

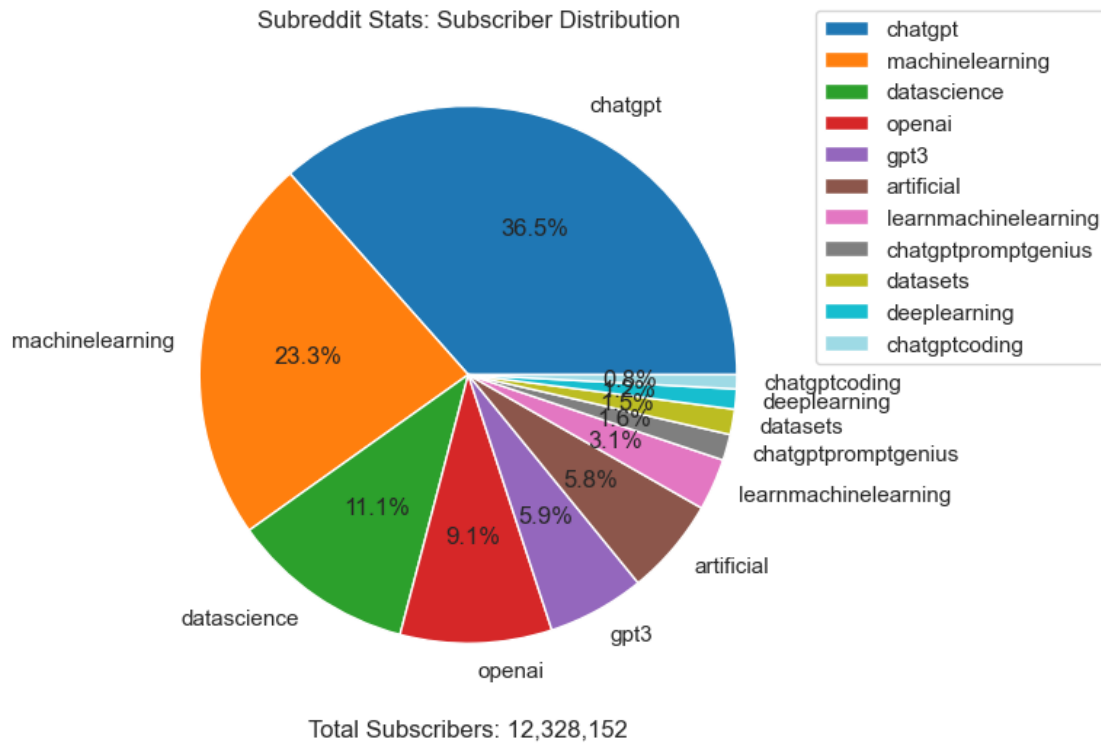
# sentiment-analysis

February 28, 2024

## 1 Reddit Subreddit API analysis

Subreddit Stats

name	subscribers	subscribers_pct
chatgpt	4505500	36.55
machinelearning	2870456	23.28
datascience	1366348	11.08
openai	1124730	9.12
gpt3	722076	5.86
artificial	720095	5.84
learnmachinelearning	384514	3.12
chatgptpromptgenius	193280	1.57
datasets	187702	1.52
deeplearning	149090	1.21
chatgptcoding	104361	0.85



