



Exploring Public Opinion on AI Applications in Healthcare and Medicine

Ilyas Yeskenov(24444243)

ECON7055: Projects for Data Analytics

8th May, 2025

Introduction

- AI technologies are widely adopted
- Main applications:
 - Medical imaging
 - AI diagnostics
 - GenAI as consultant
- Main issues:
 - Data privacy
 - Safety
 - Bias



Introduction

Project Question

How Public sentiment on AI
Applications in Healthcare and
Medicine change over time?



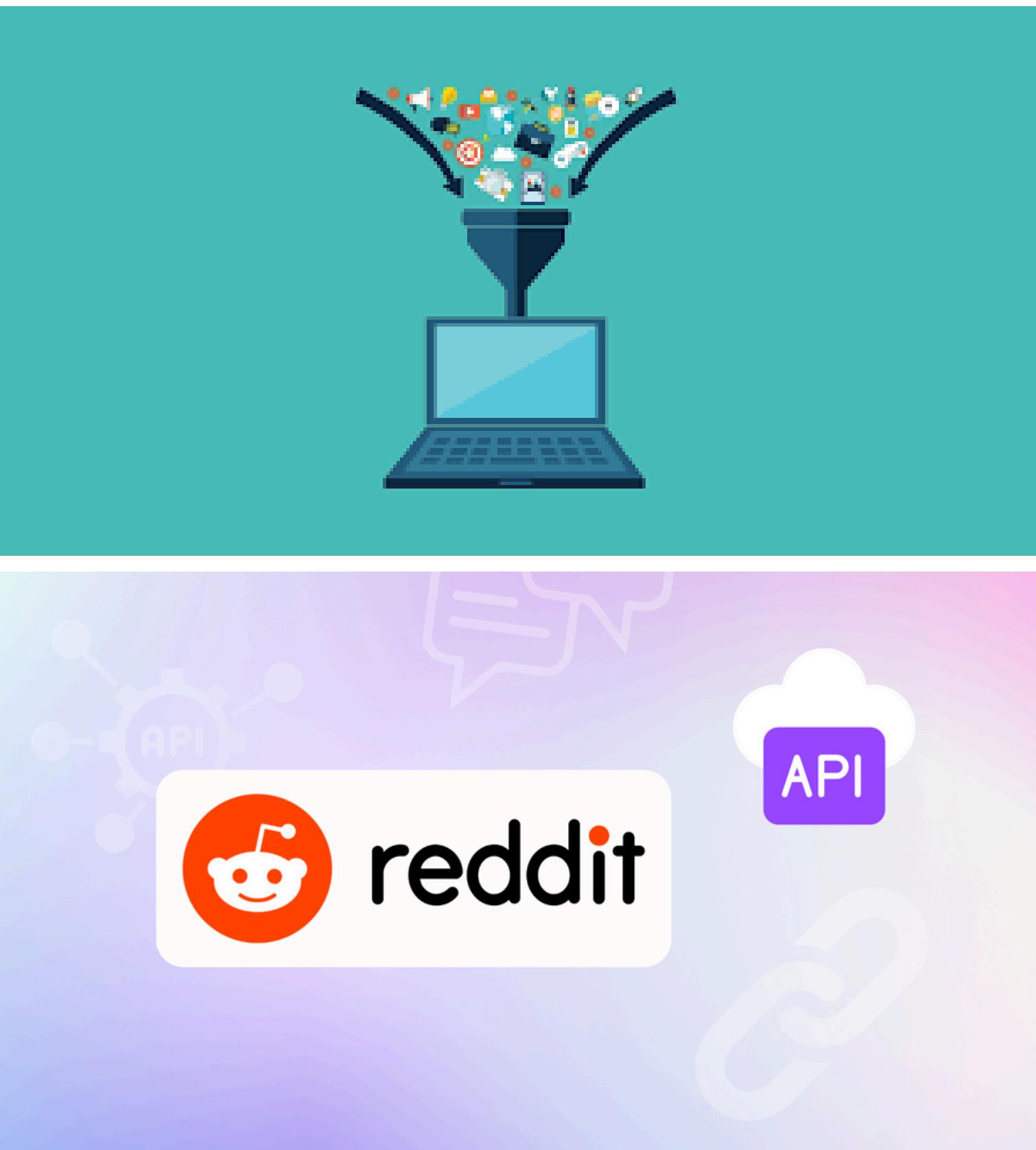
Interesting Aspects

1. Uncover general attitudes toward AI technologies.
2. Analyze public sentiment through social media insights.
3. Look how sentiment changes according to some events,

Data collection

- Scraping the comments from Reddit
- Using PRAW Python Library
- Focusing on the keywords like:
 - *Medical AI*
 - *AI in medicine*
 - *Trust in AI diagnostics*
- Merging everything into one dataframe
- Final dataset had 20972 comments
- 16 columns (time, score, user, ...)

