Mini Project

EQUISUMM

A BIAS-AWARE FRAMEWORK FOR INCLUSIVE SUMMARIZATION

Chaitanya Wanjari (2022IMT - 034)
Jessica Kamal (2022IMT - 054)
Riddhi Jain (2022IMT - 097)
Samruddhi Kurhe (2022IMT - 100)

Under the supervision of : Dr.Roshni Chakraborty

PRESENTATION OUTLINE

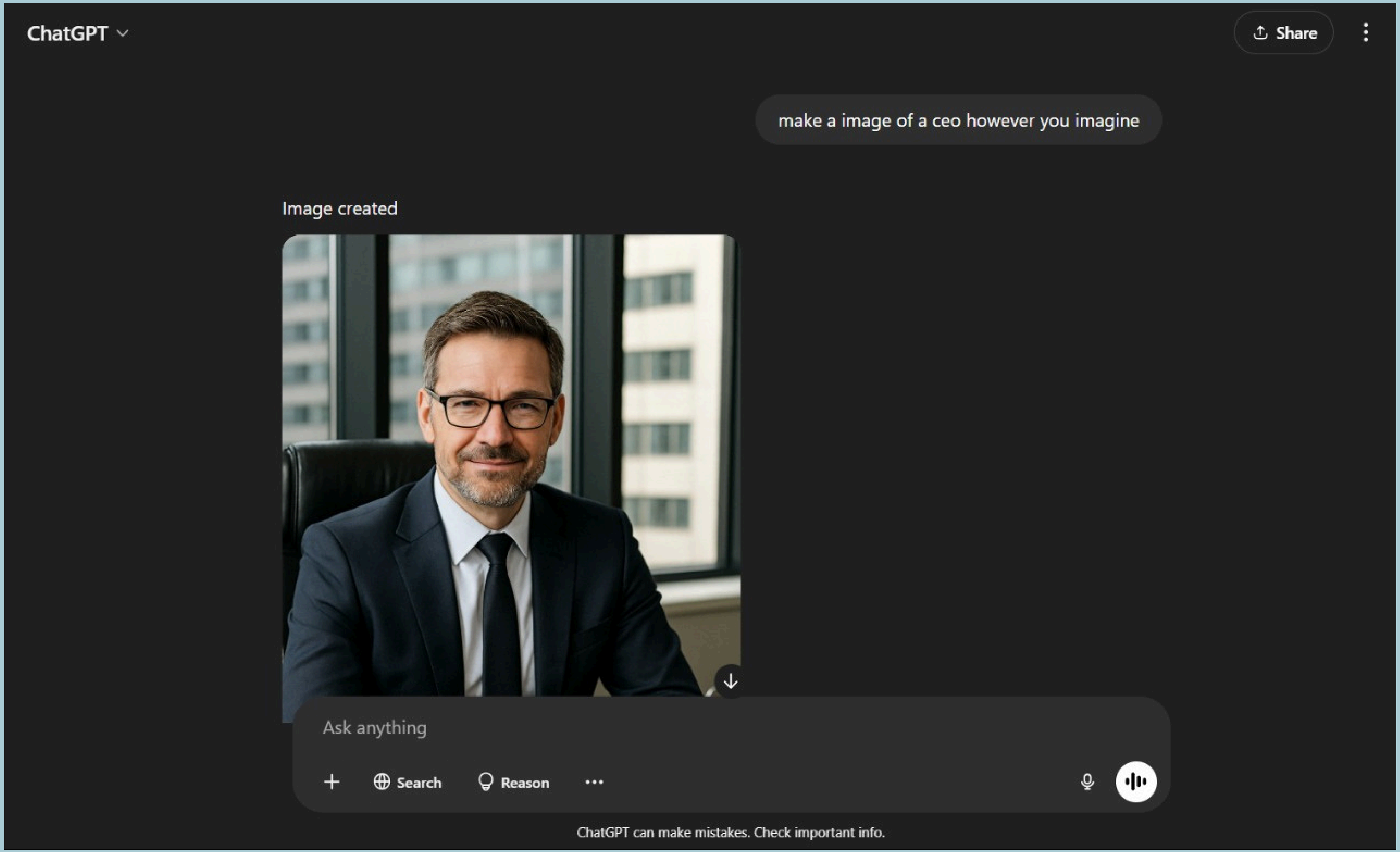
01 Introduction

02 Proposed Method

03 Topic Wise Analysis

04 Metrics

05 Conclusion



AI AND STEREOTYPES: A QUICK GLIMPSE

What is Gender Bias?

Gender bias is when individuals are treated unfairly based on their gender—often favoring one over another.

It can be conscious or unconscious, showing up in language, behavior, expectations, or decisions.

In a meeting, a man's idea might be praised while a woman's similar idea gets overlooked—that's bias in action.

Motivation

Examples

"She’s too emotional to handle a leadership position."
"We need someone assertive for this job, maybe a man would be better."
"She's probably not interested in tech, it's a guy thing."
‘A bold and assertive man took charge of the meeting.’

WHY IS REPRESENTATION NECESSARY?

Equality: Fair representation leads to equal opportunities.

Diversity: Enriches perspectives and drives innovation.

Social Impact: Prevents the reinforcement of harmful stereotypes.

Automation Fairness: Automated systems should generate summaries that are unbiased and representative of all voices.

Summarization: Information Overload

Challenge:

- There is a plethora of information available for every topic.

Solution:

- Effective summarization condenses key points, making large volumes of information accessible and manageable.

Summarization Challenges

Existing summarization tools does not focus on gender based fairness even if the event is gender based

Tweet 1:
“When I told HR he made me uncomfortable, they said he was just being friendly. I started doubting myself”

Tweet 2:
"As a man, I tried speaking up about harassment I witnessed, but was told to stay out of 'women’s matters"

Methodology

01 Analyzing existing summarization approaches for gender fairness

02 Proposal of a gender fair summarization approach

03 Visualization of Topic-wise Summarization

04 Analyzed existing metrics to measure bias.

05 Proposed a gender inclusion based metrics

Challenges with Existing Metrics

Limitations of Summarization based metrics

- Based on ROUGE scores
- Does not consider demographic attributes

ROUGE-N Formula:

$$\text{ROUGE-N} = \frac{\sum_{\text{ref} \in \text{References}} \sum_{\text{gram}_n \in \text{ref}} \text{Count}_{\text{match}}(\text{gram}_n)}{\sum_{\text{ref} \in \text{References}} \sum_{\text{gram}_n \in \text{ref}} \text{Count}(\text{gram}_n)}$$

Where:

- `gram_n` refers to n-grams (e.g., unigrams for ROUGE-1, bigrams for ROUGE-2)
- `Count_match` is the number of n-grams that match between generated and reference
- The result is a recall score indicating how much of the reference is covered

Summarization

Existing research works

- LexRank
- LSA (Latent Semantic Analysis)
- Community Detection + Lexrank

Proposed Methodology

Gender + Lexrank

A Deep Dive into Semantic Similarity, Summarization & Bias Detection

Datasets (tweets):
MeToo
Legalisation of Abortion

Gender + Lexrank Pipeline

Classification on basis of gender

Calculation of pair wise similarity scores

Obtain summaries for different gender tweets

Obtain a combined summary from the prior obtained summaries

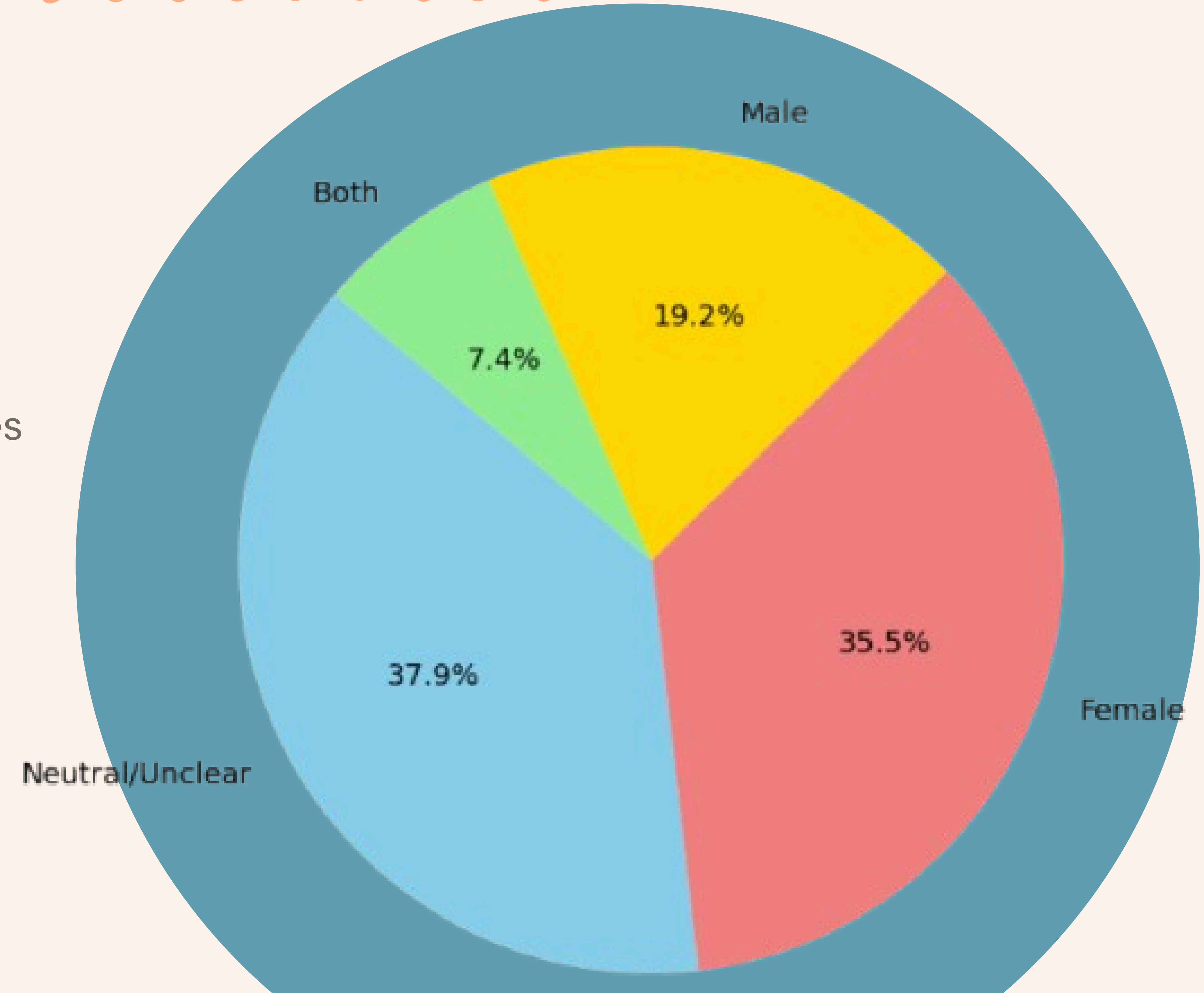
Calculate Inclusion Bias score in the final summary using proposed metric

Gender Classification

- **MeToo Dataset**

  Who's Being Talked About?

