of sentiment.

METHODOLOGY



TEXT ANALYSIS OF TWEETS

Cleaning and Pre-processing textual data with NeatText library

NeatText:

Python library for text preprocessing and cleaning.

	text	clean_tweet
2	RT @DarrellLerner: ChatGPT Plugins are the fas...	RT ChatGPT Plugins are the fastest way to ge...
3	Get an intelligent chatbot for your website in...	Get an intelligent chatbot for your website in...
4	🔥Hey Guys, #ZenithSwap has launched at just \$...	🔥Hey Guys, has launched at just \$ 55,000 USD...
5	RT @sinsonetwork: Now! Join #SINSO DataLand^Ch...	RT Now! Join DataLand^ChatGPT \n 3.23-4.6...
9	The plagiarism detector will introduce its #AI...	The plagiarism detector will introduce its d...
...
40995	RT @solanaturbo: Hey everyone, have you heard ...	RT Hey everyone, have you heard about the ne...
40996	Can ChatGPT write my SoHo House application? W...	Can ChatGPT write my SoHo House application? W...
40998	RT @YGPT_official: 🚀 YGPT LAUNCH 🚀\n\nTime to ...	RT 🚀 YGPT LAUNCH 🚀\n\nTime to show you what ...
41000	RT @chatgpt_issac: Lets see which community is...	RT Lets see which community is bigger and I...
41001	@simamaung ChatGPT	ChatGPT

19711 rows × 2 columns



Noise

- remove mentions
- remove hashtags
- remove urls
- remove emojis
- remove special characters
- replacing contractions and their variations with their full forms (ex. “wouldn’t” to “would not”)
- remove numbers
- remove stopwords (ex. “and”, “of”, “is”) and converting tweets to lower case

TEXT ANALYSIS OF TWEETS

Cleaning and Pre-processing textual data with NeatText library

Removing hashtags

	text	clean_tweet
4	🔥Hey Guys, #ZenithSwap has launched at just \$...	🔥Hey Guys, has launched at just \$ 55,000 USD...
5	RT @sinsonetwork: Now! Join #SINSO DataLand^Ch...	RT @sinsonetwork: Now! Join DataLand^ChatGPT...

Removing mentions

	text	clean_tweet
2	RT @DarrellLerner ChatGPT Plugins are the fas...	RT ChatGPT Plugins are the fastest way to ge...
40998	RT @YGPT_official 🚀 YGPT LAUNCH 🚀\n\nTime to ...	RT 🚀 YGPT LAUNCH 🚀\n\nTime to show you what ...
41000	RT @chatgpt_issac Lets see which community is...	RT Lets see which community is bigger and I'...

TEXT ANALYSIS OF TWEETS

Cleaning and Pre-processing textual data with NeatText library

Removing special characters (symbols, whitespace characters (like tabs and line breaks))

index	text	clean_tweet
31216	RT @crispinhunt: Bravo @eu_comission : “Companies deploying generative AI tools, such as ChatGPT, will have to disclose any © material us...	RT Bravo Companies deploying generative AI tools such as ChatGPT will have to disclose any material us
15095	RT @BrianRoemmele: This is another in a series of SuperPrompts™ for ChatGPT (and other AI) from the https://t.co/MmjROXv0yF archive. In t...	RT This is another in a series of SuperPrompts for ChatGPT and other AI from the archive In t

Removing emojis

index	text	clean_tweet
4	🔥 Hey Guys, #ZenithSwap has launched at just \$ 55,000 USD Marketcap. The ChatGPT of DEX - Reimagining DeFi with AI-Powered Yield Farming. Now at 4X. Lot of up potential at such low marketcap.🔥 😊 \$ARB \$ZSP #Arbitrum https://t.co/9VWYtYzAJD	Hey Guys has launched at just \$ 55000 USD Marketcap The ChatGPT of DEX Reimagining DeFi with AIPowered Yield Farming Now at 4X Lot of up potential at such low marketcap \$ARB \$ZSP
5	📅 17 Mar 23-4.6 📅 Tasks ①Log in to<https://t.co/Hlwqa7HG40> ②Try SINSO #ChatGPT& twe...	RT Now Join DataLand^ChatGPT 32346 Tasks Log in tolt Try SINSO twe

TEXT ANALYSIS OF TWEETS

Cleaning and Pre-processing textual data with NeatText library

Removing url

index	text	clean_tweet
3	Get an intelligent chatbot for your website in minutes with Chatbase AI. Train ChatGPT on your data and let it answer any question your users have. Simply upload a document or link and add the chat widget - it's that easy! Make Money using AI: https://t.co/yLHeqn4w9T https://t.co/ba54JvoRsM	Get an intelligent chatbot for your website in minutes with Chatbase AI. Train ChatGPT on your data and let it answer any question your users have. Simply upload a document or link and add the chat widget - it's that easy! Make Money using AI:

Removing punctuations

index	text	clean_tweet
1848	RT @McaleerStephen: Forget plugins. ChatGPT can solve general computer tasks using a keyboard and mouse!! The trick? Recursively criticizi...	RT Forget plugins ChatGPT can solve general computer tasks using a keyboard and mouse The trick Recursively criticizi...
2236	@margal @RomanRussy I guess now we could just ask chatgpt to write a letter and then no one's the "orchestrator"!!	I guess now we could just ask chatgpt to write a letter and then no one's the "orchestrator"

