sentiment-analysis

February 26, 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     |
+--------------------+------------+----------------+
```

Subreddit Stats: Subscriber Distribution

Total Subscribers: 12,328,152
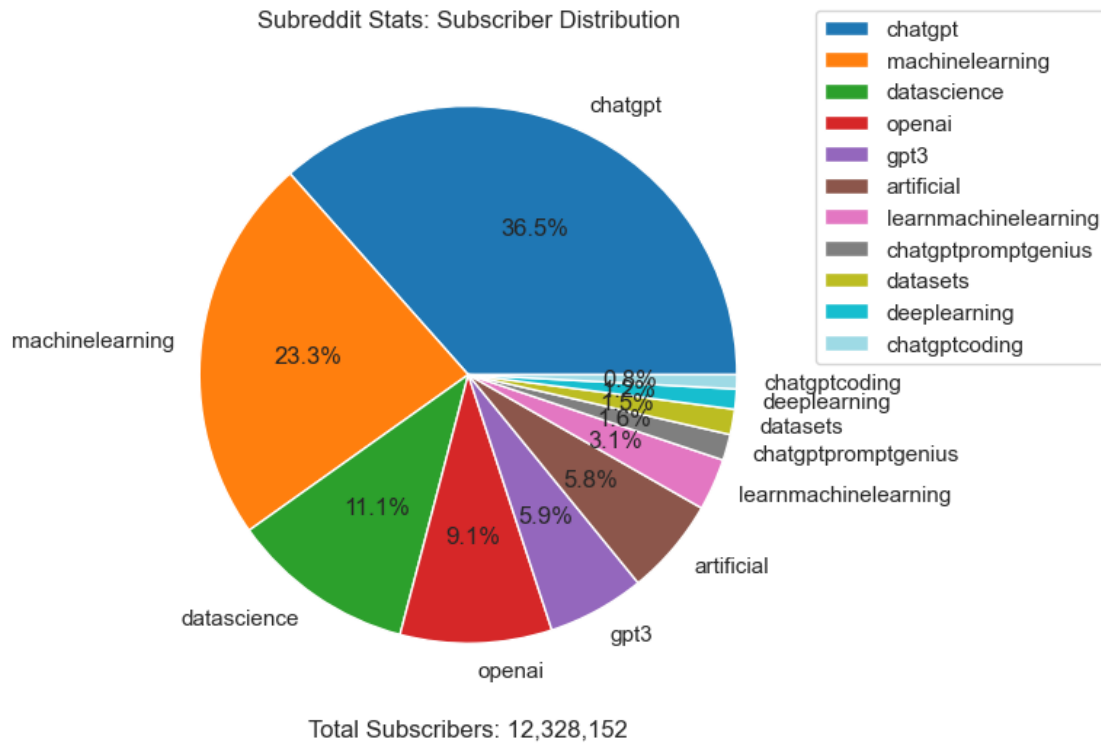
2 Reddit Search API data analysis

```
Total imported Reddit posts: 9822