- Used spaCy's Named Entity Recognition to detect gender references
- Tweets categorized into:
  1) Male
  2) Female
  3) Neutral
  4) Both

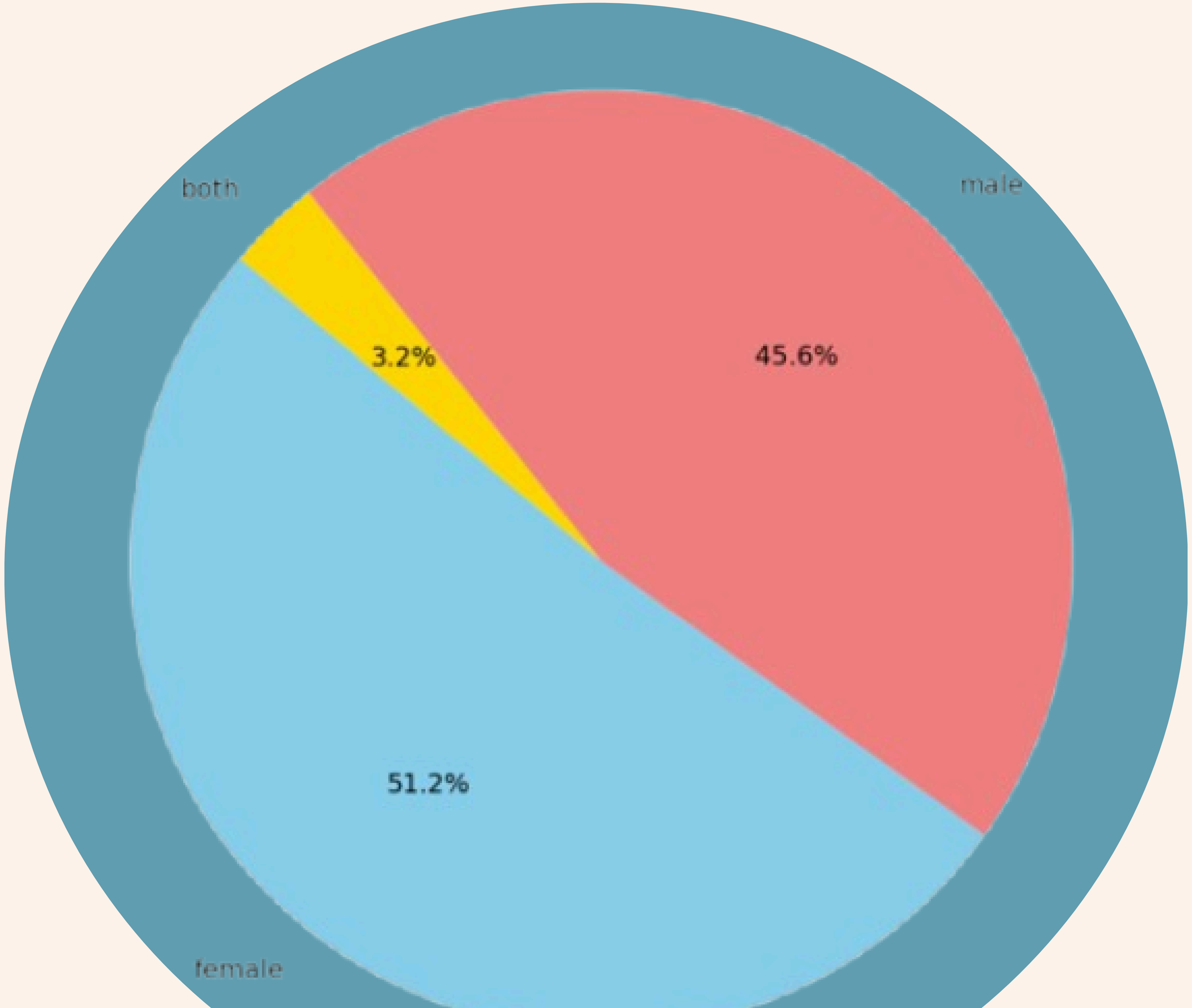


Gender Classification

- **Legalisation of Abortion Dataset**

  Who's Being Talked About?

- Classification of tweets on the basis of Clustering + Similarity
- Sentence embedding for all tweets
- Compare each unknown tweets embedding to centroid of male and female tweets
- Classify an unknown tweet as male or female



Semantic Similarity

What Are They Talking About?

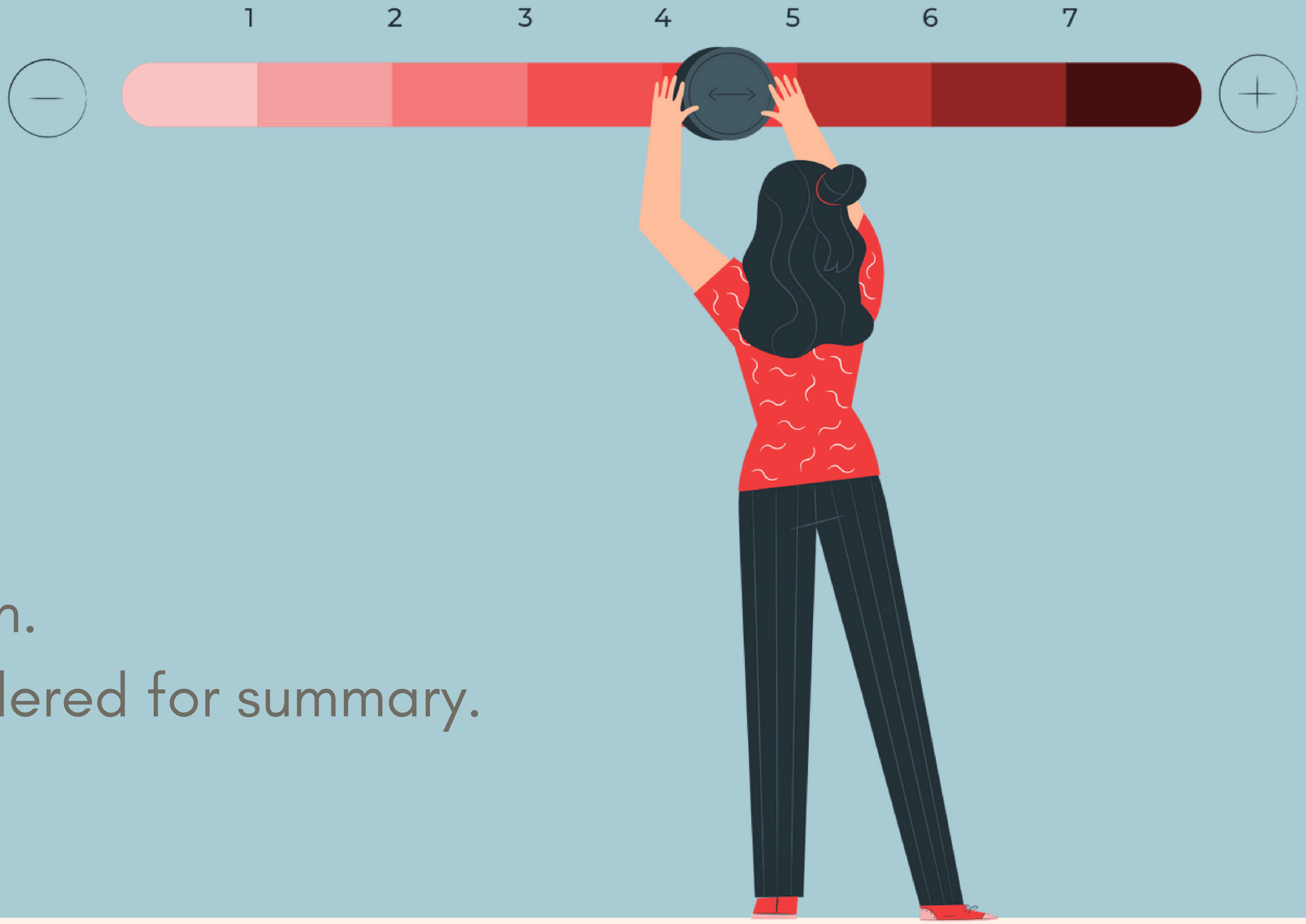
- **MeToo dataset**

  Calculated pairwise semantic similarity using Bert Embeddings

- **Legalisation of Abortion dataset**

  Classification of tweets on the basis of Clustering and Similarity

Threshold



How Similar Is Similar Enough?

- Manually analyzed similarity scores.
- Set a threshold for semantic connection.
- Only tweets above the threshold considered for summary.

Weighted LexRank Summarization

From Tweets to Summary

- Applied Weighted LexRank to each gender category:
- Male
- Female
- Both
- Created even-sentence summaries for each.
- Merged & summarized again for the final output.

Summarization

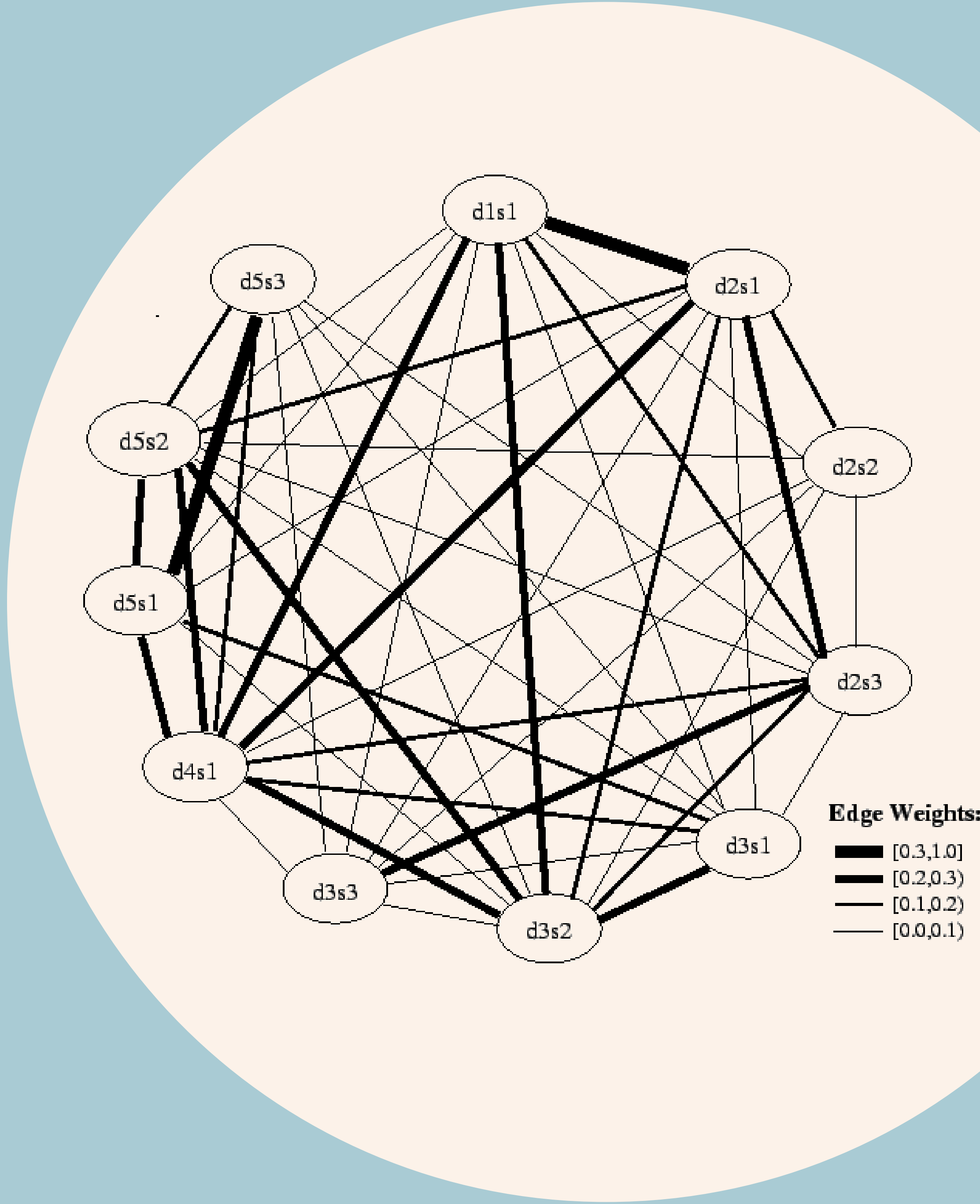
Male tweets Summary

Female tweets Summary

Both tweets Summary

Weighted Lexrank

Combined Summary



Topic-wise Summarization: BERTopic Analysis

Concept:

- Organize summaries based on specific topics or clusters.
- Advantage: Enables a clearer view of thematic emphasis, including potential bias in different sections of the text.

What is BERTopic?