TEXT ANALYSIS OF TWEETS

Cleaning and Pre-processing textual data with NeatText library

Replacing contractions and their variations with their full forms

index	text	clean_tweet
350	Most people in the world will look at generative AI and dismiss it, And i couldn't be happier that means I have a small headstart to learn more about it get into better roles and build better business. #AI #chatGPT	Most people in the world will look at generative AI and dismiss it And i could not be happier that means I have a small headstart to learn more about it get into better roles and build better business
220	@Daniel_Rubino Very weird. I wouldn't be surprised if Apple "invents" AI after the issues are worked out with Bing Chat, ChatGPT, and Bard though.	Very weird I would not be surprised if Apple invents AI after the issues are worked out with Bing Chat ChatGPT and Bard though

Removing numbers

47	RT @rowancheung: AI prompting is the best skill to learn right now. Companies are now paying up to \$335,000/year for Prompt Engineers. H...	RT AI prompting is the best skill to learn right now Companies are now paying up to year for Prompt Engineers H
95	🔥 Hey Guys, #ZenithSwap has launched at just \$ 55.000 USD Marketcap. The ChatGPT of DEX - Reimagining DeFi with AI-Powered Yield Farming. Now at 4X. Lot of up potential at such low marketcap. 🔥😊 \$ARB \$ZSP #Arbitrum https://t.co/eWLsdVy3Sm	Hey Guys has launched at just USD Marketcap The ChatGPT of DEX Reimagining DeFi with AIPowered Yield Farming Now at X Lot of up potential at such low marketcap ARB ZSP

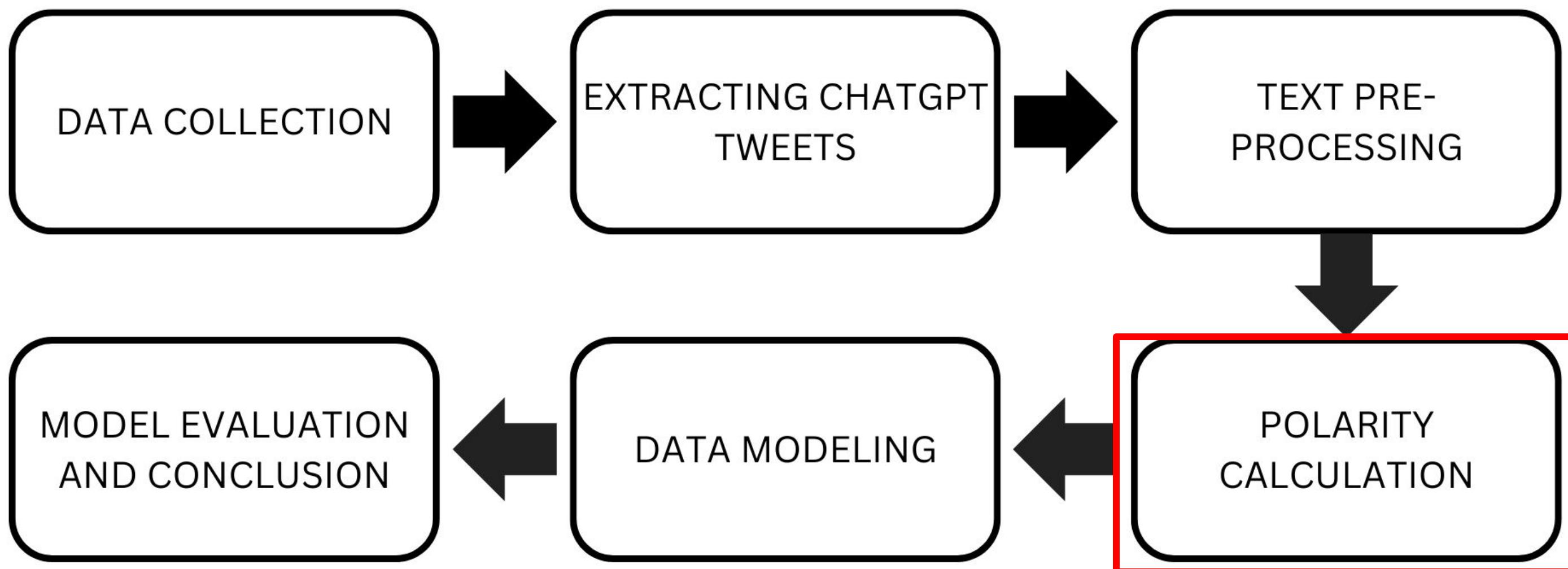
TEXT ANALYSIS OF TWEETS

Cleaning and Pre-processing textual data with NeatText library

Stopwords removal and changing to lower case for consistency and standardization

	text	clean_tweet
2	RT @DarrellLerner: ChatGPT Plugins are the fas...	rt chatgpt plugins are the fastest way to get ...
3	Get an intelligent chatbot for your website in...	get an intelligent chatbot for your website in...
4	🔥Hey Guys, #ZenithSwap has launched at just \$...	hey guys has launched at just usd marketcap ...
5	RT @sinsonetwork: Now! Join #SINSO DataLand^Ch...	rt now join datalandchatgpt taskslog in toltr...
9	The plagiarism detector will introduce its #AI...	the plagiarism detector will introduce its det...
...
40995	RT @solanaturbo: Hey everyone, have you heard ...	rt hey everyone have you heard about the new m...
40996	Can ChatGPT write my SoHo House application? W...	can chatgpt write my soho house application wr...
40998	RT @YGPT_official: 🚀 YGPT LAUNCH 🚀\n\nTime to ...	rt ygpt launch time to show you what weve be...
41000	RT @chatgpt_issac: Lets see which community is...	rt lets see which community is bigger and im g...
41001	@simamaung ChatGPT	chatgpt

METHODOLOGY



Polarity Score using VADER LEXICON

Concept: Measures the sentiment of a text, ranging from -1 (most negative) to +1 (most positive).