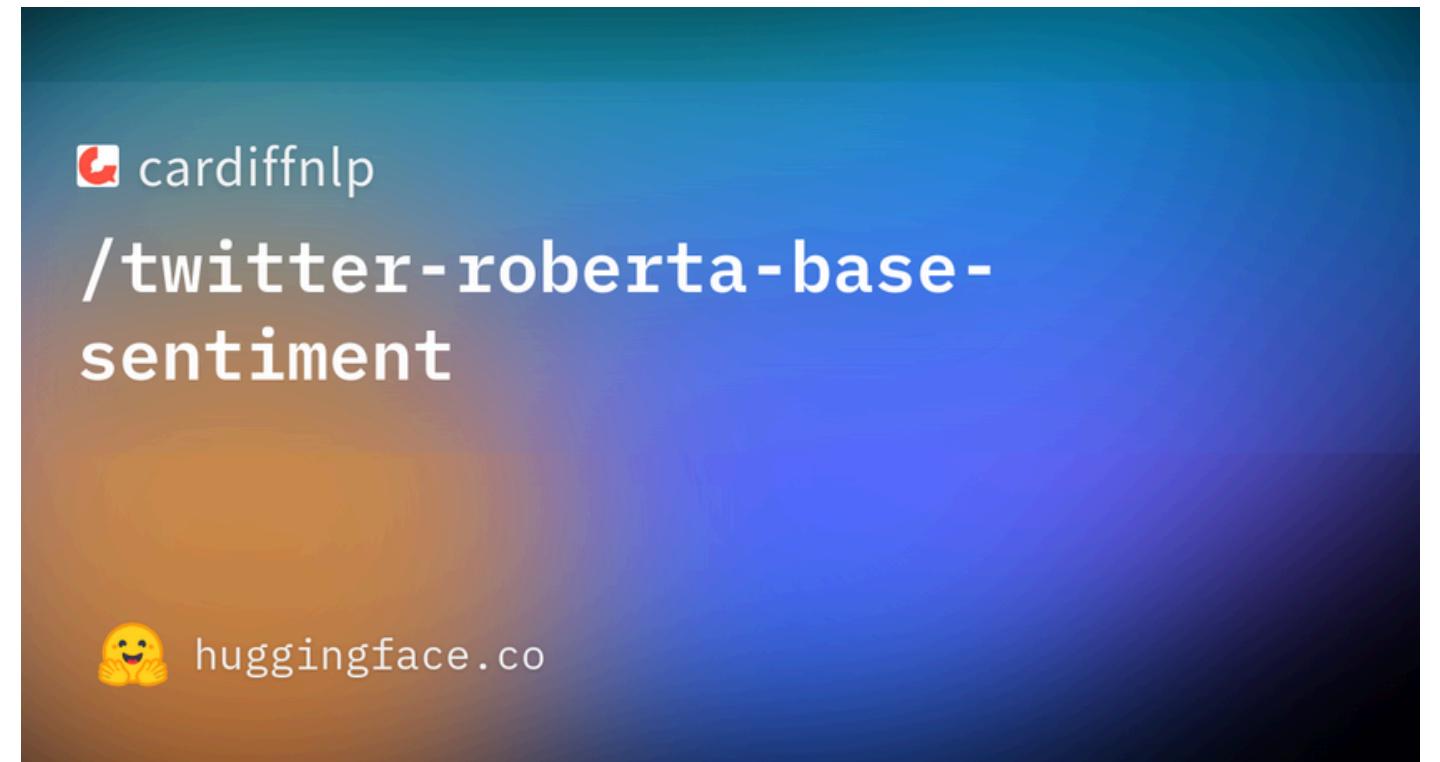
• • • •



Data labelling

1. Pre-trained model was applied (3 sentiments)
2. Sampling **3500** comments only from positive and negative comments.
3. Manual labelling with a friend (2000:1500)
4. The comment were on various topics

• • • •



- Trained on 58 million tweets
- Fine-tuned for TweetEval benchmark
- 3 sentiments