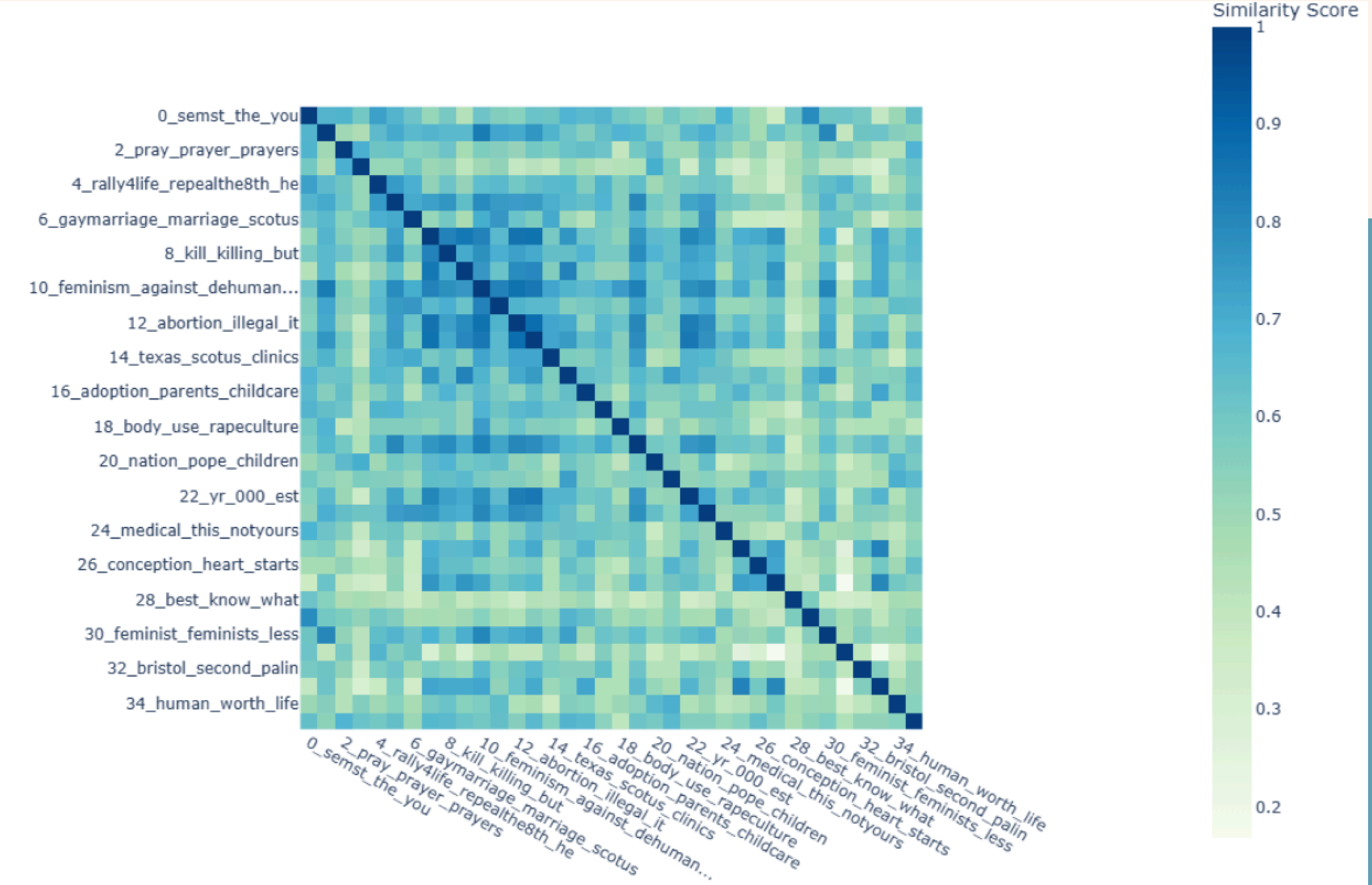
It is a topic modeling technique that leverages transformers and c-TF-IDF to create dense clusters allowing for easily interpretable topics whilst keeping important words in the topic descriptions.

Why Preprocessing is Critical for Meaningful BERTopic Results?

Challenges:

- Grammatical keywords affecting the topic division
- Most frequent keywords combining the topics
- Words related to similar topics are treated as different keywords

Result before Pre-processing:



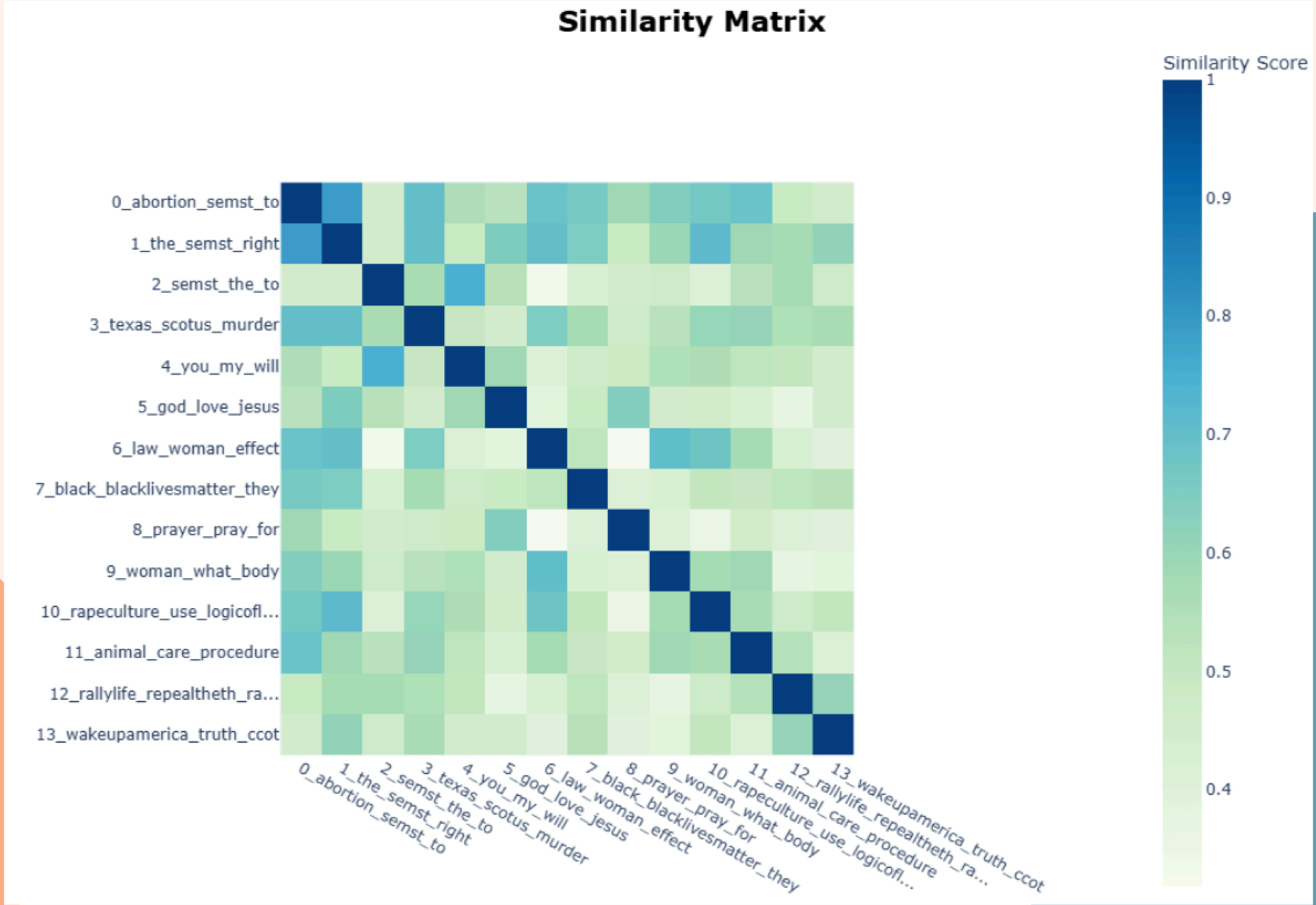
Pre-processing steps:

A. Removal of stop words (Grammatical words like articles, pronouns)

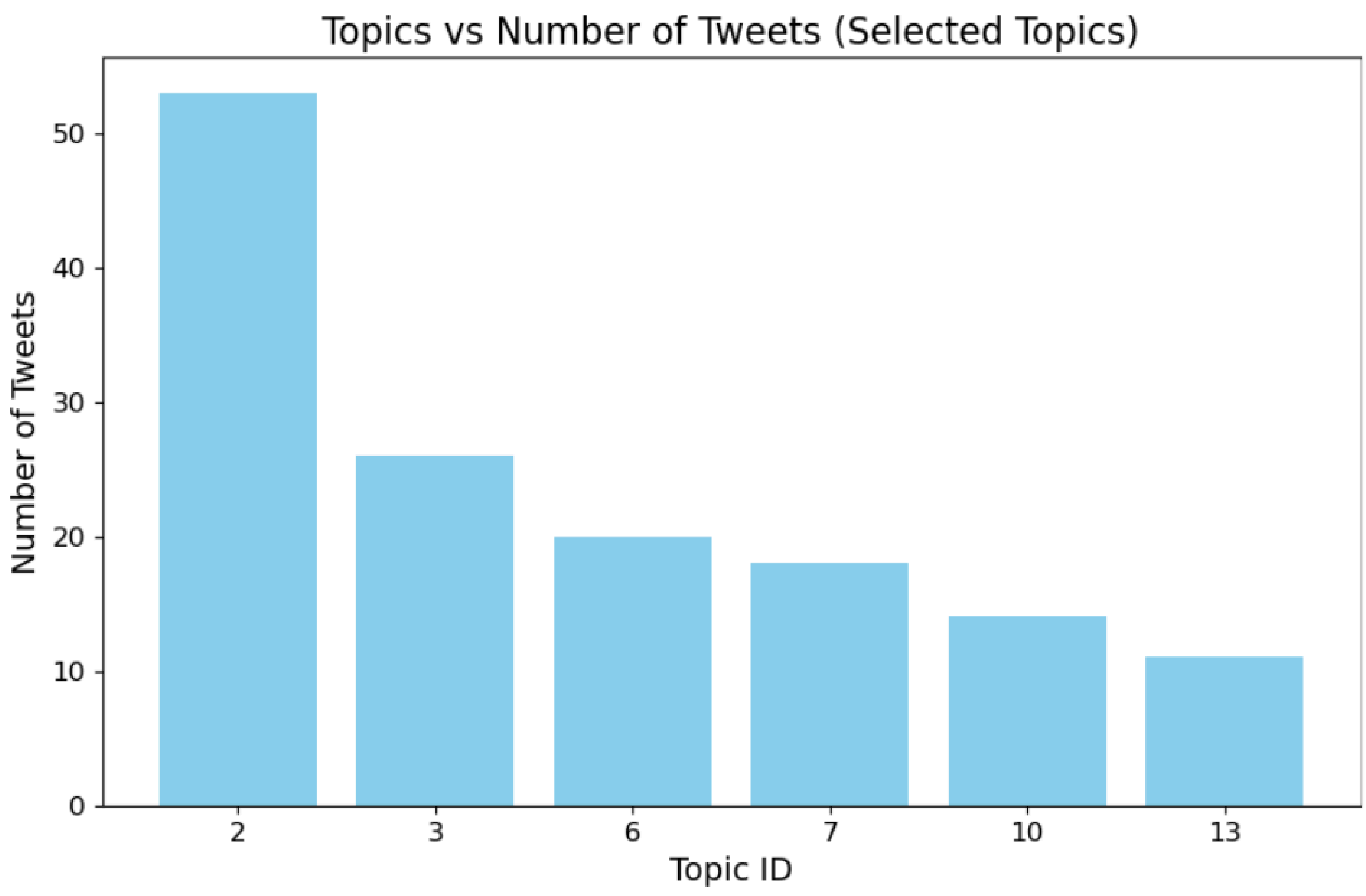
B. Filtering out with the Event Representative Keywords based on the Occurrence Ratio.
Occurrence Ratio = (No. of tweets the word appeared in) / (Total no. of tweets)

C. Unify the keywords by using Lemmatizer and Unification function

Result after Pre-processing:



Results for Legalization of Abortion Dataset:



Topics division of Tweets:

Topic Id - 3 (Texas Soctus Clinic)

- adding to the progress of this week the supreme court is also allowing texas abortion clinics to stay open ! # scotus # semst
- # scotus blocking part of our texas anti-abortion law . maybe we should call them the supreme tyrant . # tcot # tgdn # semst