## 2 Reddit Search API data analysis

Total imported Reddit posts: 9822

Duplicate cross-subreddit posts: 7345

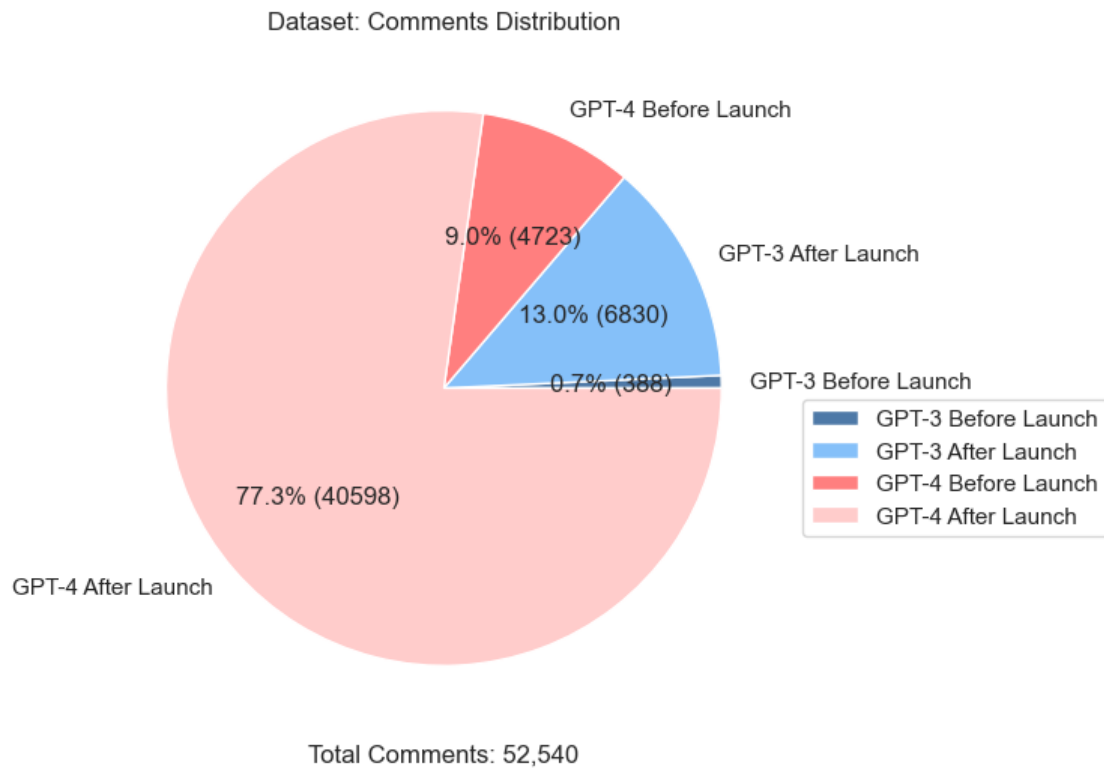