Duplicate cross-subreddit posts: 7345
De-duplicated reddit posts: 2477
```
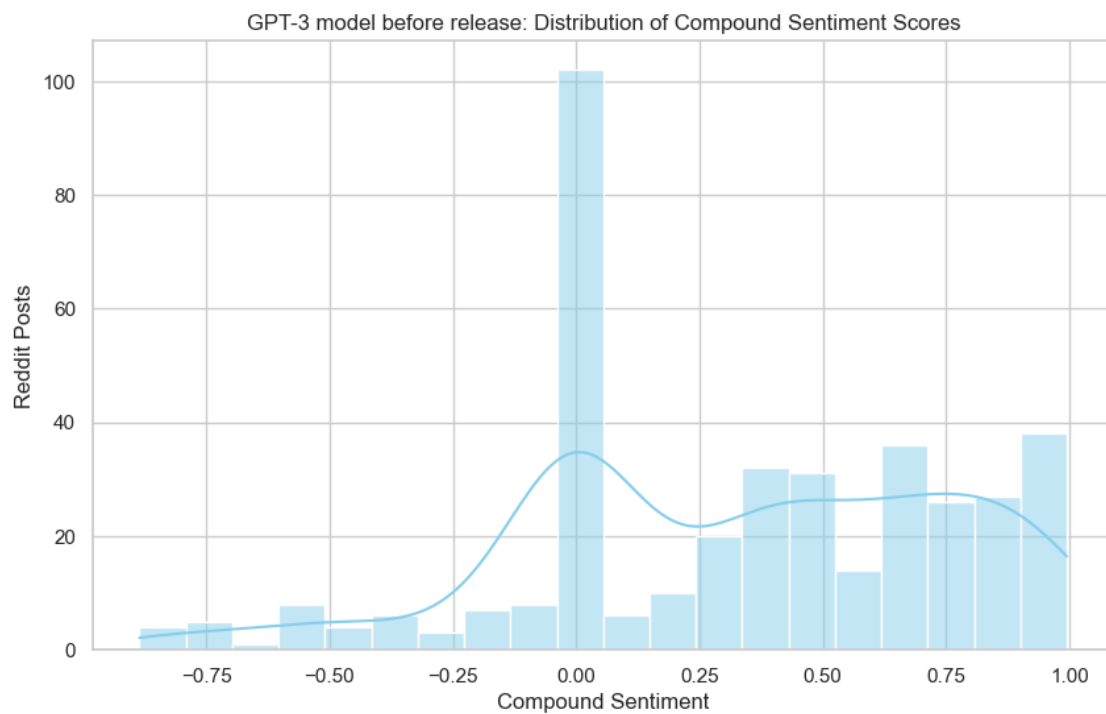
Dataset: Comments Distribution

GPT-4 Before Launch

GPT-3 After Launch

9.0% (4723)

13.0% (6830)

0.7% (388)   GPT-3 Before Launch

■	GPT-3 Before Launch
■	GPT-3 After Launch
■	GPT-4 Before Launch
■	GPT-4 After Launch

77.3% (40598)