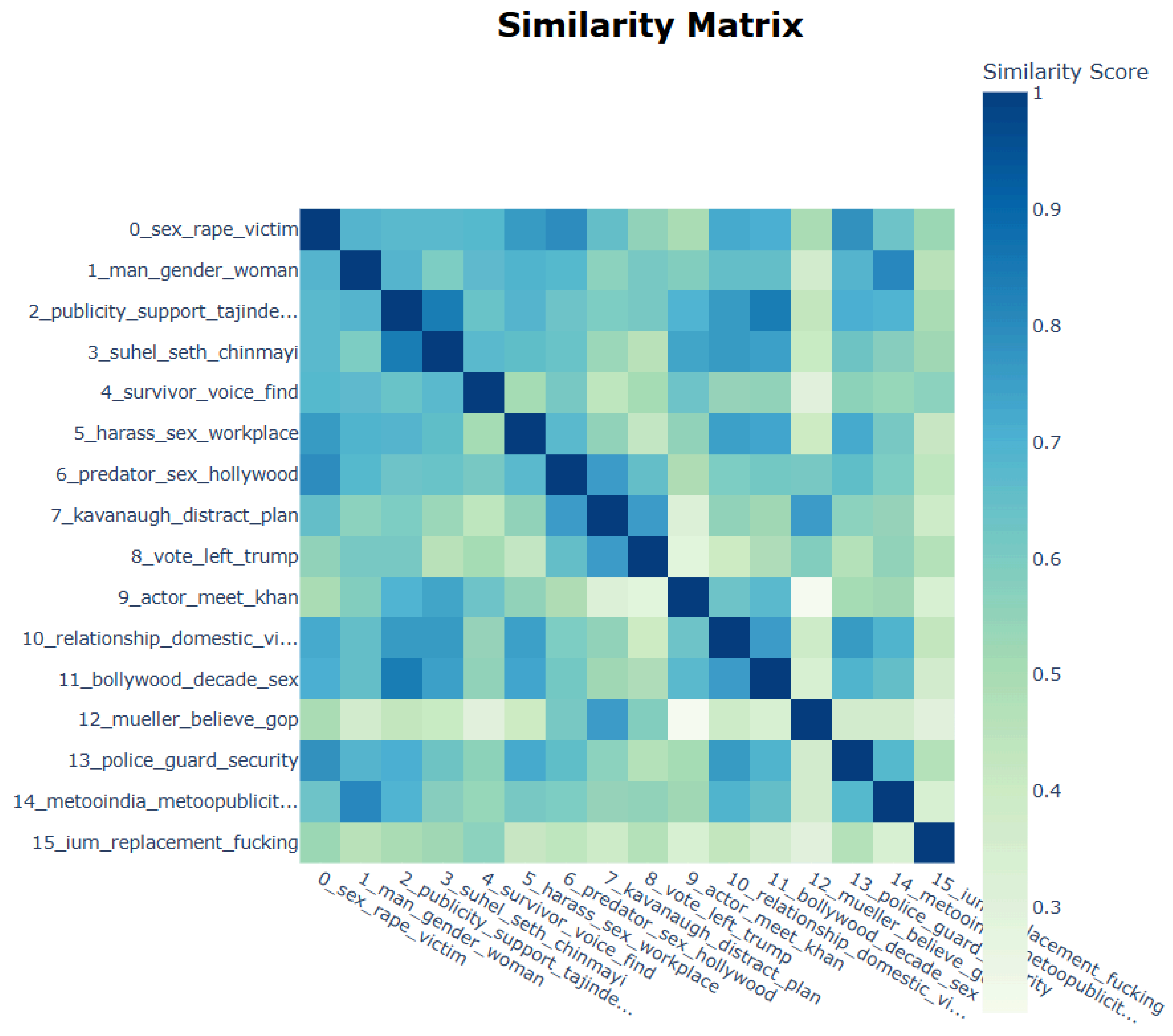
Topic Id - 5 (Religion)

- manipulation , disguise , & evil is always seen by the out-come . the ugly doors of satan appear in all forms when it 's fake love . # semst
- @ ladysandersfarm love is only love if it 's on her terms . keep your faith in jesus , he is your anchor in this storm . hebrews 6:19 # semst

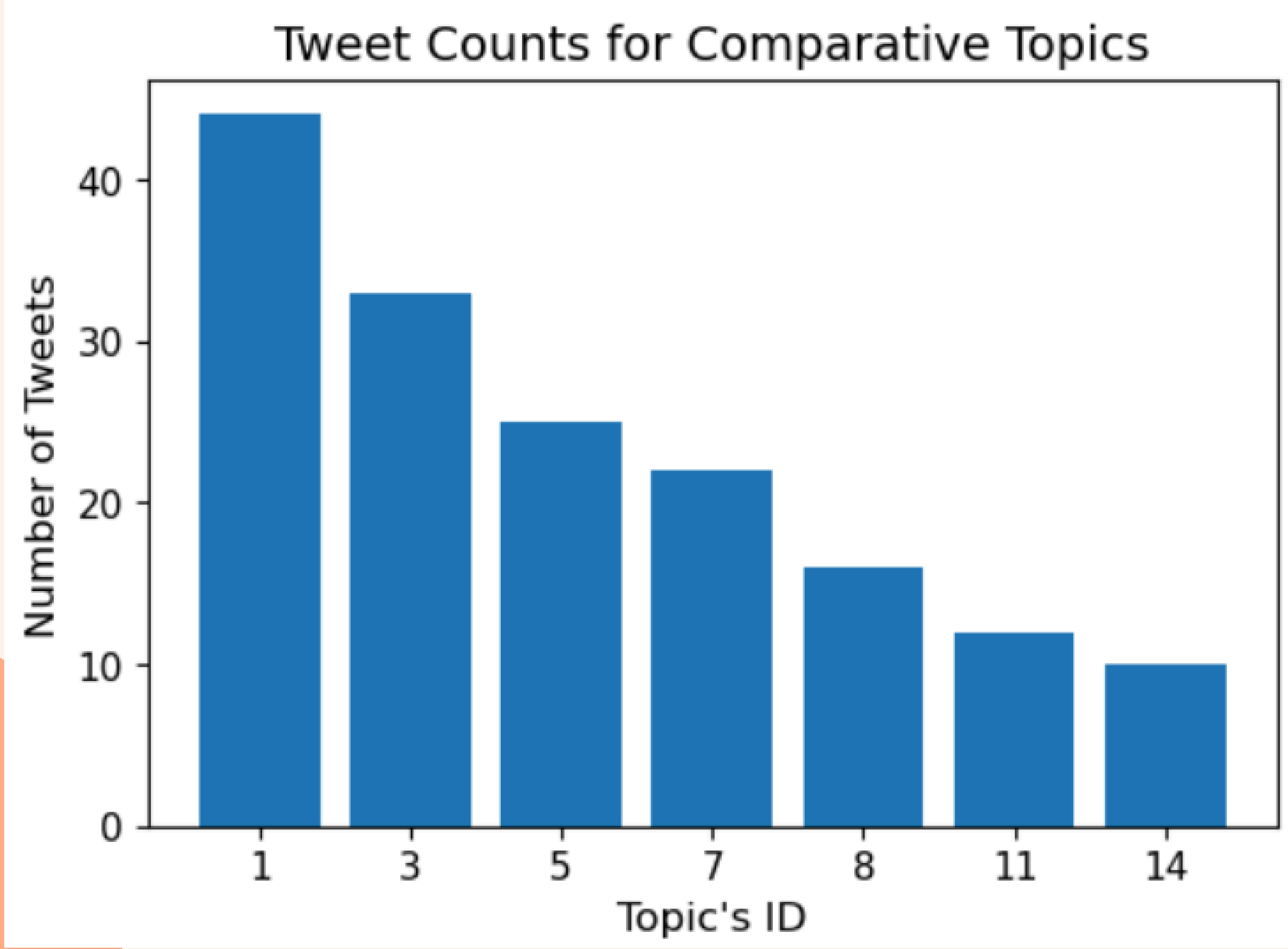
Topic Id - 10(Rape culture)

- @ azzarellijim not so great for the women you want to force to give birth against their wills . # rapeculture # semst
- @ logicoflife7 i don't confuse `` right to life '' with 'right-to-use-someone's-body-without-consent ' do you ? # rapeculture # semst

Similarity Matrix for MeToo Dataset:



Results for MeToo Dataset:



Topics division of tweets:

Topic Id - 5 (Workplace Sex Harassment)

- Home Minister Rajnath Singh who is making fun of #MeToo campaign in this video is ironically heading a committee to look into cases of sexual harassment at workplace.
- Thank you @LeanneWood Together we can tackle sexual harassment. We need a national conversation and a national survey to expose how widespread it is in Welsh society so we can move to eliminate it. A support network is an excellent idea. #MeToo

Topic Id - 14 (MeToo India)

- Respect females, not because they deserve. ,Do it because your single mistake (according to her) will ruin your life totally, after that social media , women's commission, aliens all would be against you. #MeghaSharma #MeTooIndia #MeToo #MeTooControversy
- Dear #feminists of India, see this. #MeToo #MeTooIndia #MeTooControversy #MeToo4Publicity   #Police and Judiciary must watch this because no media will show you this and make it popular. #feminist and biased society must change their biased mentality. #MenToo #HeToo https://t.co/7MYtpopqbS

INCLUSION BIAS

WHY DO WE NEED IT?

Understanding Inclusion Bias

Inclusion Bias refers to the unequal representation of different demographic groups (such as gender) in the output summaries relative to their presence in the source text.

Inclusion bias measures whether gendered identifiers are represented in summaries proportionally to their presence in the source text, rather than assuming a uniform distribution. It highlights if certain voices are under- or over-represented relative to the original content.

$$\text{Count-Based Score} = \frac{Count_{female} - Count_{male}}{Count_{male} + Count_{female}}$$

METRICS

Summary: "She is brave. He is strong. The woman and the man both spoke up."

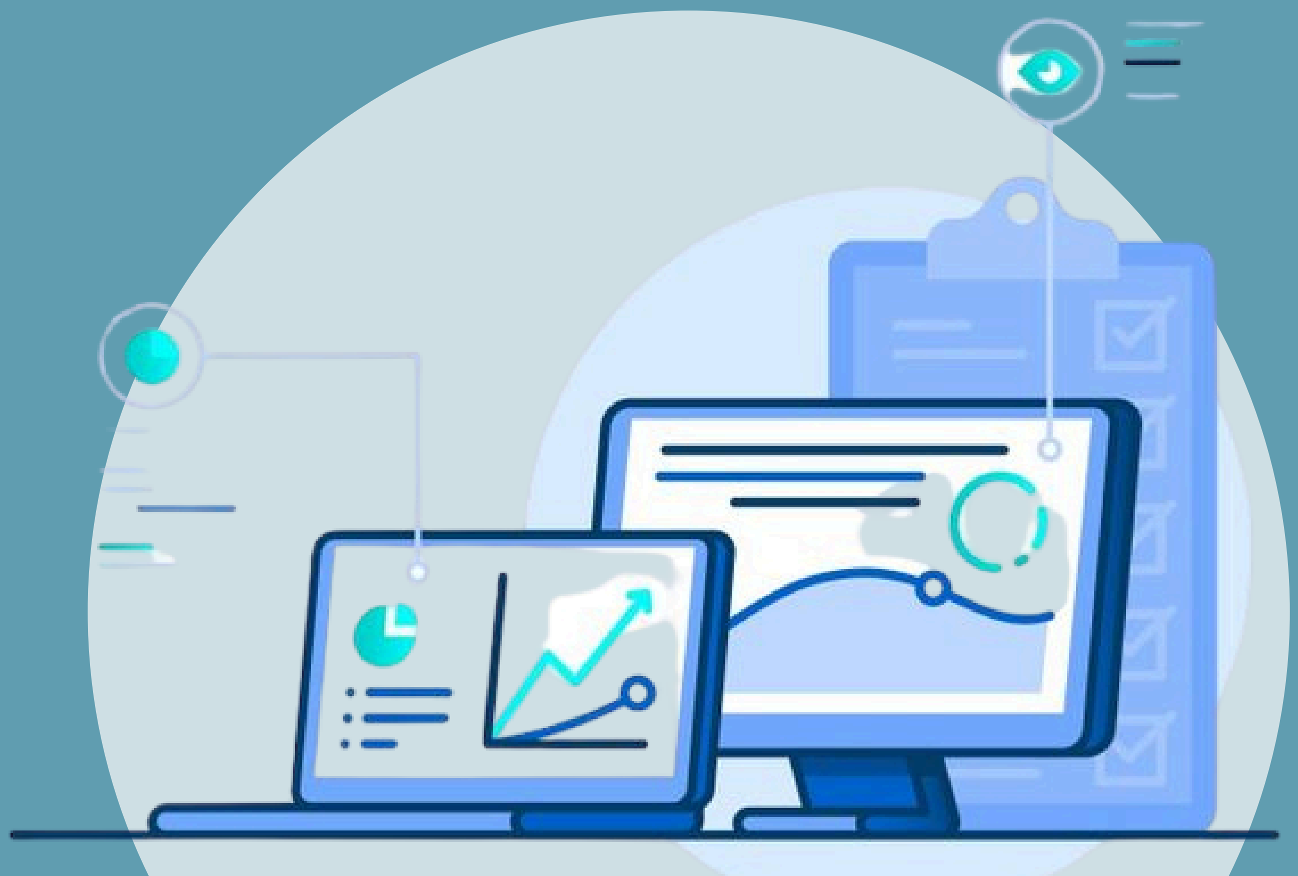
Female: ["she", "woman"] → 2 unique, 2 total

Male: ["he", "man"] → 2 unique, 2 total

Summary: "She is a woman. The woman inspired another woman. He was there."