De-duplicated reddit posts: 2477

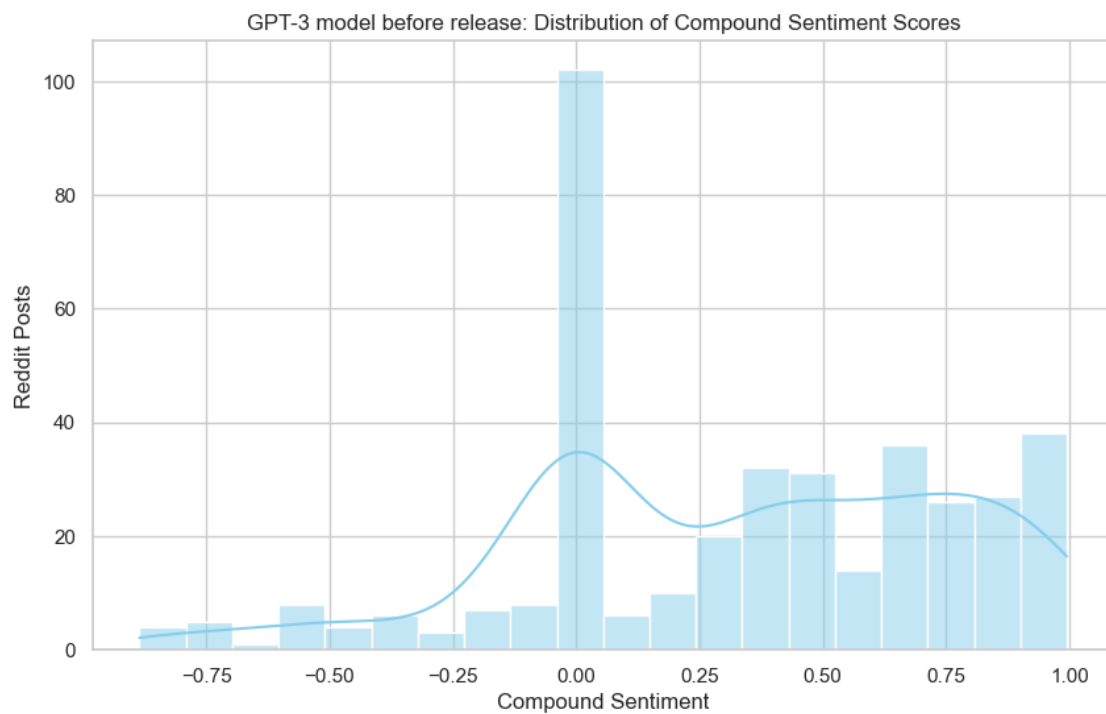
### 3 Reddit Comment API data analysis

### 4 Sentiment analysis

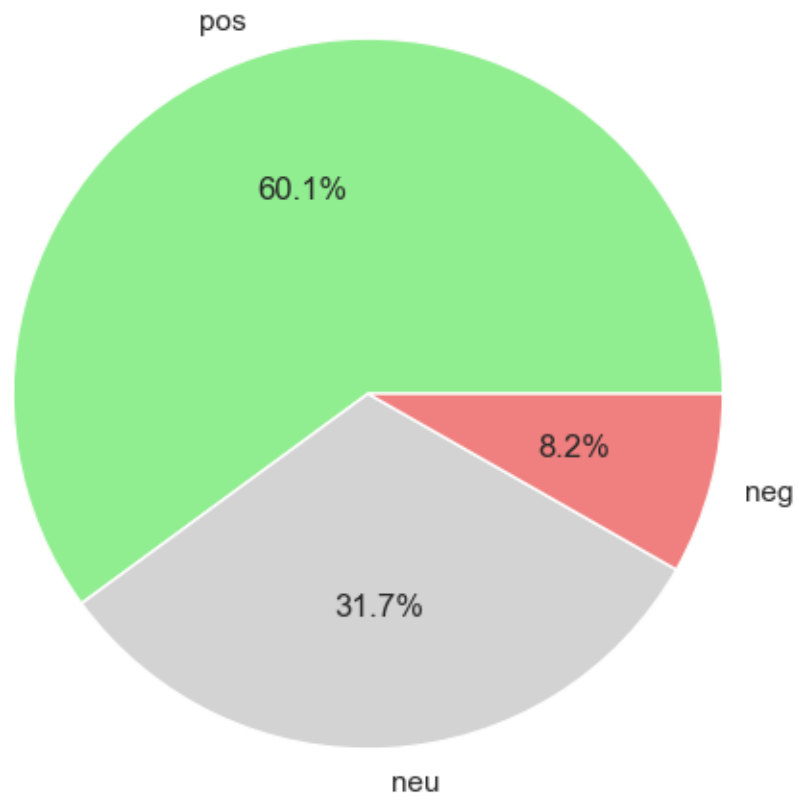
