

# **Exploring Public Attitudes towards AI Moratorium: A Topic Modeling and Sentiment Analysis Study**

**(1358 words)**

**Indiana University Bloomington**

**Social Media Mining**

**Instructor: Ali Ghazinejad**

Laxmi Soujanya Taduru  
Spring 2023

## 1. Introduction

Artificial Intelligence has the potential to bring significant benefits to society, but also raises ethical and societal concerns. One proposed solution is an AI moratorium, a temporary ban on the development and deployment of certain types of AI until their implications can be better understood. However, these opinions are divided.

Previous studies, [1] Jobin et al. (2019) conducted a content analysis of online media coverage of AI governance and identified several themes, such as privacy, topics that were frequently discussed. [2] Floridi et al. (2018) proposed a set of ethical principles for AI that consider both individual and societal values. Existing studies have largely overlooked the role of public attitudes and perceptions in shaping AI governance and have not used advanced text analysis methods. This study aims to fill these gaps by analyzing public discussions of AI governance on Reddit using topic modeling and sentiment analysis.

*Keywords: Artificial Intelligence, AI governance, societal concerns, AI moratorium, development, implications, individual values, public attitudes, public perceptions, topic modeling, sentiment analysis.*

## 2. Research question

What are the main topics and how are the sentiments expressed in Reddit discussions related to AI moratorium?

## 3. Method

### 3.1 Data

Reddit is an excellent platform to study public attitudes and perceptions of AI moratorium because it is one of the largest online communities where people can discuss and share their opinions on AI-related issues. On Reddit, users are free to express their opinions and ideas, which makes it an ideal platform for gathering data on public attitudes and perceptions towards AI moratorium.

To answer the research question, I have collected the data from the comments under posts on Reddit related to AI moratorium, using keyword AI Moratorium in the search bar. This was selected as the single search word to ensure that the comments are completely in relevance with AI moratorium.

Also I have manually read the title and browsed through the comments in order to ensure that the data being collected is appropriate. The time frame for the data collected is past one year.

The data was collected using the PRAW module and the Reddit API. The data collection process involves retrieving comments from various posts, and respective sub comments. The collected data includes the post ID, title, date, and text. Based on the engagement activity on each post, a total of 1075 comments were retrieved from 14 different posts.

### 3.2 Analysis

To perform analysis on this data, it was essential to preprocess the data to ensure that the results obtained are accurate and meaningful.

This involved stop word removal which was carried out to eliminate commonly used words that do not contribute to the meaning of the text. Regular expressions (regex) were also applied to remove special characters and symbols that do not add value to the analysis.

After the data cleaning process, the processed data was subjected to sentiment analysis using the EMPATH model, a Python-based tool for sentiment analysis, to identify the positive and negative sentiments expressed in the text data. The sentiment analysis process allows to classify each comment as either positive, negative, or neutral. This classification provided insights about the overall sentiment of the Reddit community towards the topic of AI Moratorium.

After analyzing the sentiment of users, I aimed to better understand the data by performing topic modeling. By conducting topic modeling on the overall sentiment of the user comments and separately on the positive and negative sentiments I was able to gain insights into the main topics for each category. This approach allowed me to identify the most relevant and frequent topics with positive and negative sentiment, providing a more nuanced understanding of the public's attitudes and perceptions towards AI moratorium. Overall, the combination of sentiment analysis and topic modeling enabled me to extract valuable insights from the data and gain a deeper understanding of the public discourse on AI moratorium.

## 4. Results

## 4.1 Sentiment Analysis

After analyzing 1075 comments on the AI moratorium from Reddit using sentiment analysis, it was found that 621 comments were neutral, 361 comments were negative, and 93 comments were positive. This indicates that the overall sentiment on AI moratorium is leaning towards neutrality, with a relatively higher number of negative comments compared to positive ones. From the analysis it could be after manually verifying some subset comments it could be inferred that users on Reddit are in favor of AI development and deployment, while positive sentiment indicates that users are against AI.