- 1. Tokenization** - split into individual words or tokens
- 2. Lexicon lookup** - VADER has a predefined list of words, each associated with a sentiment score
- 3. Compound Score Calculation** – the total score is normalized using a mathematical squashing function.
- 4. Adjustments** - VADER considers entire context of the sentence.

VADER LEXICON - VADER (Valence Aware Dictionary and sEntiment Reasoner)

Using `SentimentIntensityAnalyzer`, we can compute for the polarity scores

Comp Score: Positive

		clean_tweet	sentiment	compound	comp_score
2		rt chatgpt plugins are the fastest way to get rich in ive created a stepbystep guide showing you how to earn	{'neg': 0.0, 'neu': 0.763, 'pos': 0.237, 'compound': 0.6808}	0.6808	pos
3		get an intelligent chatbot for your website in minutes with chatbase ai train chatgpt on your data and let it answer any question your users have simply upload a document or link and add the chat widget its that easymake money using ai	{'neg': 0.0, 'neu': 0.932, 'pos': 0.068, 'compound': 0.4588}	0.4588	pos

Comp Score: Negative

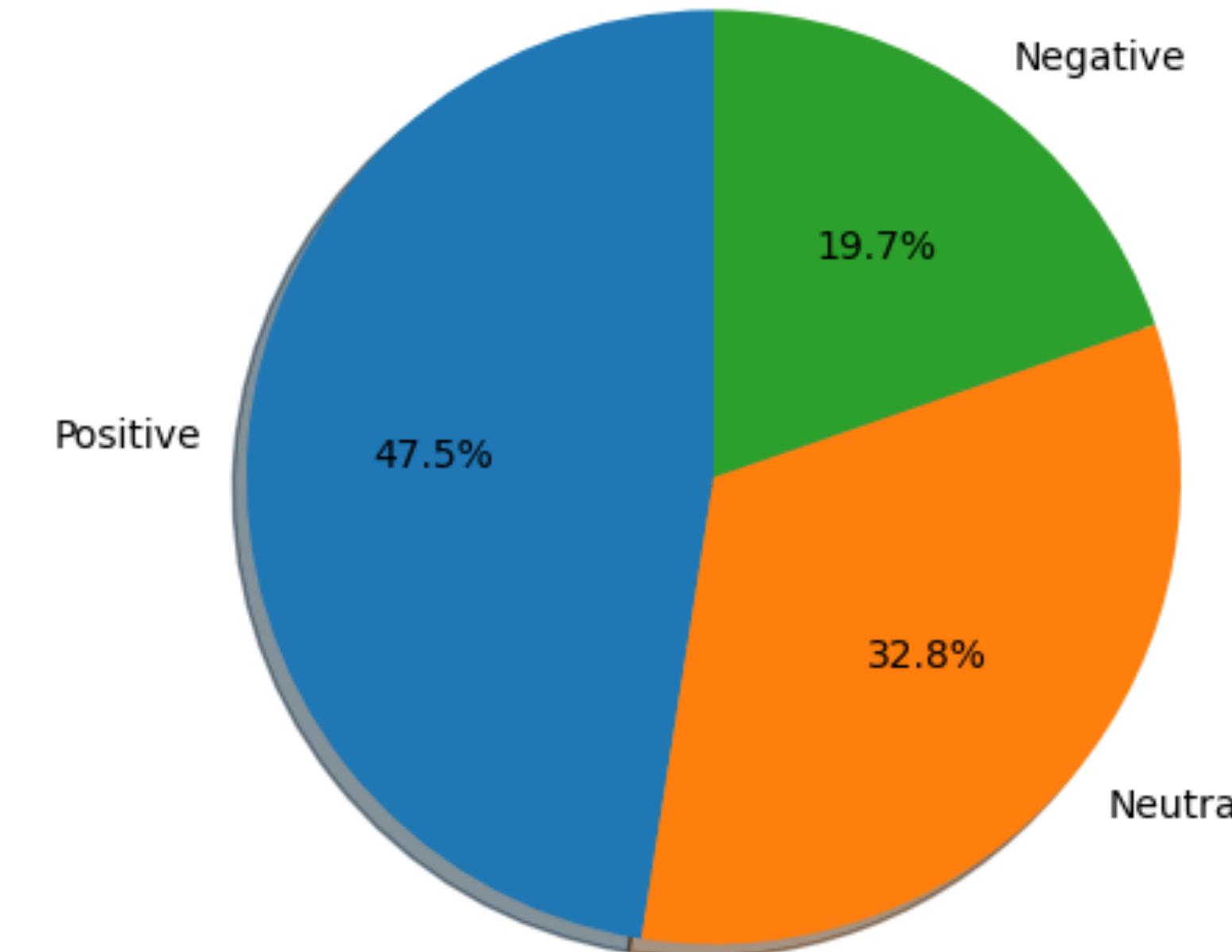
index		clean_tweet	sentiment	compound	comp_score
10		germany could follow in italys footsteps by blocking chatgpt over data security concerns	{'neg': 0.163, 'neu': 0.688, 'pos': 0.15, 'compound': -0.0516}	-0.0516	neg