Female: ["she", "woman", "woman", "woman"] → 2 unique, 4 total

Male: ["he"] → 1 unique, 1 total



PROPOSED METRIC

WEIGHTED INCLUSION BIAS

- Counts all occurrences (including repetitions) of gendered terms.
- Emphasizes amplification due to word repetition in summaries.

$$\text{Frequency-Based Score} = \frac{\sum_{i=1}^{n} freq(f_j)}{\sum_{i=1}^{n} freq(m_i) + \sum_{j=1}^{k} freq(f_j)} - \frac{\sum_{j=1}^{k} freq(m_i)}{\sum_{i=1}^{n} freq(m_i) + \sum_{j=1}^{k} freq(f_j)}$$

Datasets

MeToo

- Source: Collection of tweets related to the #MeToo movement.
- Content Type: Social media posts expressing personal experiences, opinions, news headlines, and public responses.
- 485 unique tweets

"When will women realise they are digging a well for themselves by false #MeToo allegations?"

Legalization of Abortion

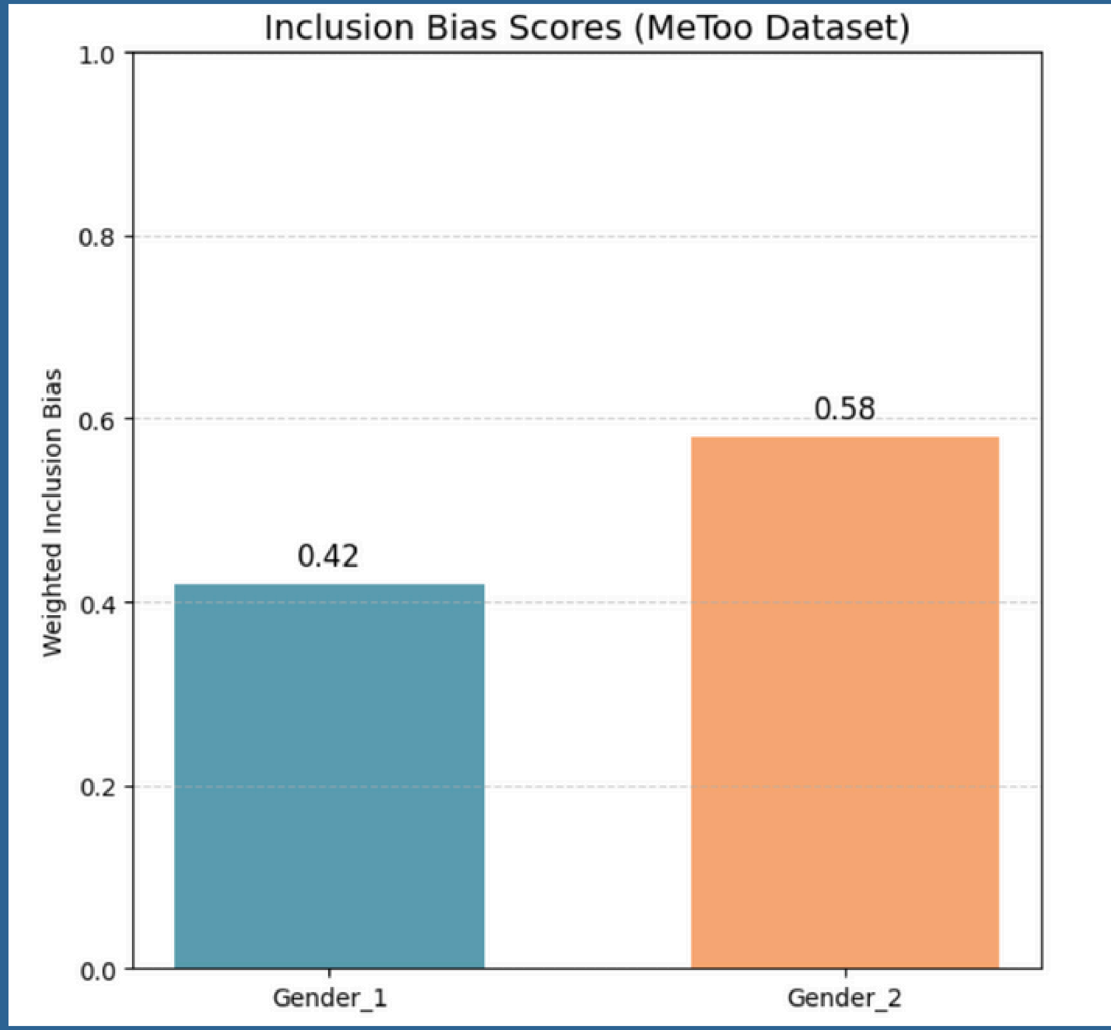
- Source: Tweets discussing the topic of abortion legalization.
- Content Type: Political, ethical, and religious arguments both for and against abortion.
- 934 unique tweets

How many abortion doctors have told a woman 'no, an abortion is not required in your case'?

Datasets

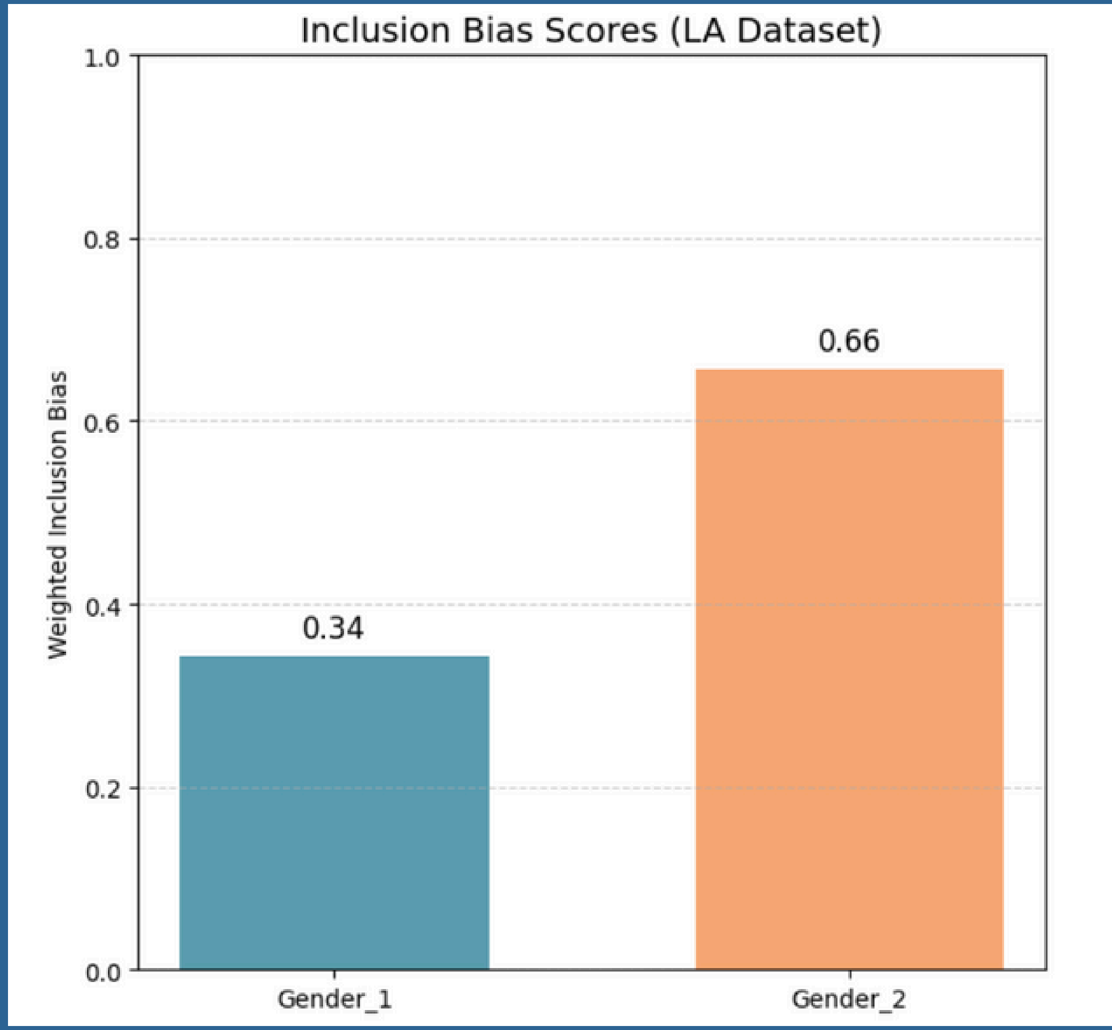
MeToo

- Weighted Gender 1 Score: 0.41
- Weighted Gender 2 Score: 0.58
- Weighted Inclusion Bias Score (gender 1- gender 2): 0.16



Legalization of Abortion

- Weighted Gender 1 Score: 0.34
- Weighted Gender 2 Score: 0.65
- Weighted Inclusion Bias Score (Gender 1 - Gender2): 0.31

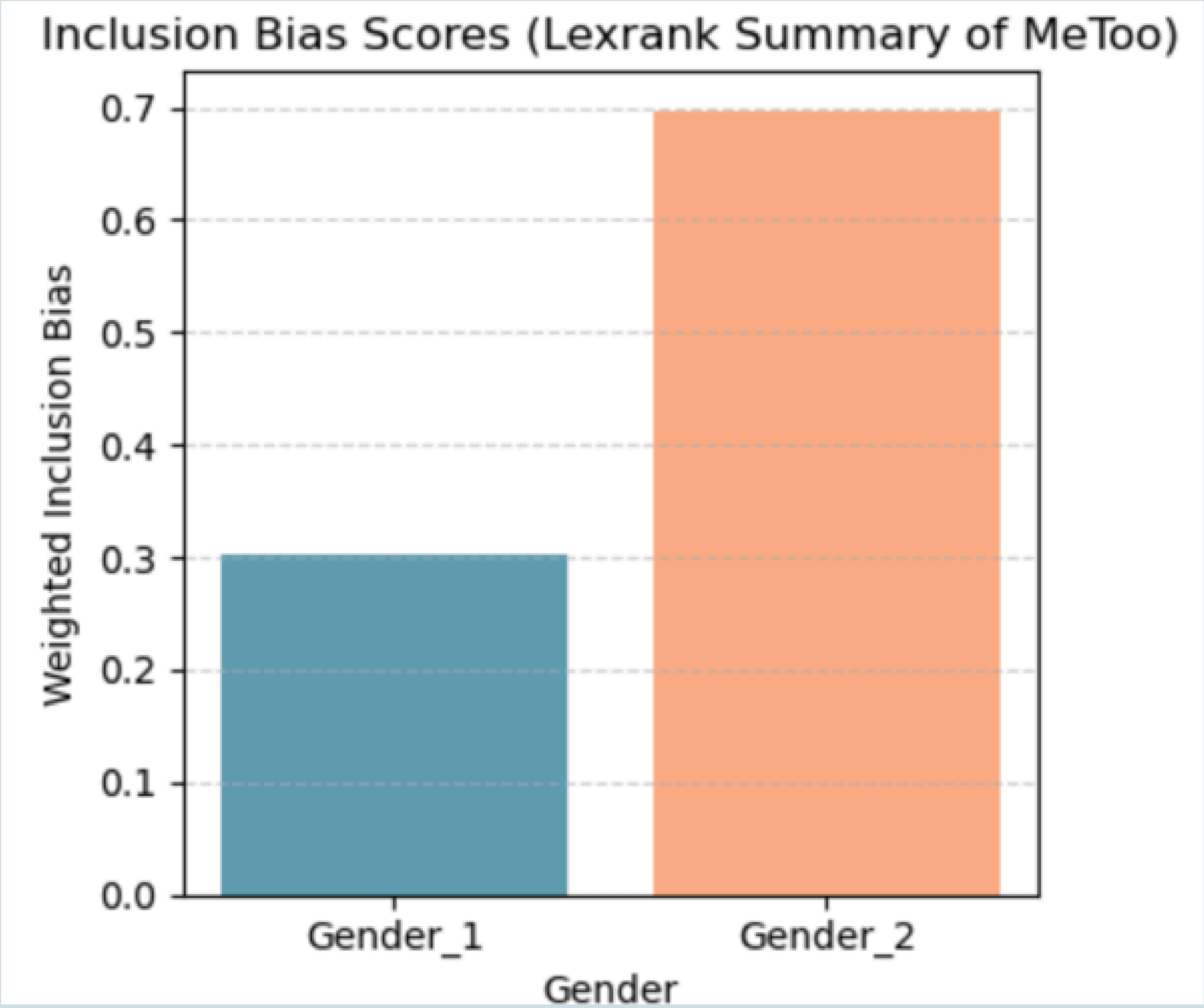


Results(MeToo)

Lexrank

Weighted gender 1 Score: 0.30
Weighted gender 2 Score: 0.69
Weighted Inclusion Bias Score
(gender1 - gender2): 0.39

Summary is biased toward gender 2.