GPT-4 After Launch

Total Comments: 52,540

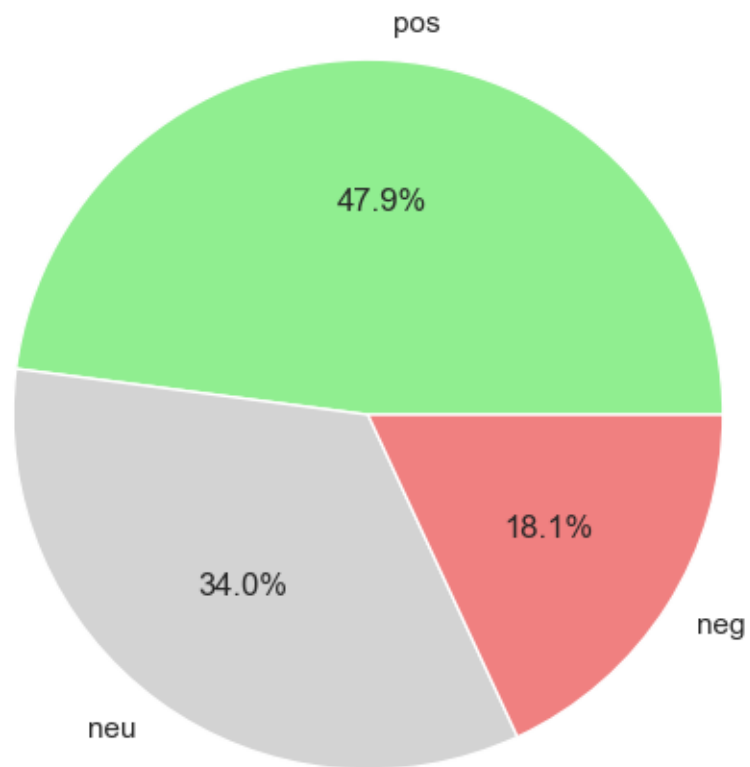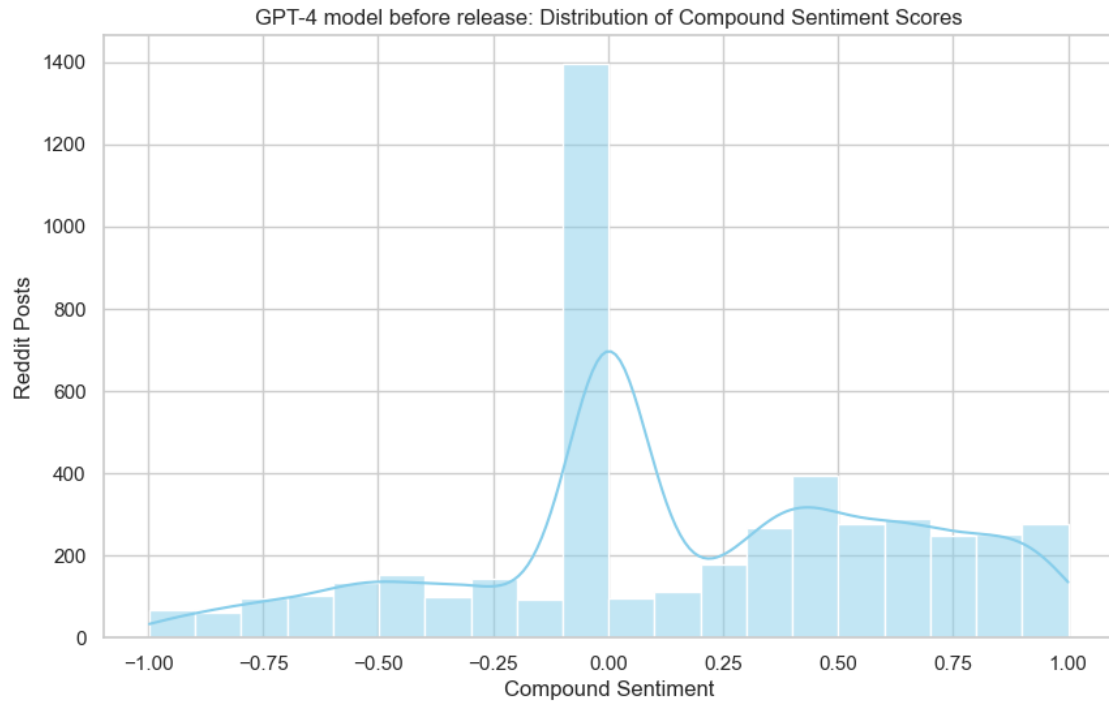GPT-3 model before release: Distribution of Compound Sentiment Scores
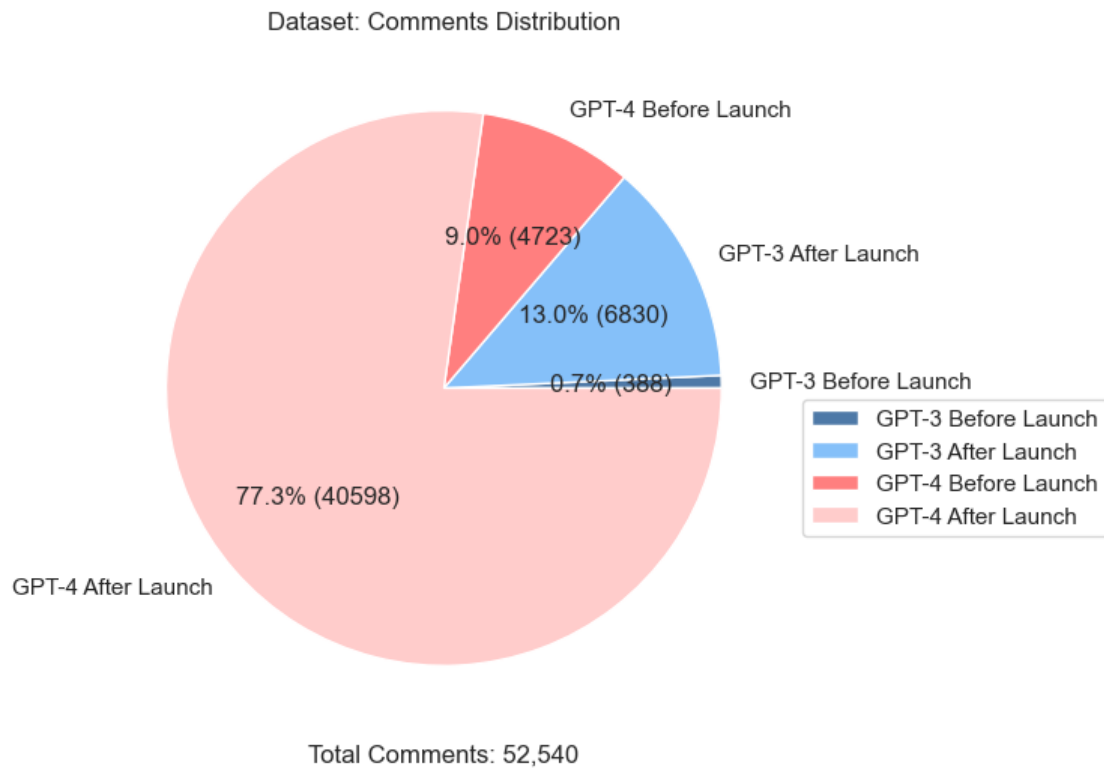
GPT-3 model before release: Sentiment Distribution

GPT-3 model after release: Distribution of Compound Sentiment Scores