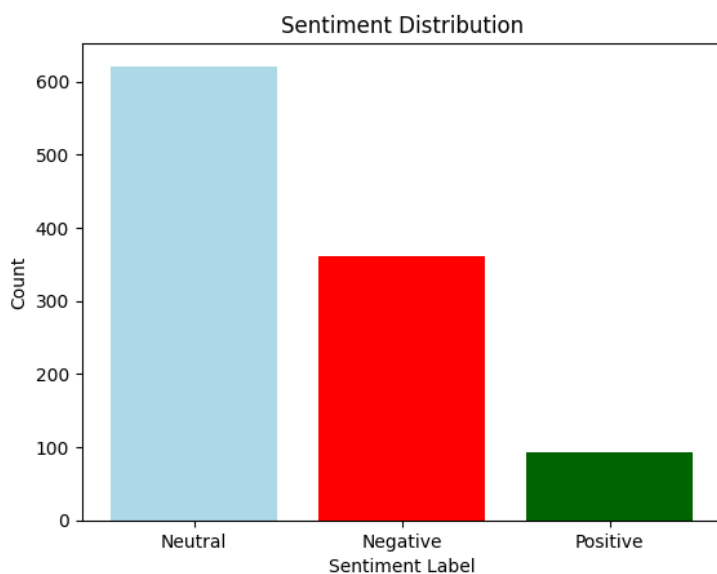


Fig 4.1.1 Sentiment Distribution

## 4.2 Topic Modeling - Overall sentiment analysis

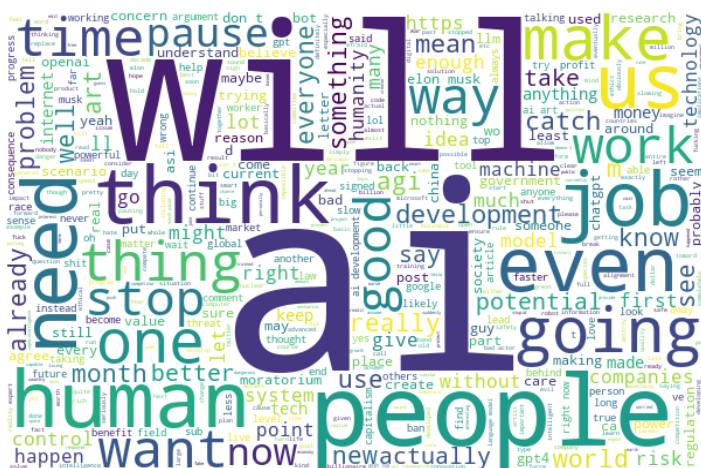


Fig 4.2.1 Overall sentiment Word cloud

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
0	ai	ai	ai	ai	ai
1	would	people	people	us	people
2	like	like	think	people	going
3	people	even	like	like	think
4	want	human	could	think	even
5	development	one	open	could	would
6	art	would	going	would	us
7	need	development	catch	agi	like
8	think	jobs	us	humans	need
9	going	get	work	stop	something

Fig 4.2.2 Top 10 words from each topic

The term "AI" is common across all topics, indicating that it is a prominent topic of discussion. Other common words such as "people", "like", "think", "even", "development", "us", and "work" suggest that the discussions also focus on the potential benefits and drawbacks of AI, job displacement, the need for regulation, and the development of artificial general intelligence (AGI).