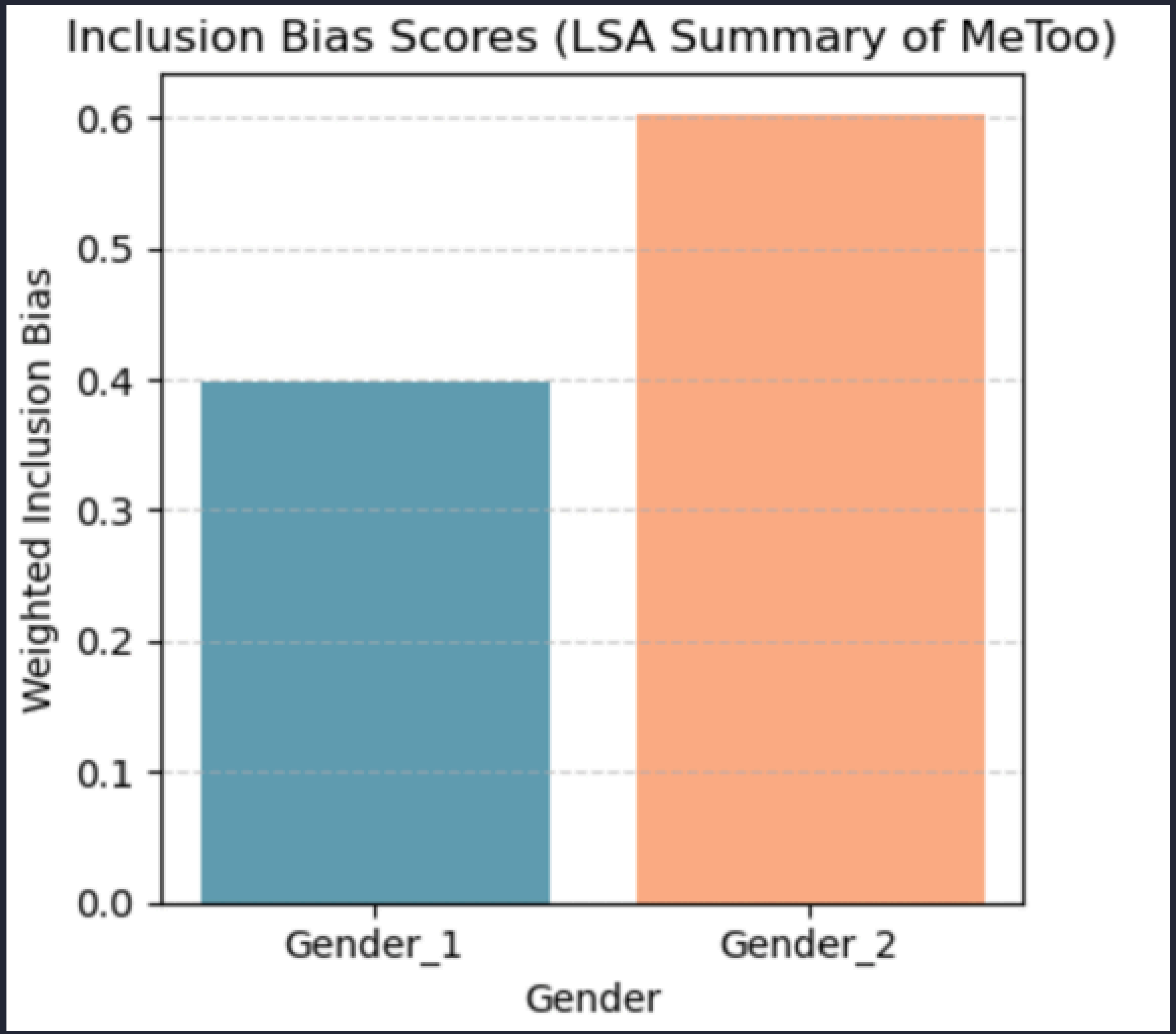


Results(MeToo)

LSA

Weighted gender 1 Score: 0.397260
Weighted gender 2 Score: 0.602740
Weighted Inclusion Bias Score
(gender1 - gender2): 0.205479

Summary is biased toward gender 2.

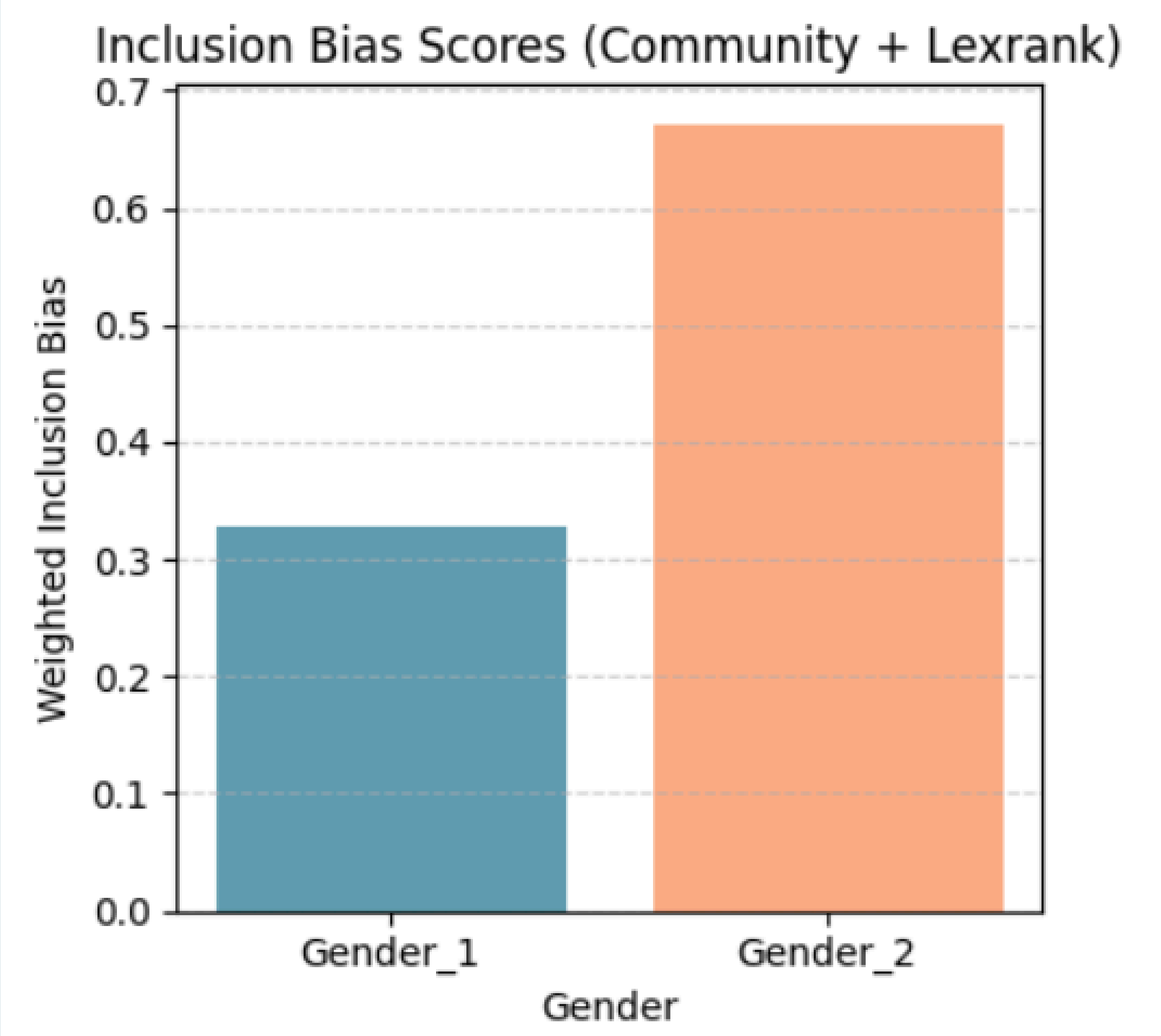


Results(MeToo)

Community+Lexrank

Weighted gender 1 Score: 0.32
Weighted gender 2 Score: 0.67
Weighted Inclusion Bias Score
(gender1 - gender2): 0.34

Summary is biased toward gender 2.



Results(MeToo)

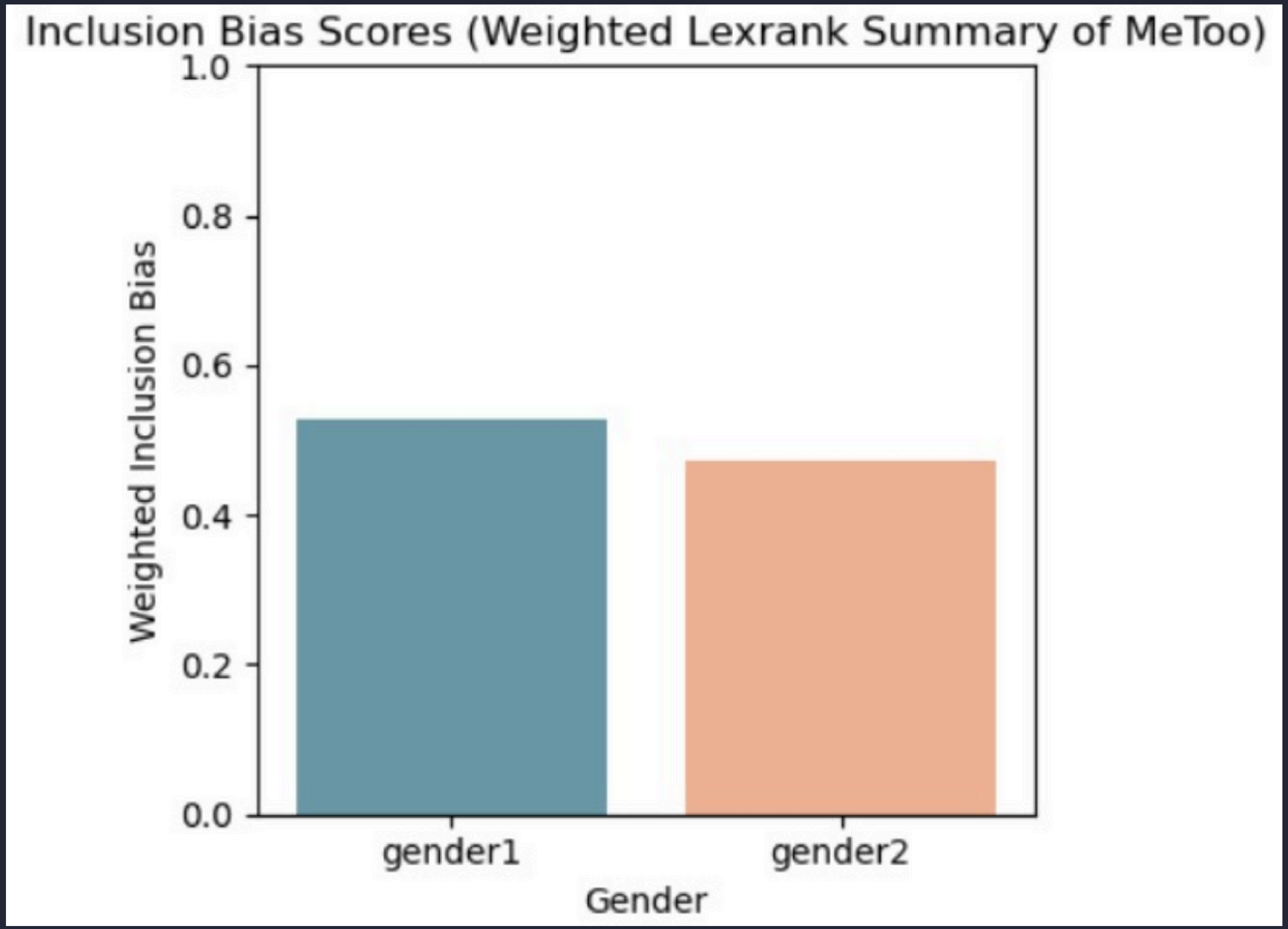
Gender + Lerank

Weighted gender 1 Score: 0.57

Weighted gender 2 Score: 0.44

Weighted Inclusion Bias Score (gender1 - gender2): 0.05

Summary is biased toward gender 2.