Comp Score: Neutral

index		clean_tweet	sentiment	compound	comp_score
64		rt the rise of the developer	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}	0.0	neu

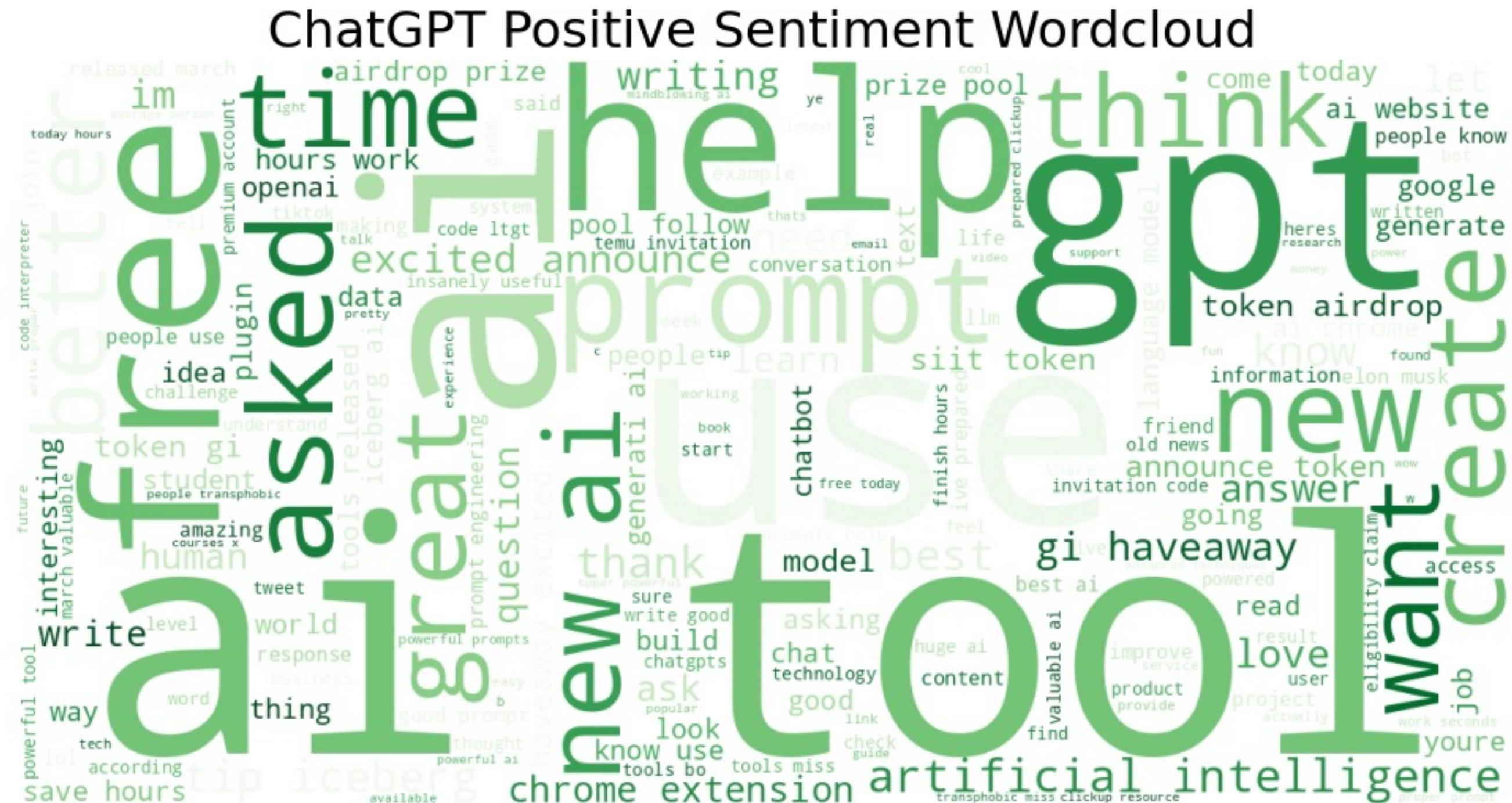
SENTIMENT LEVELS