### 5 Data preparation

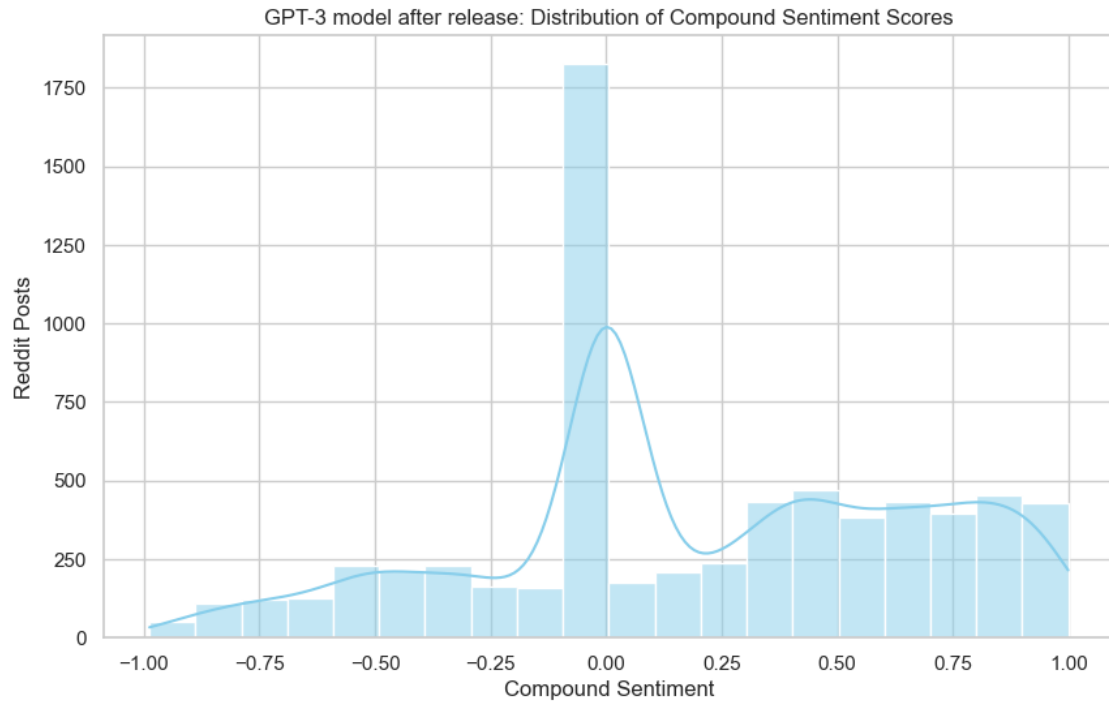
### 6 Dataset split on launch dates



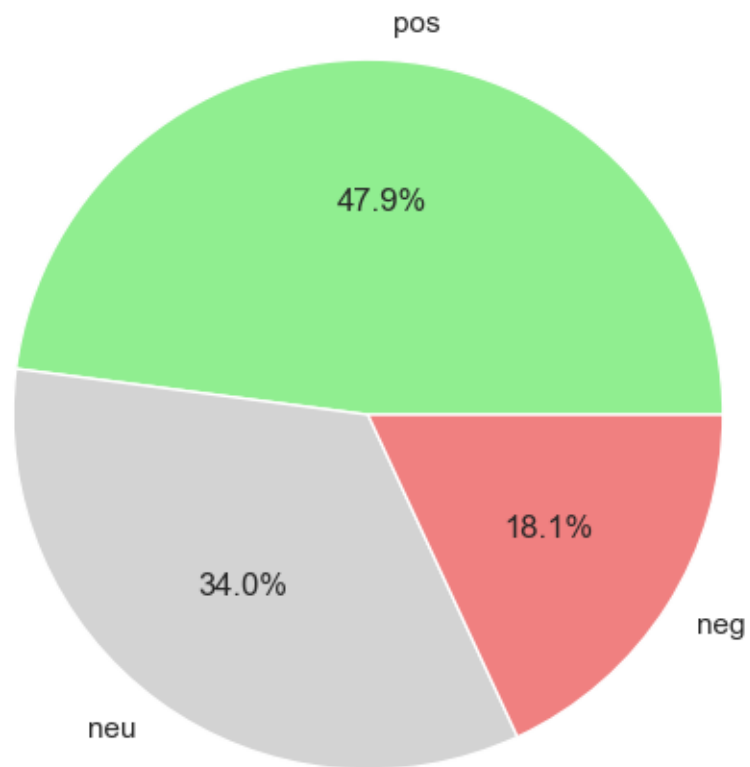


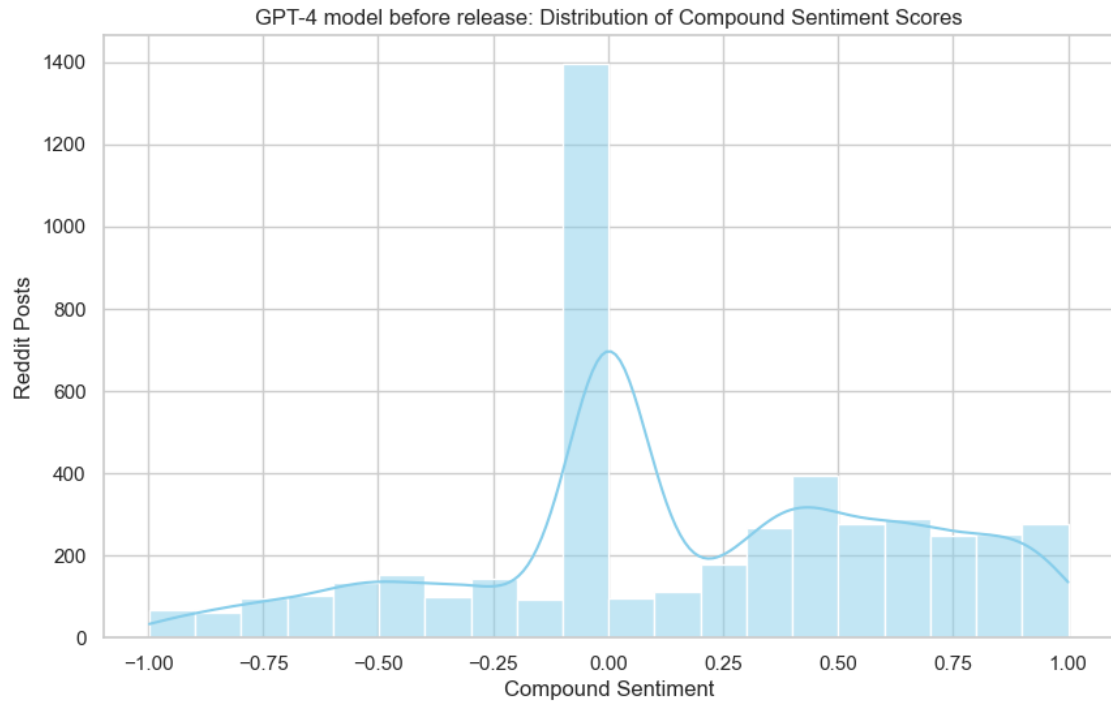