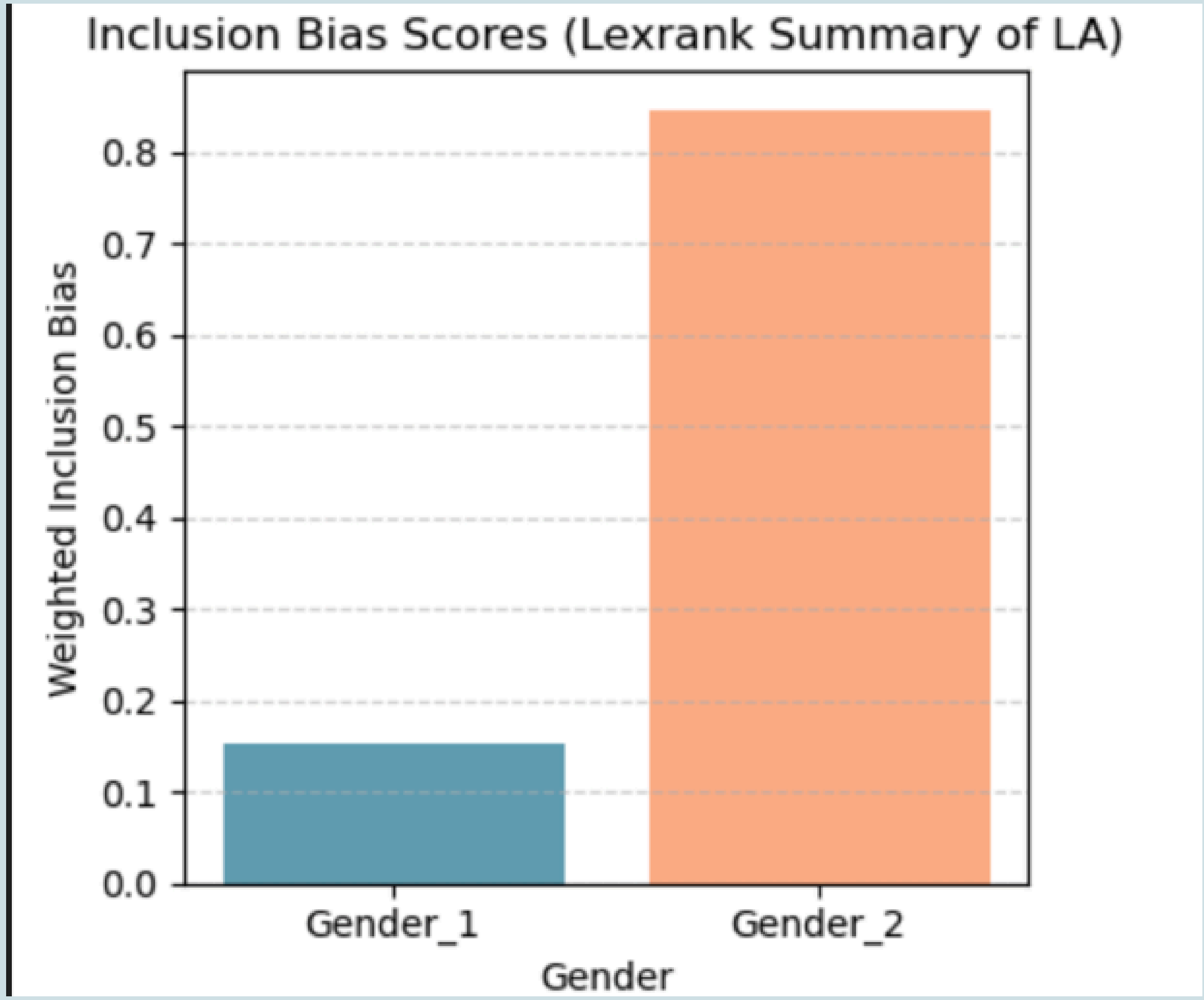


Results(LA)

Lexrank

Weighted gender 1 Score: 0.15
Weighted gender 2 Score: 0.84
Weighted Inclusion Bias Score
(gender1 - gender2): 0.69

Summary is biased toward gender 2.

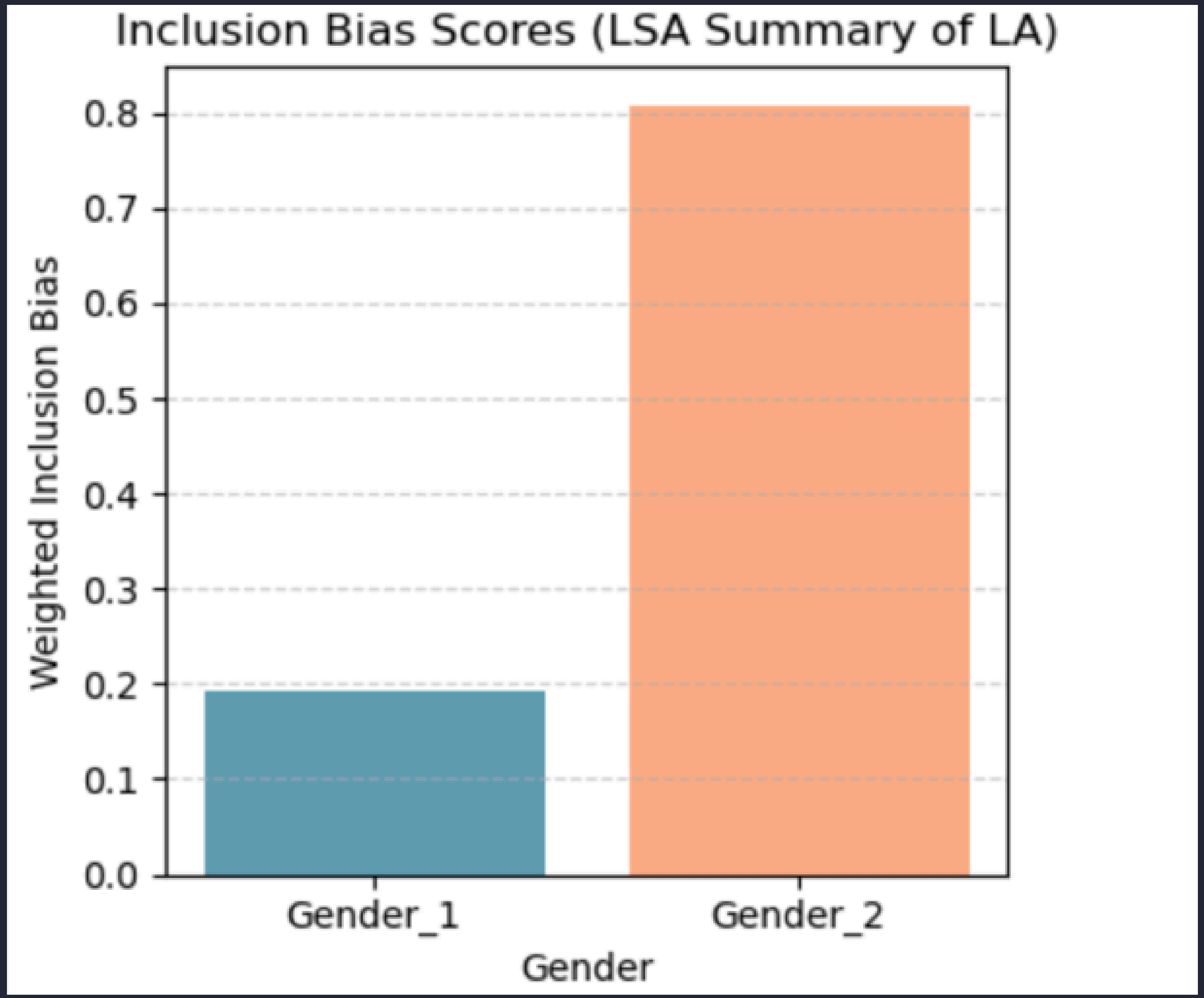


Results(LA)

LSA

Weighted gender 1 Score: 0.19
Weighted gender 2 Score: 0.80
Weighted Inclusion Bias Score
(gender1 - gender2): 0.61

Summary is biased toward gender 2.

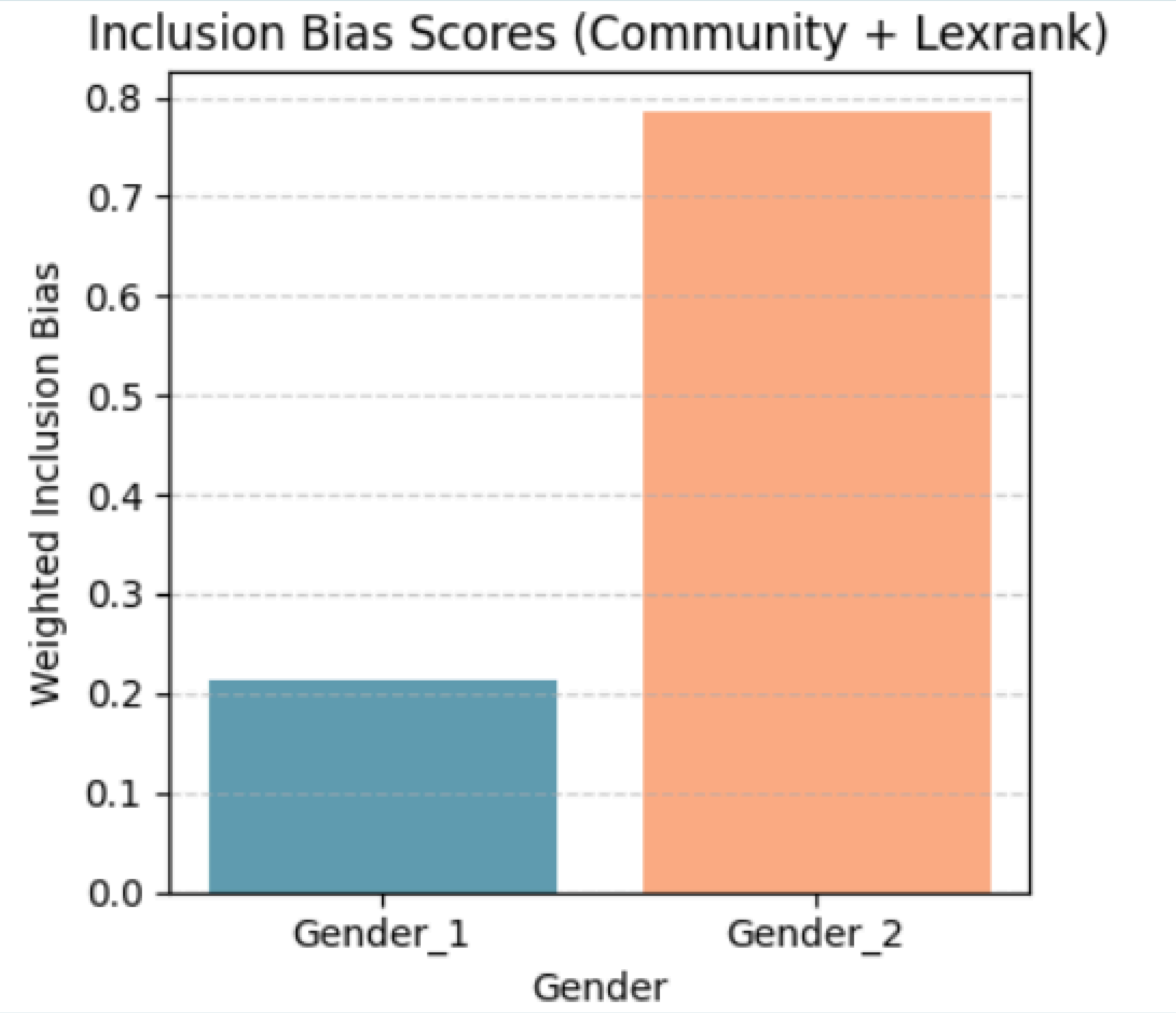


Results(LA)

Community+Lexrank

Weighted gender 1 Score: 0.21
Weighted gender 2 Score: 0.78
Weighted Inclusion Bias Score
(gender1 - gender2): 0.57

Summary is biased toward gender 2.