Text processing

1. Cleaning and Lowercasing

2. Tokenization

3. Lemmatization

4. Vectorisation (Scikit-learn):

a. TF - CountVectorizer()

b. TF-IDF - TfidfVectorizer()

c. LatentDirichletAllocation() with TF



Natural Language Analysis
with Python NLTK

$$w_{x,y} = tf_{x,y} \times \log \left(\frac{N}{df_x} \right)$$

TF-IDF

Term x within document y

$tf_{x,y}$ = frequency of x in y
 df_x = number of documents containing x
 N = total number of documents

Model training

- Logistic Regression (LR)
- Support Vector Machine (SVM)
- Random Forest (RF)
- Multinomial Naïve Bayes (MNB)
- XGBoost (XGB)
- Accuracy used as metrics
- Overall results are around **80%**
- LDA didn't perform good
- GridSearchCV for LR and MNB

	LR	SVM	RF	MNB	XGB
TF	0.791	0.785	0.805	0.824	0.784
TF-IDF	0.811	0.795	0.781	0.755	0.794
TF+LDA	0.63	0.644	0.628	0.63	0.644



Data analysis

- Negative - **71%**
- Positive - **28%**
- Medical AI - **63%**
- AI in medicine - **20%**
- Trust in AI diagnostics - **16%**