GPT-3 model before release: Sentiment Distribution





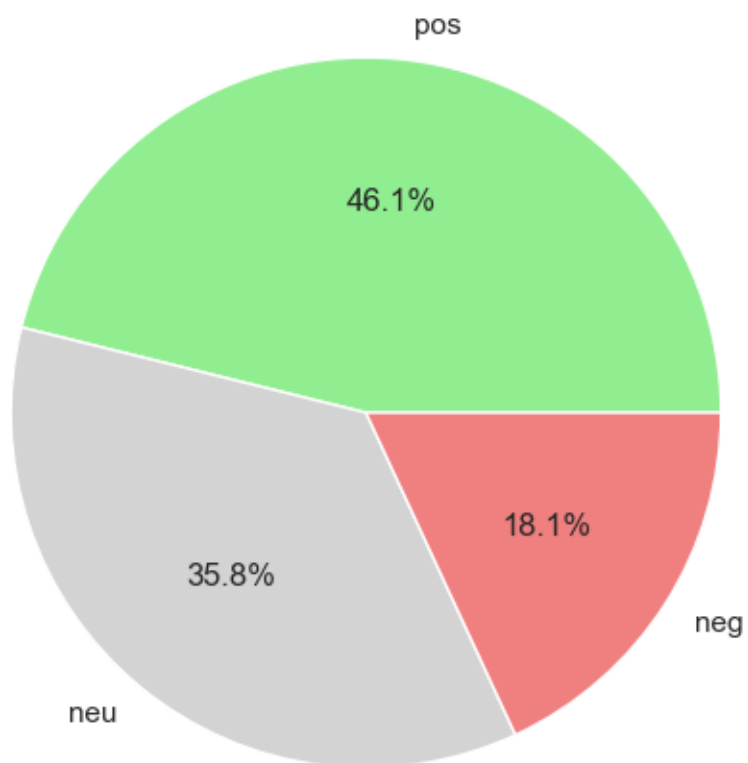
GPT-3 model after release: Sentiment Distribution