GPT-3 model after release: Sentiment Distribution

GPT-4 model before release: Distribution of Compound Sentiment Scores
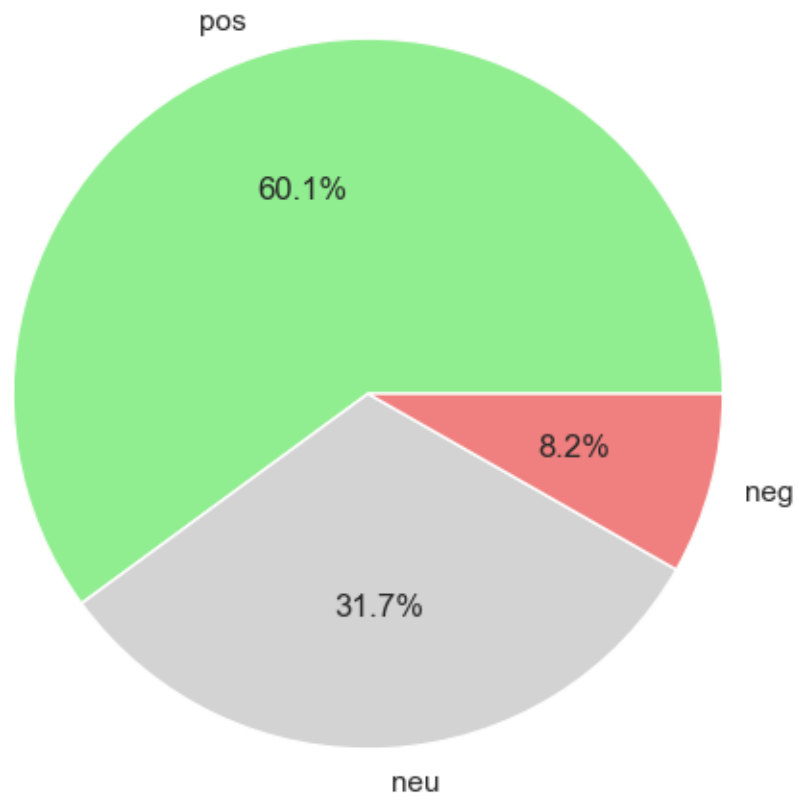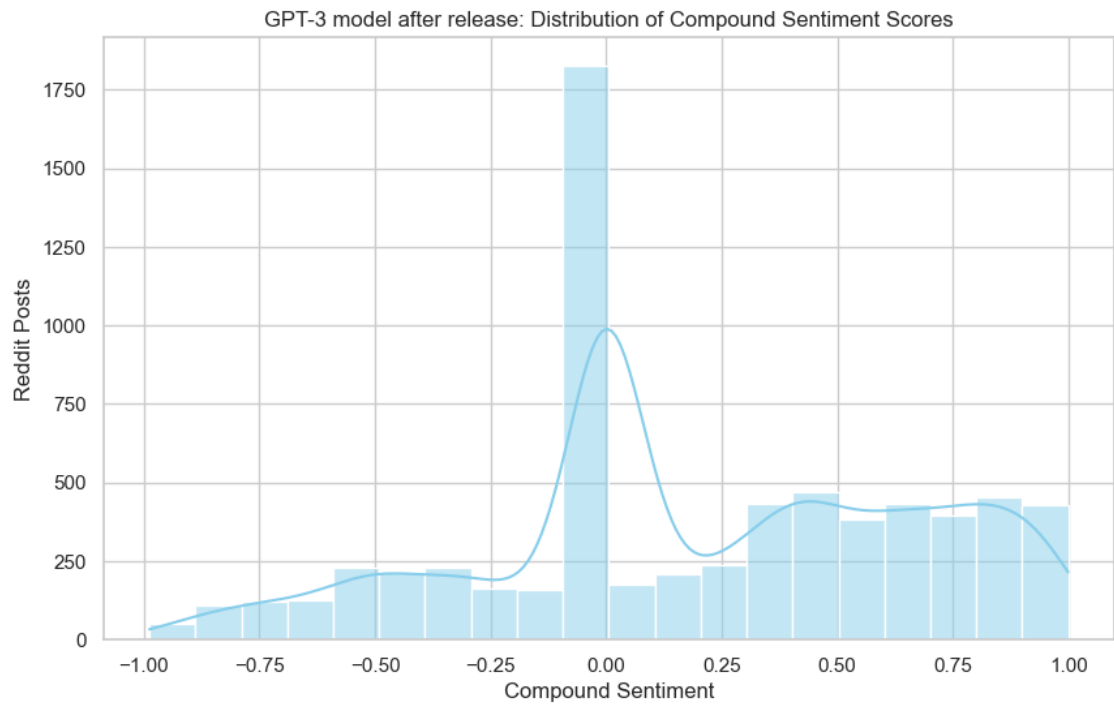
GPT-4 model before release: Sentiment Distribution

GPT-4 model after release: Distribution of Compound Sentiment Scores