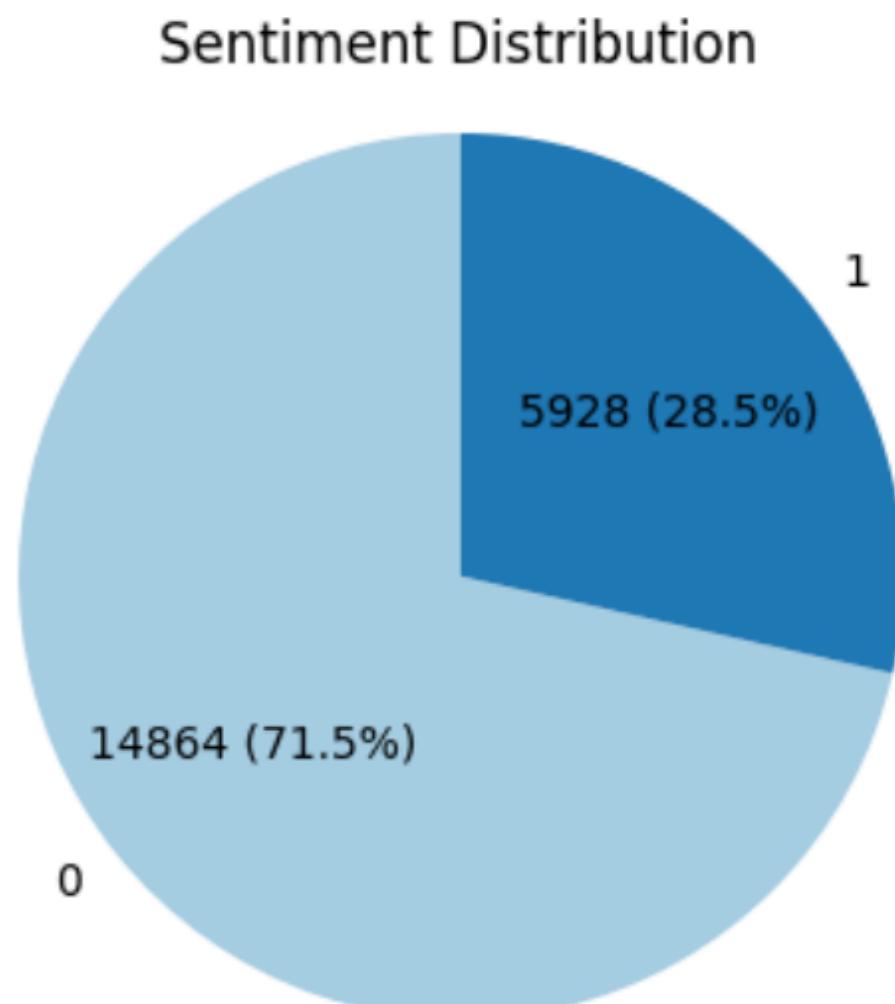
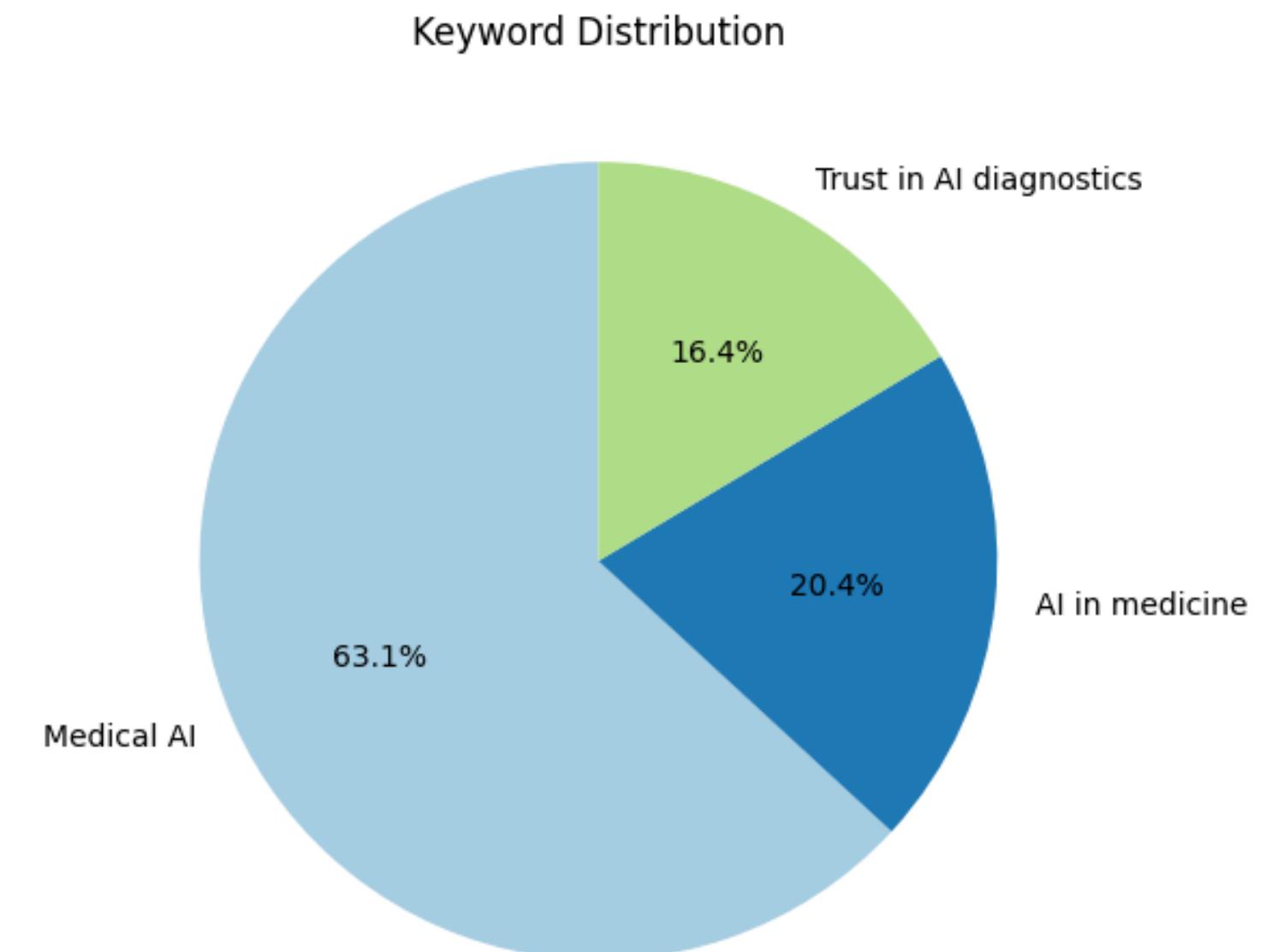


Figure 4



Data analysis

- No high correlation for numerical values

	score	length	mysentiment
score	1.000000	0.010927	-0.031357
length	0.010927	1.000000	-0.019034
mysentiment	-0.031357	-0.019034	1.000000

- Average lengths are similar for

both pos and neg

For length:

	Min	Mean	Median	Max
Positive	2	239	115	8112
Negative	1	254	131	7980

- Negative comments have

higher score than positives.