Percentage of Sentiment Levels in Tweets



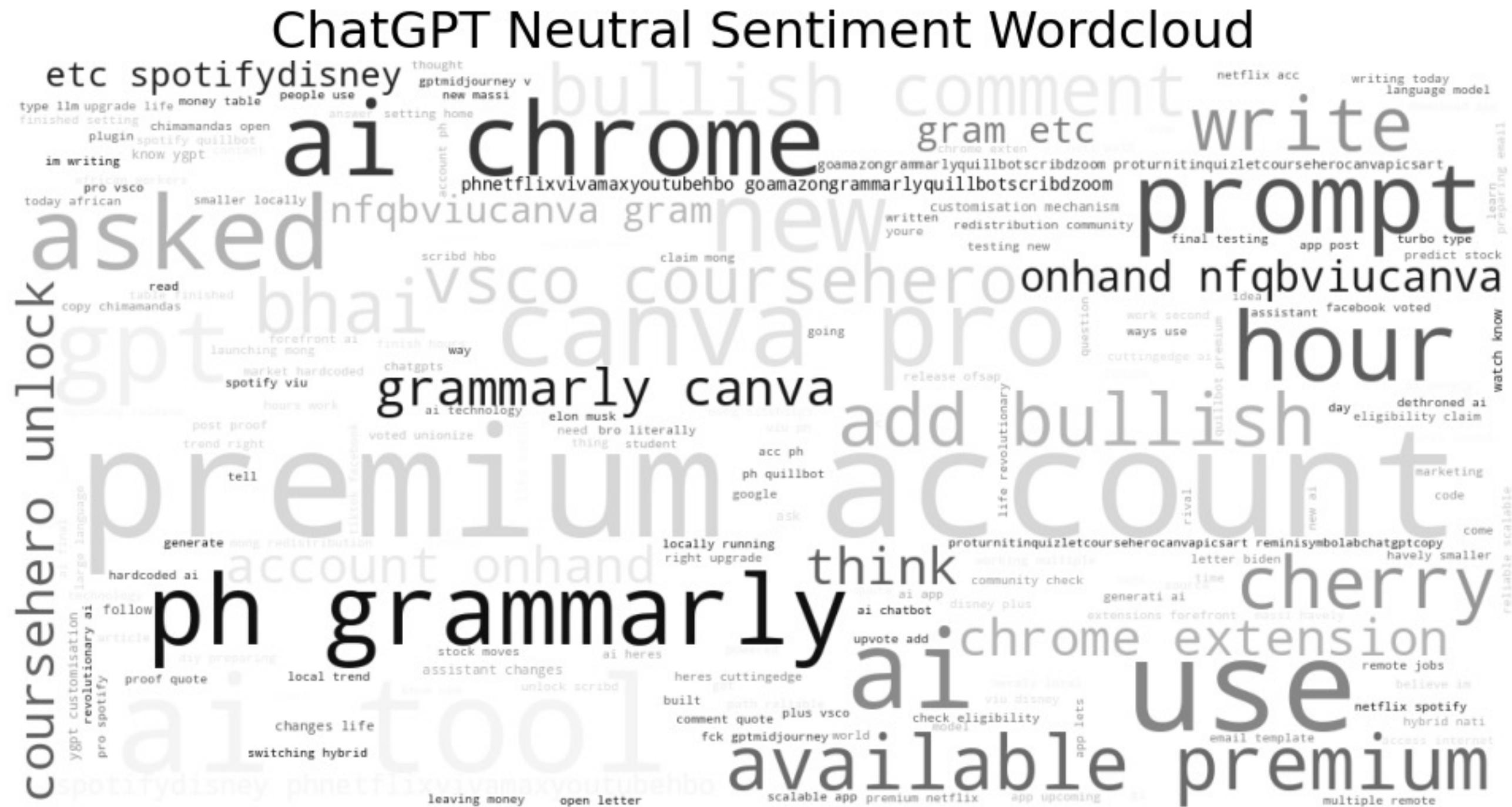
WORD CLOUD

- visually represents text data
 - size of each word in the visualization indicates its frequency or importance