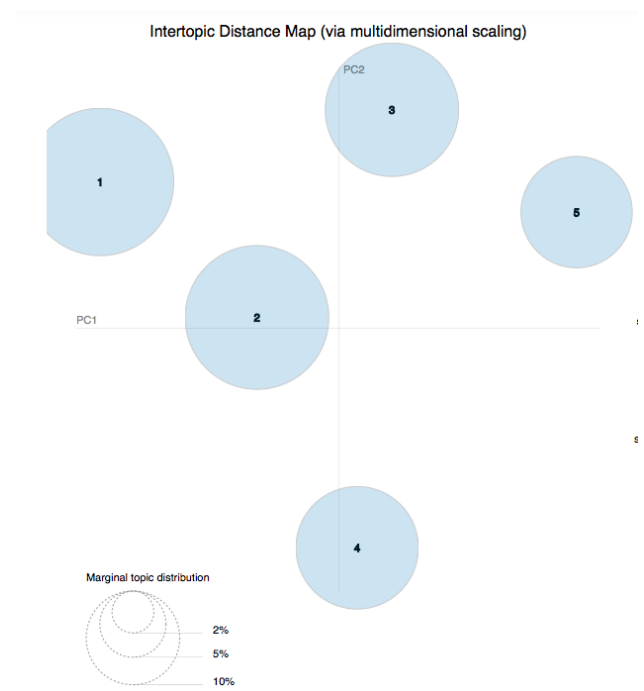


Fig 4.2.3 Intertopic Distance Map

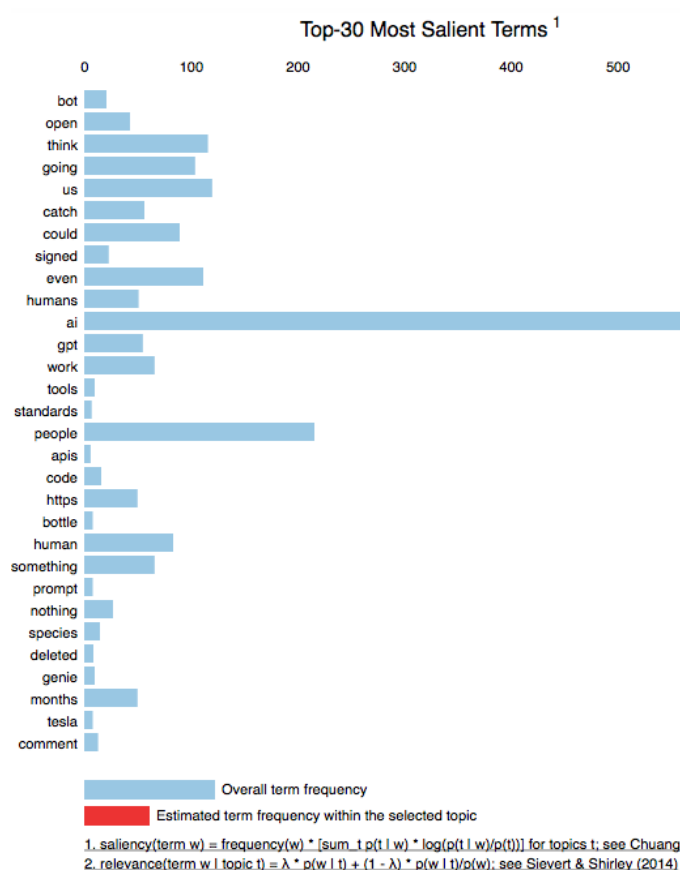


Fig 4.2.4 Top 30 Most Salient Terms

### 4.3 Topic Modeling - Positive sentiment analysis



Fig 4.3.1 Positive sentiment Word cloud

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
0	ai	ai	ai	ai	ai
1	asi	would	better	even	agi
2	could	better	could	need	human
3	better	jobs	control	keep	model
4	would	could	think	potential	job
5	agi	people	would	ambition	would
6	hope	work	good	better	jobs
7	love	think	hope	people	good
8	us	make	humans	jobs	capitalism
9	something	art	like	work	way

Fig 4.3.2 Top 10 words from each topic

Based on the above topics we can infer that people who have a positive sentiment towards AI moratorium see the potential benefits of AI and believe that it can be better utilized with proper governance and control. The words like control are also present, indicating that they recognize the importance of regulating AI to prevent negative consequences. The appearance of the word "job" in the positive sentiment topic modeling suggests that some people may view the development of AI as a potential threat to employment and job opportunities.