For scores:

	Min	Mean	Median	Max
Positive	-564	39	3	15461
Negative	-105	76	3	21662

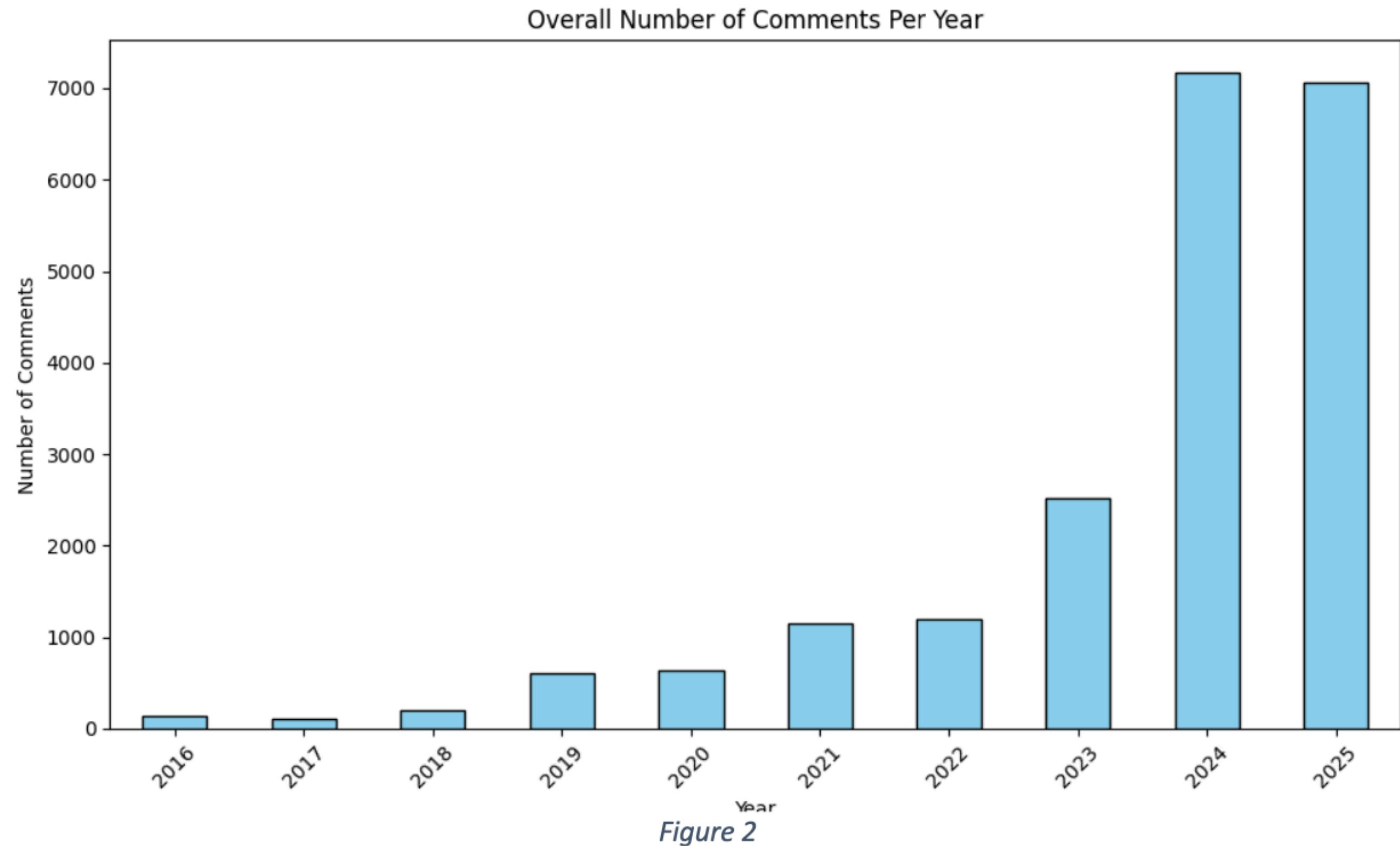
Word Cloud

- In general:
 - AI, just, https, people, like, know, people, ...
 - For positives:
 - Thank, good, love, AI, amazing, will, ...
 - For negatives:
 - AI, reddit, https, think, comments, people, ...