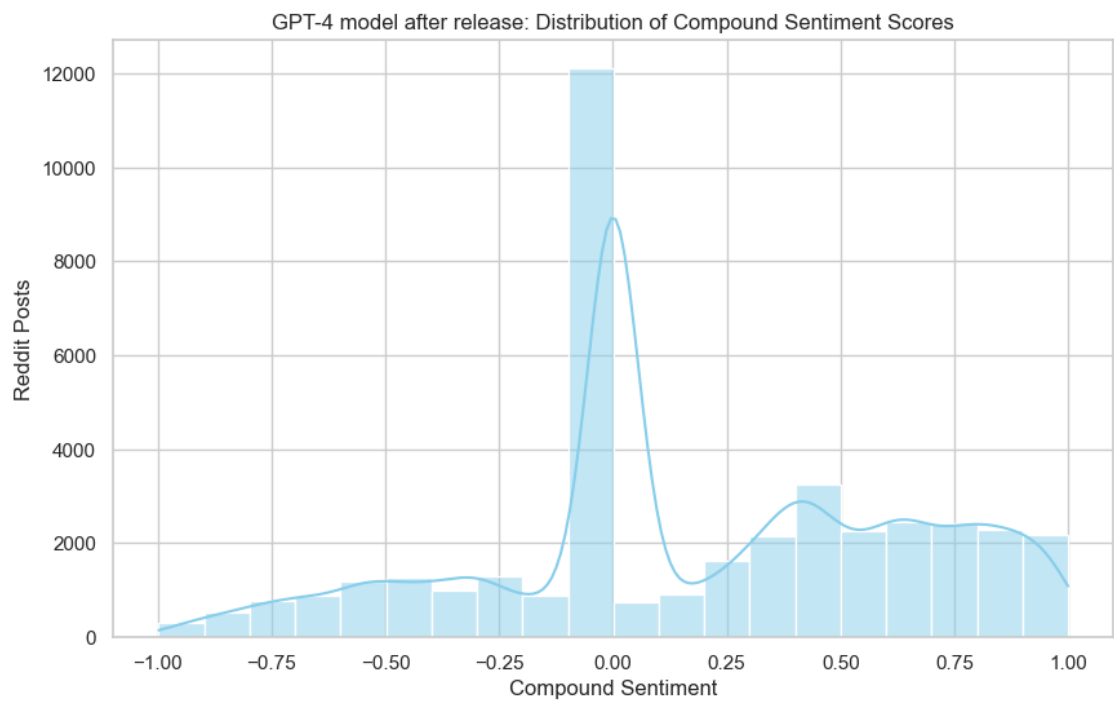




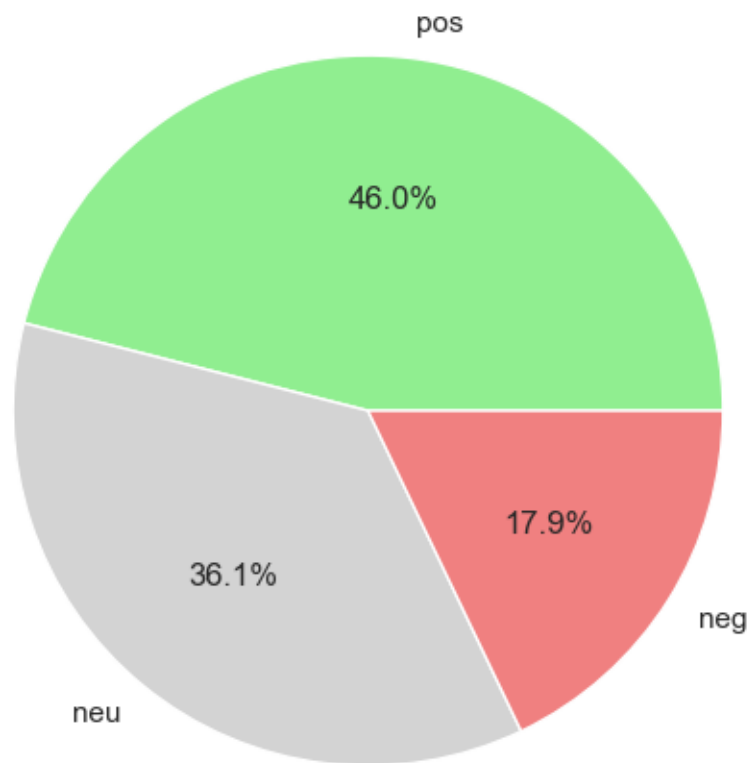


GPT-4 model before release: Sentiment Distribution





GPT-4 model after release: Sentiment Distribution



## 7 Categorical Distributions