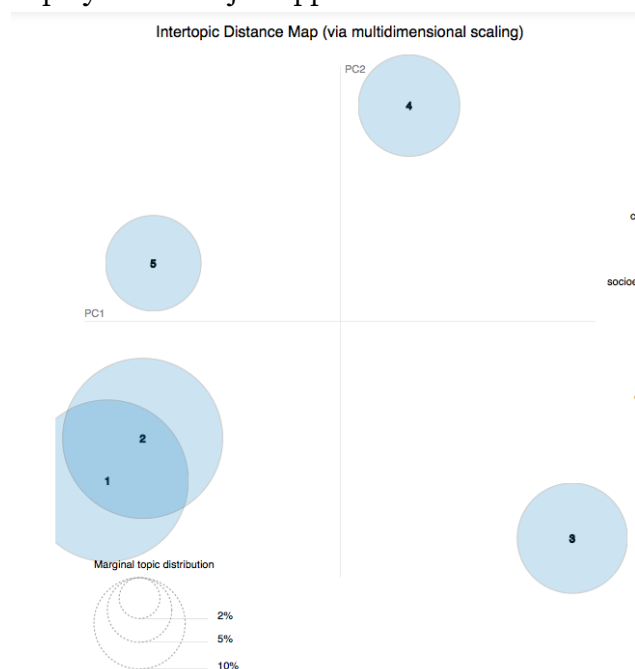


Fig 4.3.3 Intertopic Distance Map

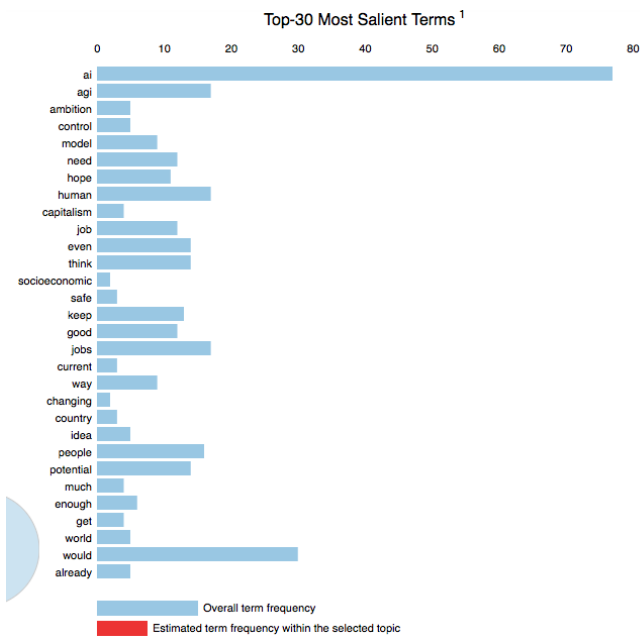


Fig 4.3.4 Top 30 most Salient terms

#### 4.4 Topic Modeling - Negative sentiment analysis



Fig 4.4.1 Negative sentiment Word cloud

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
0	ai	ai	ai	ai	ai
1	people	people	people	people	people
2	us	even	like	stop	even
3	like	us	stop	like	like
4	could	would	think	going	think
5	need	development	would	get	open
6	time	like	us	think	stop
7	jobs	think	even	good	us
8	think	could	get	something	would
9	bad	one	development	one	going

Fig 4.4.2 Top 10 words from each topic

From topic 1,2 and 3 words like ai, people, even, us, would, stop suggesting that those who do not support an AI moratorium are more focused on the potential benefits of AI and its impact on people, rather than the potential risks. They believe that even with the potential risks, the benefits of AI outweigh them, and do not see a need to stop or slow down the development of AI.