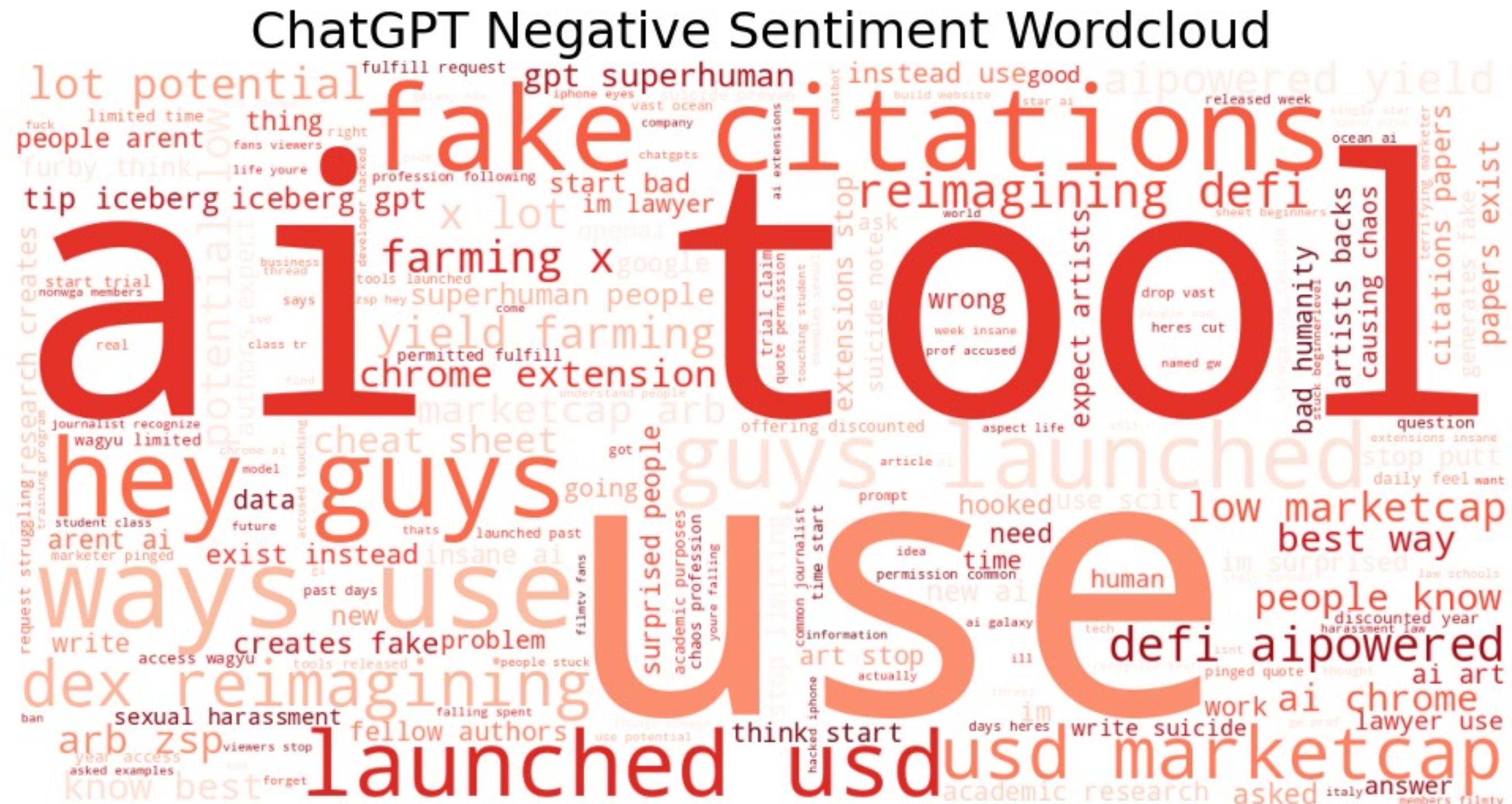
WORD CLOUD

- visually represents text data
 - size of each word in the visualization indicates its frequency or importance.



WORD CLOUD

- visually represents text data
 - size of each word in the visualization indicates its frequency or importance.



FEATURE ENGINEERING



TF-IDF

- Term Frequency-Inverse Document Frequency
- Gives higher weight to words that are frequent in a document, emphasizing words that are more unique and relevant

Example:

Imagine you're in a library and you want to find books on "dinosaurs":

TF (Term Frequency): This measures how frequently a term appears in a document. So if "dinosaurs" appears 10 times in a 100-word book, its TF is 0.1 (or 10%).

IDF (Inverse Document Frequency): This measures how important a term is. If every book in the library was about "dinosaurs", then the term wouldn't be special or useful to distinguish books on the topic. But if only a few books are about "dinosaurs", then it's a significant term.

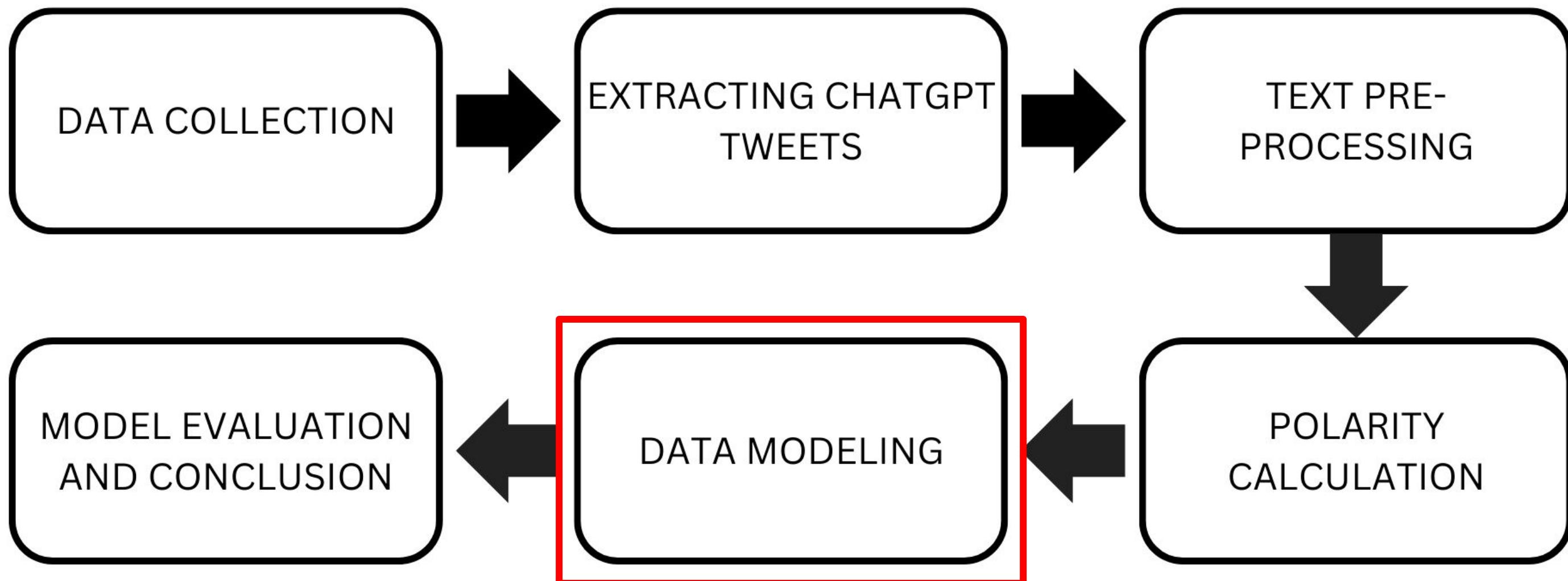
INPUT AND TARGET VARIABLES

Train test split - method where data is divided into two subsets: one to train a machine learning model and the other to test its performance.