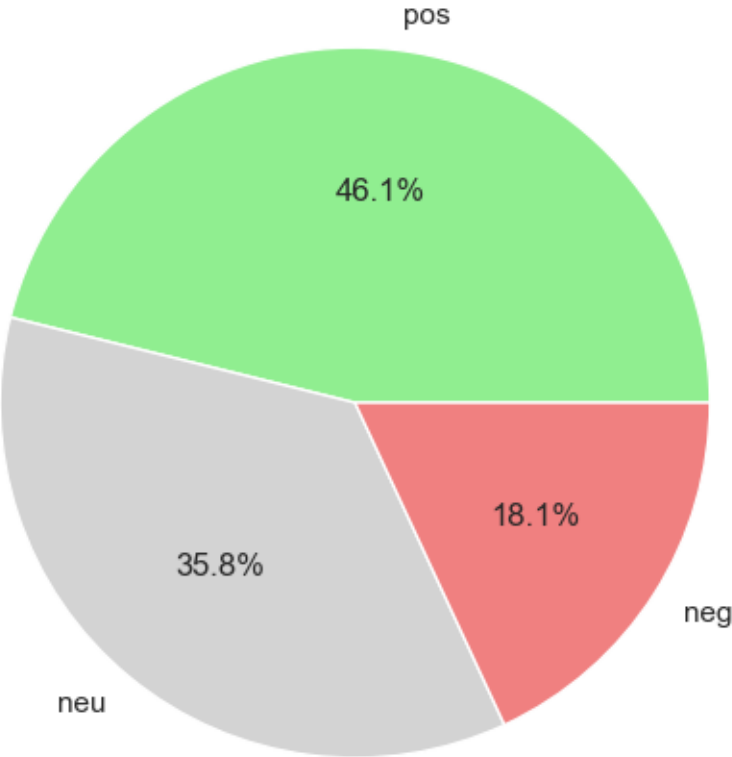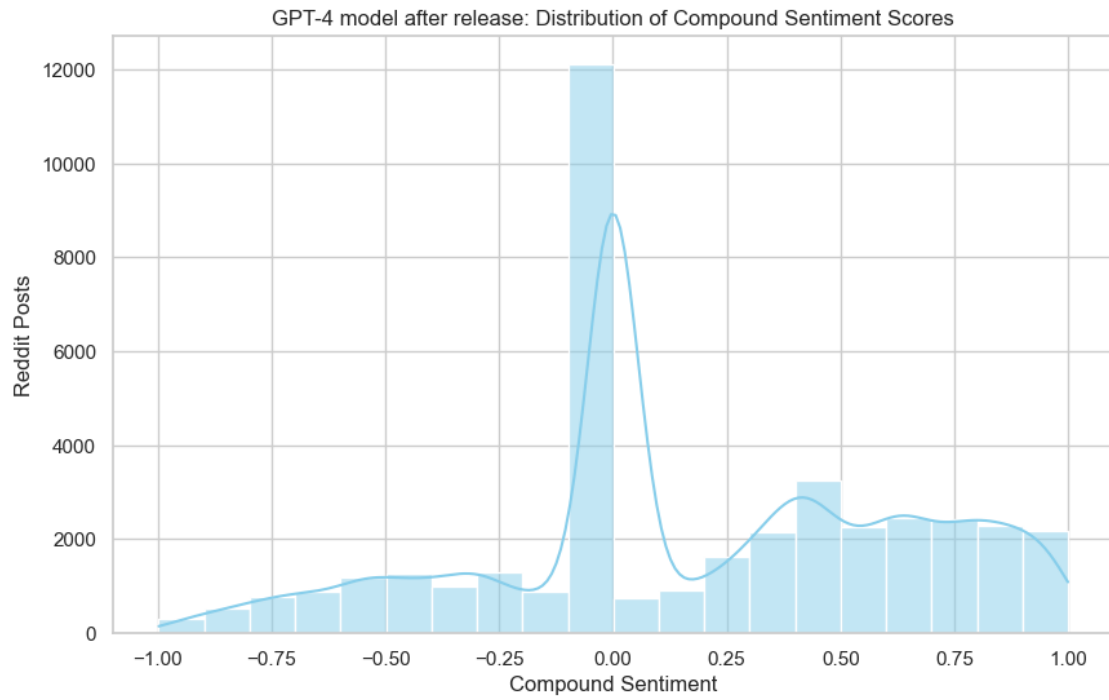
GPT-4 model after release: Sentiment Distribution
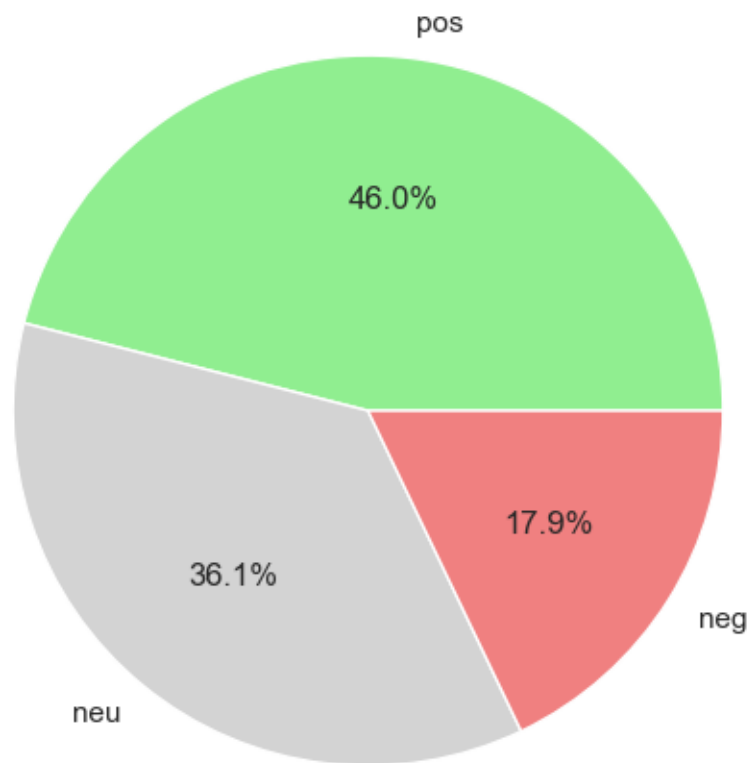
7   Categorical Distributions

Distribution of subreddit in GPT-3 model before announcement

Distribution of subreddit in GPT-3 model after announcement

Distribution of subreddit in GPT-4 model before announcement

Distribution of subreddit in GPT-4 model after announcement

Distribution of query in GPT-3 model before announcement

Distribution of query in GPT-3 model after announcement

Total: 6830

Distribution of query in GPT-4 model before announcement

Distribution of query in GPT-4 model after announcement