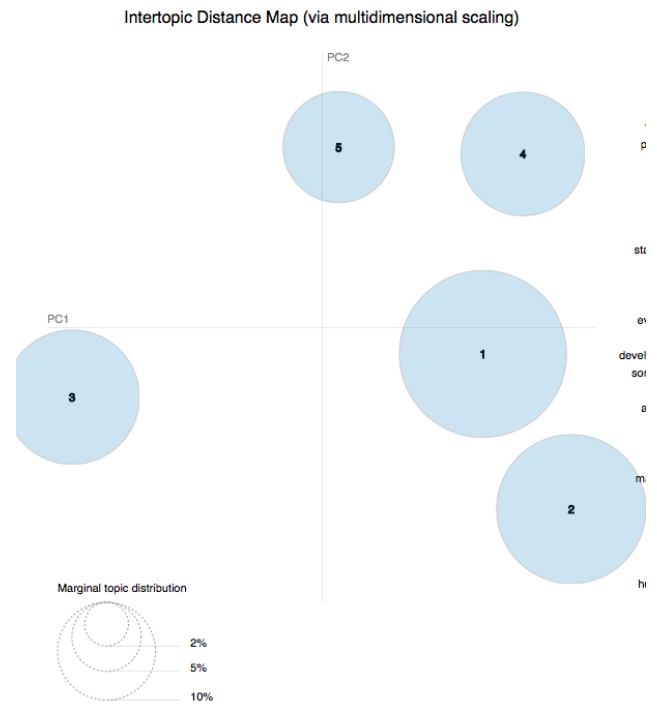


Fig 4.4.3 Intertopic Distance Map

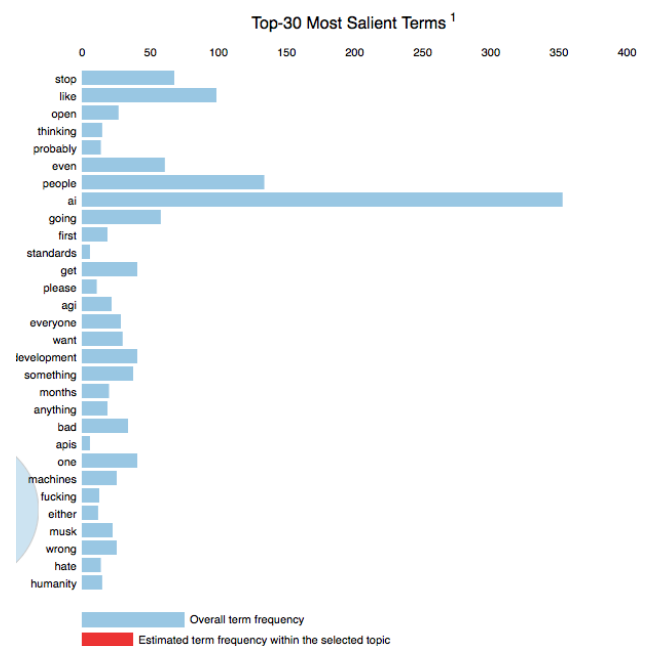


Fig 4.4.4 Top 30 Most Salient Terms

## 5. Conclusion and Limitations

*Some key findings:* Overall, the combination of sentiment analysis and topic modeling provided valuable insights into the public discourse on AI moratorium. The common themes that emerged from the topics included the potential benefits and drawbacks of AI, job displacement, the need for regulation, and the development of AGI. The analysis revealed that the overall sentiment towards AI moratorium was neutral, with a higher number of negative comments compared to positive ones. Users on Reddit generally seemed to be in favor of AI development and deployment, while those with negative sentiment towards AI moratorium were more focused on the potential benefits of AI and its impact on people, rather than the potential risks.

*Limitations:* One limitation of this analysis is that it is based solely on comments from Reddit users and may not be representative of the general public's views. Additionally, the analysis relies on machine learning algorithms, which may not always accurately capture the nuances of human language and meaning. Therefore, we cannot make any conclusive inferences solely based on the sentiment analysis and topic modeling results. This suggests that there is a need for further research on public attitudes and perceptions towards AI moratorium and for a better understanding of the complexities of the topic.

## 6. References

- [1] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*.
- [2] Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P (2018). *AI4People—An ethical framework for a good AI society*.
- [3] Babao Zhang, Allan Dafoe (2019) US public opinion on the governance of artificial intelligence.