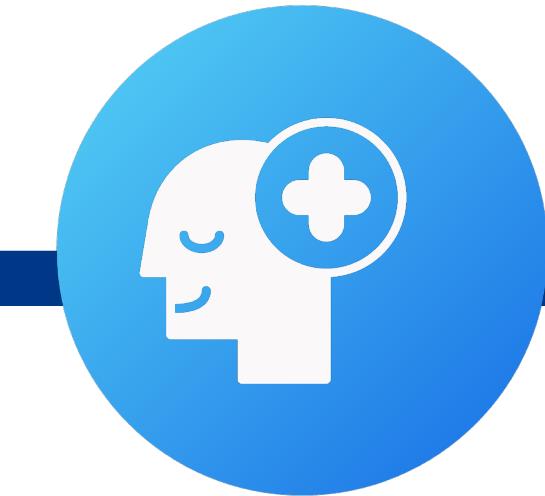
Compound score – target variable

Clean tweets – input variable

METHODOLOGY



DATA MODELING



SMOTE

- Synthetic Minority Over-sampling Technique used to address class imbalance problems in the dataset

Example:

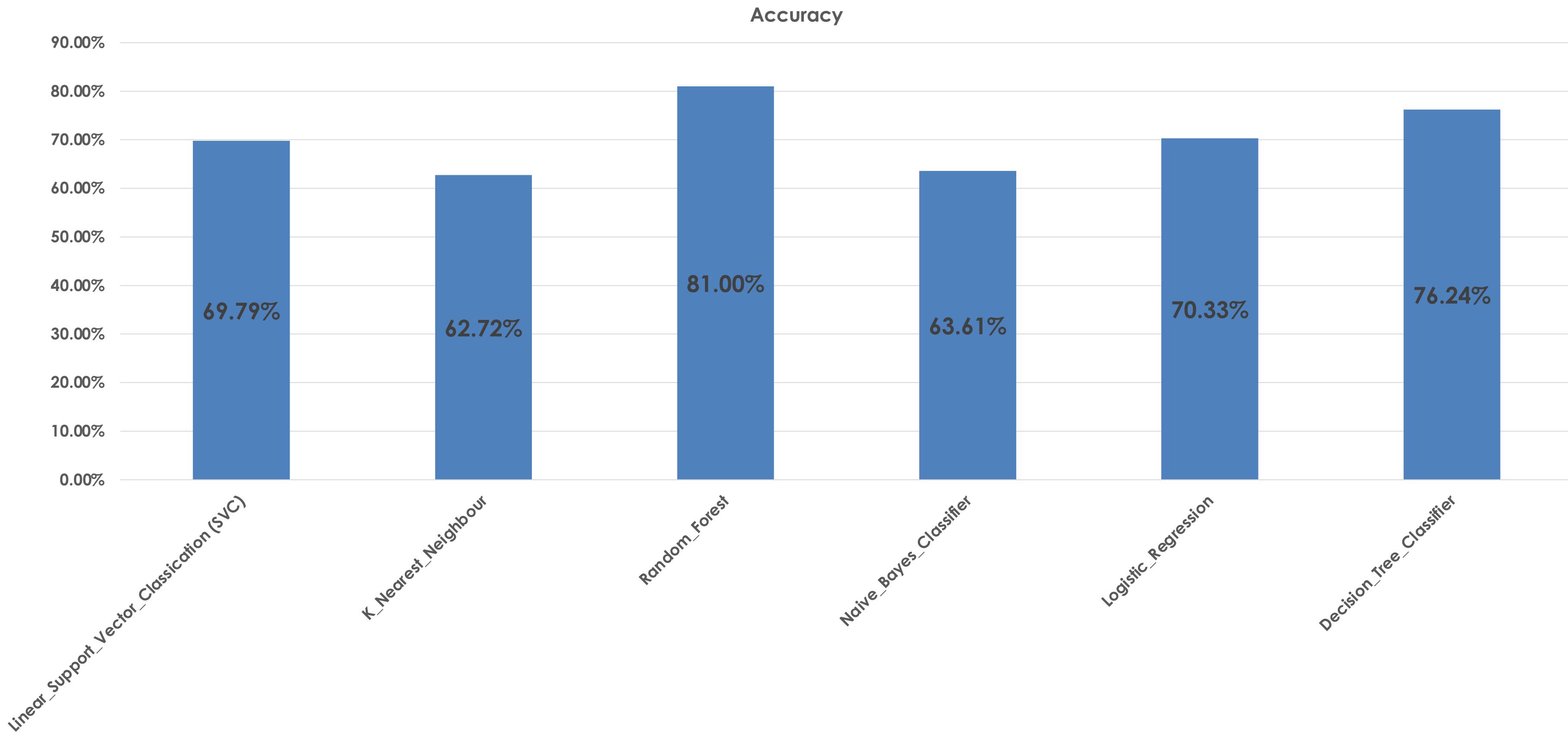
Imagine a school's basketball team trying out players wherein out of 100 students trying out, 95 are tall and 5 are short.

SMOTE helps by creating "synthetic" short players.