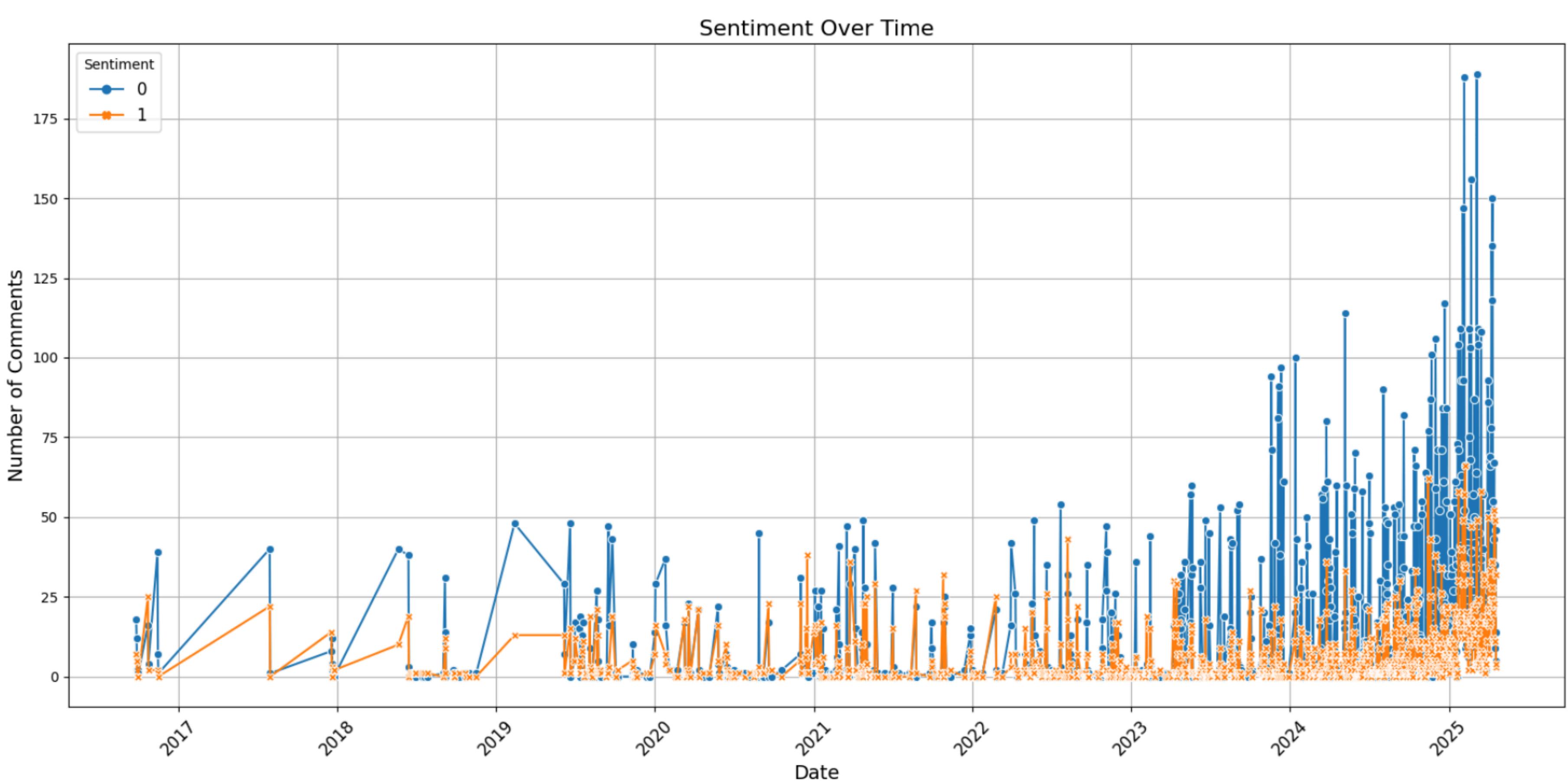


Comments every year

- COVID19 in 2020-22
- Introduction of GenAI
in Nov 2022



Comments every year



Comments every year

- Negative comments consistently around **60-70%**
- Positive comments peaked in 2020 → **46%**.
- 2020-22 - Positive is more than **30%**
- 2022-23 - Significant rise in negative comments
- 2023 - Positive sentiment dropped to **24%** (the lowest)

	Year	Negative	Positive	Total	Negative (%)	Positive (%)
0	2016	102	42	144	70.83	29.17
1	2017	67	41	108	62.04	37.96
2	2018	135	62	197	68.53	31.47
3	2019	413	193	606	68.15	31.85
4	2020	342	295	637	53.69	46.31
5	2021	709	443	1152	61.55	38.45
6	2022	800	400	1200	66.67	33.33
7	2023	1911	606	2517	75.92	24.08
8	2024	5261	1904	7165	73.43	26.57
9	2025	5124	1942	7066	72.52	27.48

Conclusion

- We can see some changes in sentiments after events
- The study is limited to Reddit (small dataset, scope, data quality)
- Overall, sentiment mainly negative towards AI in healthcare
- The results are similar with other studies
- Further study and research is needed