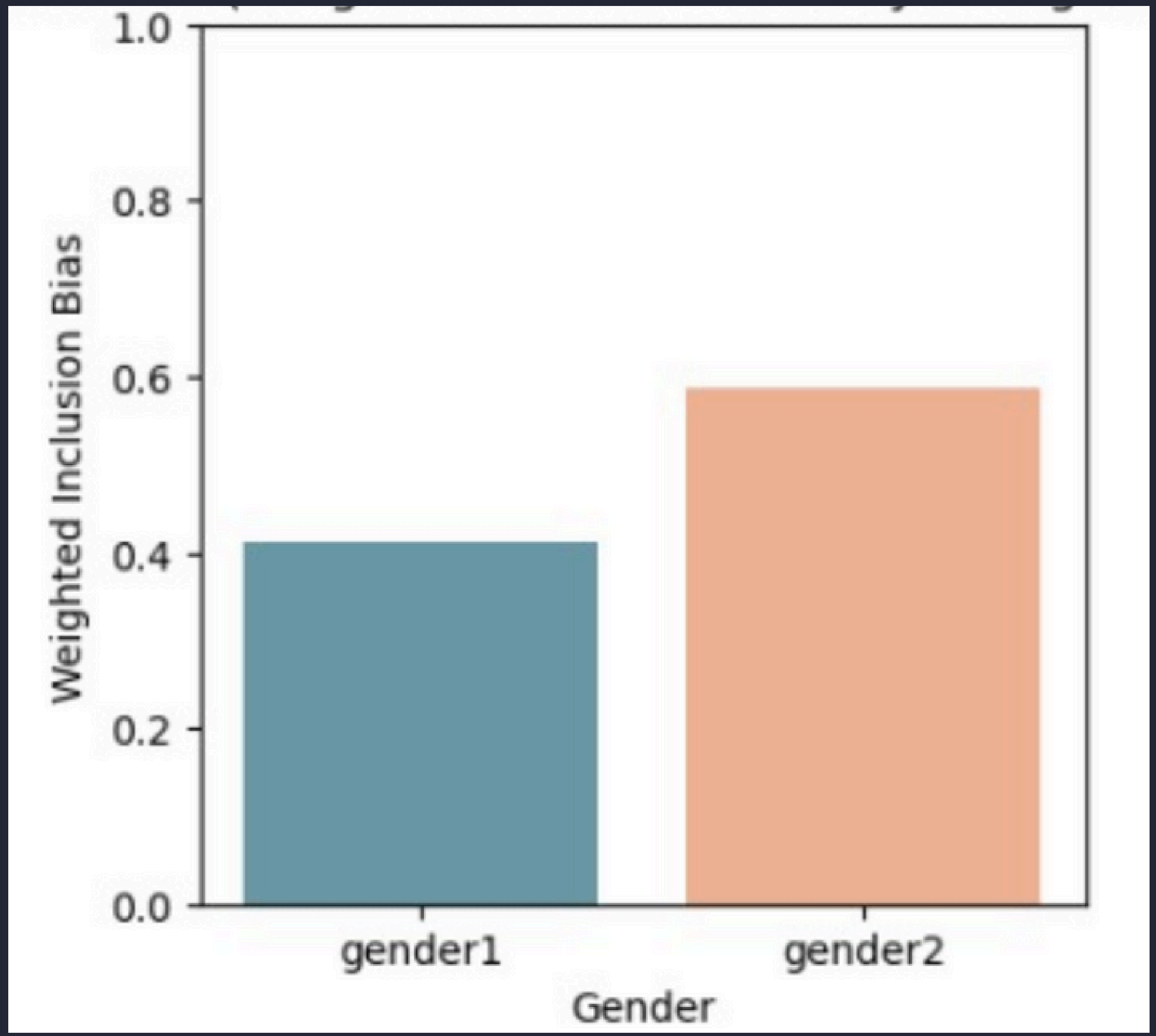


Results(LA)

Gender + Lerank

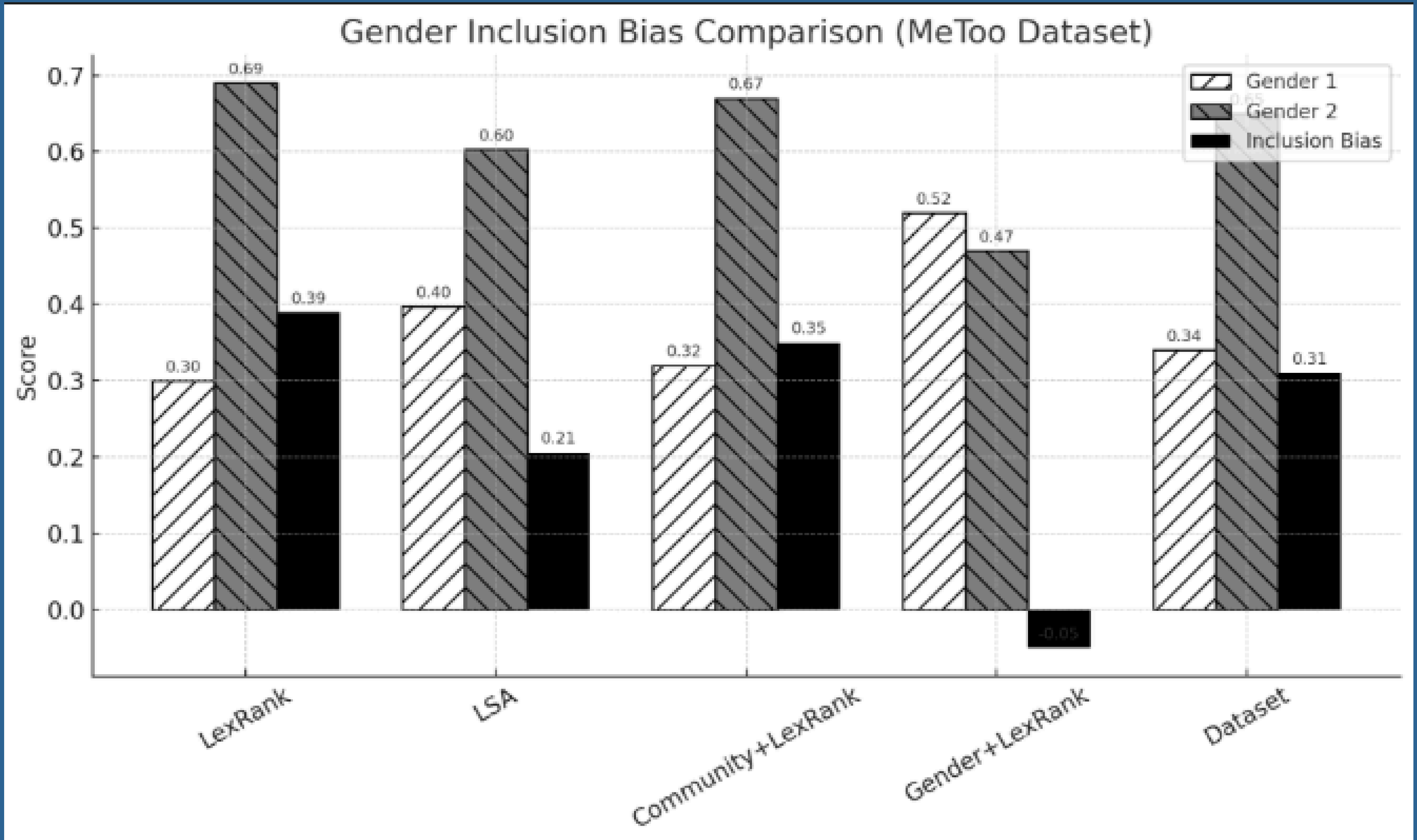
Weighted gender 1 Score: 0.43
Weighted gender 2 Score: 0.57
Weighted Inclusion Bias Score
(gender1 - gender2): 0.1 8

Summary is biased toward gender 2.



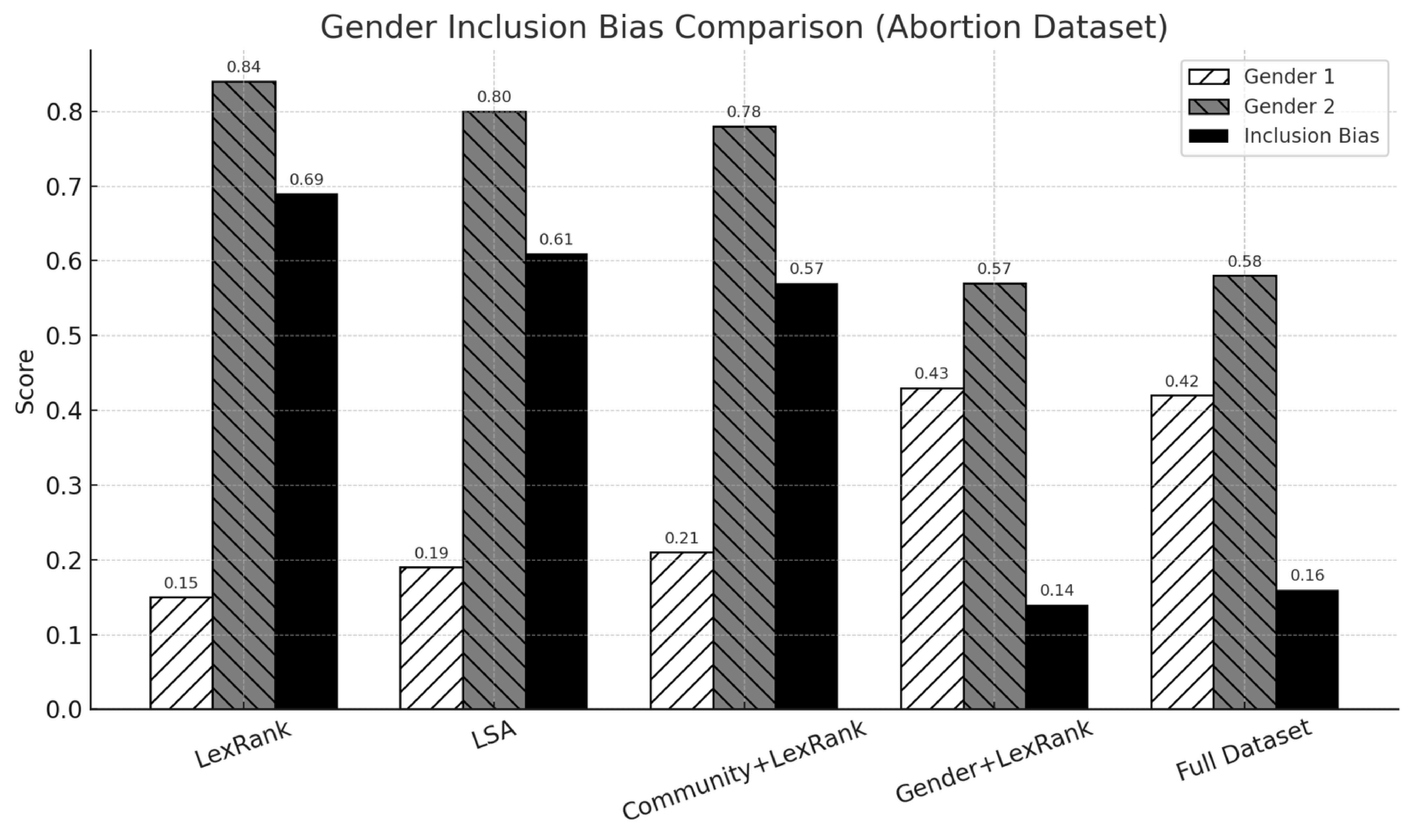
Results

MeToo



Results

LA



Gender Inclusion Bias Comparison (Abortion Dataset)
Score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Gender 1
Gender 2
Inclusion Bias
0.15
0.84
0.69
0.19
0.80
0.61
0.21
0.78
0.57
0.43
0.57
0.14
0.42
0.58
0.16
LexRank
LSA
Community+LexRank
Gender+LexRank
Full Dataset

THANKYOU!

References :
[1] AlDayel, Abeer, and Walid Magdy. ”Stance detection on social media: State of the art and trends.” Information Processing & Management 58, no. 4 (2021): 102597. [1]
[2] Dash, Abhisek, Anurag Shandilya, Arindam Biswas, Kripabandhu Ghosh, Saptarshi Ghosh, and Abhijnan Chakraborty. ”Summarizing user-generated textual content: Motivation and methods for fairness in algorithmic summaries.” Proceedings of the ACM on Human-Computer Interaction 3, no. CSCW (2019): 1-28. [2]
[3] Steen, Julius, and Katja Markert. ”Bias in News Summarization: Measures, Pitfalls and Corpora.” arXiv preprint arXiv:2309.08047 (2023). [3]
[4] Dacon, Jamell, and Haochen Liu. ”Does gender matter in the news? detecting and examining gender bias in news articles.” In Companion Proceedings of the Web Conference 2021, pp. 385-392. 2021. [4]