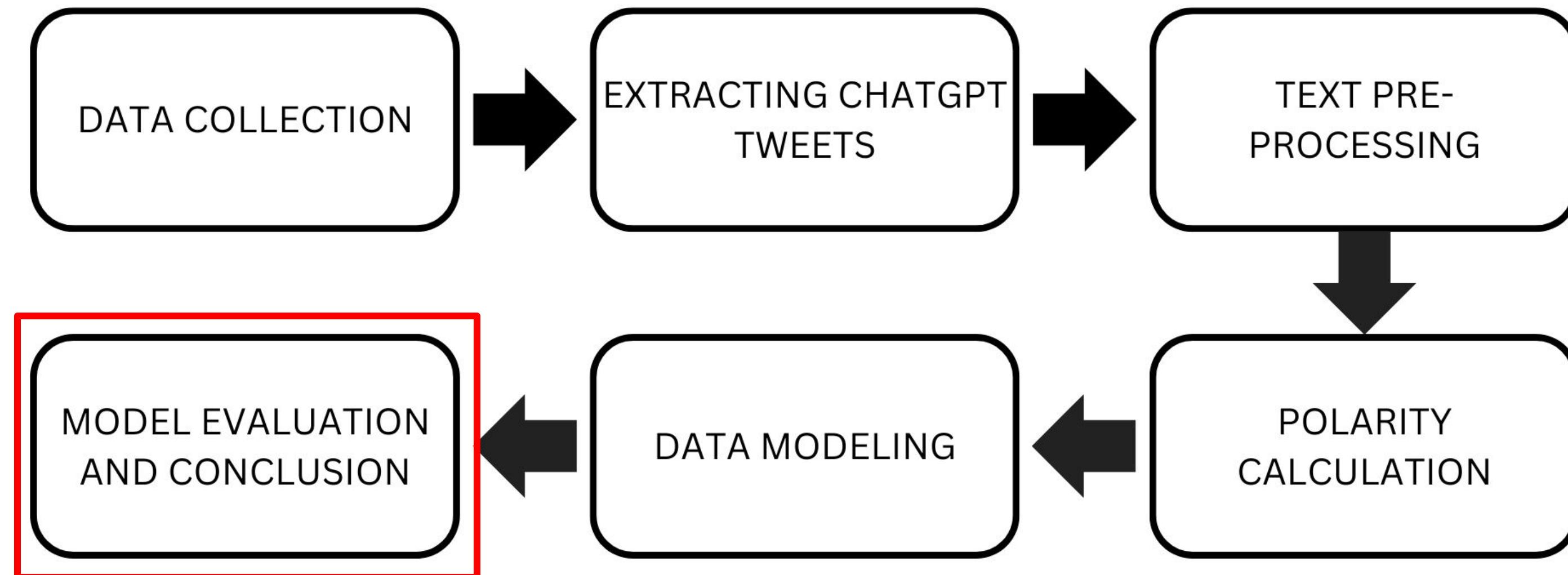
- 1.Pick a short player from the 5.
- 2.Choose one of the nearest other short players.
- 3.Create a new, synthetic short player that's a blend between the two.

For instance, if student A is 5'2" and student B is 5'4", **our synthetic student might be 5'3".**

DATA MODELING



METHODOLOGY



MODEL EVALUATION

• **Overall Accuracy:** At 81%, the model effectively classified a significant portion of the dataset into their correct sentiment classes.

• **Analysis:**

- For **Negative** sentiments, the model predicted it with a precision of 81%, indicating that the majority of tweets labeled as negative were indeed negative. However, a recall of 66% suggests that some negative tweets might have been missed.
- Neutral** sentiments, both precision and recall were above 80%, showing consistent performance in identifying and classifying neutral tweets.
- Positive** sentiments had the highest recall at 86%, implying that the model was particularly adept at catching most of the positive tweets. The precision for positive sentiments was consistent with the other classes at 81%.

• **F1-Score:** The balanced F1-scores across the classes, especially the 0.83 for positive sentiments, further underline the model's effectiveness.

Classification Summary

	Precision	Recall	F1-Score	Support
neg	0.81	0.66	0.73	812
neu	0.81	0.83	0.82	1301
pos	0.81	0.86	0.83	1830
accuracy			0.81	3943
macro	0.81	0.78	0.79	3943
weighted	0.81	0.81	0.81	3943

While the model has shown significant prowess, especially in identifying positive sentiments, there's **room for improvement** in **capturing all negative sentiments**, as indicated by the recall for the 'neg' class. Future iterations could focus on enhancing this aspect for even more balanced performance.

CONCLUSION

- The majority of tweets related to ChatGPT convey a **positive** reception, indicating user satisfaction and endorsement.
- Neutral** sentiments represent objective discussions, inquiries, or areas of ambivalence among users.
- The presence of **negative** sentiments, while lesser, highlights areas for improvement and the necessity to address user concerns.
- Sentiment suggests ChatGPT is well-received on X/Twitter, but continuous engagement and responsiveness to feedback remain paramount.

RECOMMENDATIONS

User Feedback System:

Implement a direct feedback mechanism on the platform. This allows users to provide insights on what they like and areas they feel need improvement.

Interactive Tutorials:

Introduce interactive tutorials or webinars to guide both new and existing users on maximizing their experience with ChatGPT.

Engage with Users:

Given that a significant portion of the tweets are positive and the model is good at identifying them, it suggests that users generally have a positive sentiment towards ChatGPT. Engaging with these users and encouraging them to share their positive experiences can boost the platform's reputation.

User Engagement:

Host Q&A sessions, webinars, or interactive events based on the feedback from these tweets. Given the high number of positive and neutral sentiments, there's an opportunity to further engage the community and gather more insights.

**THANK
YOU**