TABLE OF CONTENTS

LIST OF FIGURES

1. Introduction

Echo chambers in social media refer to situations where individuals are surrounded by information and opinions that reinforce their existing beliefs and perspectives. In other words, they are closed informational environments where people are more likely to be exposed to information that aligns with their preexisting views, and less likely to encounter diverse perspectives or alternative viewpoints. This can lead to the reinforcement and polarization of existing beliefs, and can make it more difficult for individuals to engage in open-minded dialogue and critical thinking.

One of the primary drivers of echo chambers in social media is the algorithmic curation of content. Platforms like Facebook, Twitter, and YouTube use complex algorithms to determine which content is displayed to which users. These algorithms are designed to maximize engagement and keep users on the platform for as long as possible, which often means showing them content that aligns with their preexisting beliefs and interests. Over time, this can create a feedback loop in which users are only exposed to content that reinforces their existing views, while content that challenges those views is hidden from view.

Reddit is a social media platform that is organized around communities or "subreddits" that focus on specific topics or interests. Users can submit content, which can be upvoted or downvoted by other users, and participate in discussions with other members of the community. Reddit is known for its large and diverse user base and while the platform has the potential to foster open dialogue and diverse perspectives, it is also believed that these subreddits are often heavily moderated by users who share similar beliefs and opinions, and dissenting views are frequently downvoted or even censored. This can create an environment in which users are only exposed to information and opinions that reinforce their existing beliefs, and can make it difficult to engage in meaningful dialogue with people who hold different views.

The topic of abortion rights is one of the most contentious and polarizing issues in modern society, with opinions and beliefs varying from pro-choice to pro-life. It is an issue that remains at the forefront of public discourse, and it is essential to study its impact on political polarization and extremism. Given the nature of the subject, it is highly probable that echo chambers exist on Reddit regarding this topic, where individuals may primarily interact with others who share similar viewpoints while dismissing or ignoring opposing opinions. Studying the existence of such echo chambers would provide valuable insights into how online communities operate and how people's beliefs and opinions are shaped in such an environment.

2. Literature Review

Authors and Year	Title	Concept/ Theoretical Model/ Framework	Methodology Used/ Implementation	Dataset details/ Analysis	Relevant Finding	Future Research/Gaps Identified
Michael Conover et al, 2011	Political Polarization on Twitter	Tweet analysis of communication of political information	Cluster analysis of networks and Manual classification of twitter users by political alignment	Twitter 'gardenhose' API	Users often retweet just other users with whom they agree politically when engaging in political discourse	Exploring more roles of technologically mediated political interaction in deliberative democracy
Rong-Ching Chang et al, 2023	#RoeOverturned: Twitter Dataset on the Abortion Rights Controversy	Keyword collection strategy	Hashtags are categorized into neutral, pro-choice and pro-life and generate a retweet network	Twitter Academic API's Full Archive Search	Opinion dynamics, Polarization and Protest mobilization	Data collection relied on a list of relevant keywords, which may not fully capture the extent of the event
David Gunnarsson Lorentzen, 2012	Polarization in Political Twitter Conversations	Analysis of relationships across profiles	Categorizing actors as left, center or right based on profile descriptions and performing social network analysis	yourTwapperKeeper	Actors follow accounts related to other groups only to some extent	Restricted to Swedish government
Jordan K. Kobellarz, Miloš Brocic, et al, 2022	Reaching the bubble may not be enough: news media role in online political polarization	Fleiss' Kappa assessment (agreement degree) and semi-supervised machine learning algorithm (similarity between nodes)	Classify users according to their political orientation, which was performed on a weekly basis	Twitter streaming API regarding 2018 Brazilian presidential elections and 2019 Canadian federal elections	Users tend to engage only with content that reinforces their convictions, regardless of the topic.	Research could be built to understand the user's reaction through discussions, when exposed to information shared by news media
Marcela Suarez Estrada et al, 2022	Toxic Social Media: Affective Polarization After	Customized query search and mixed quantitative/ qualitative	Use of machine learning models based on perspective API to provide the	Tweets published after feminist protests in	The tweets tend to instrumentalize affects, such as	Further research on how affects are instrumentalized to sustain patriarchy on

	Feminist Protests	analysis of tweets	following attributes: toxicity, severe toxicity, identity attack, insult, profanity, and threat.	the period from 16 August 2019 to 20 March 2020.	shame and hate, against the feminist mobilization to create an "enemy"	social media and how the affective feminist politics are mobilized to dispute this.
Ray Block Jr. et al, 2022	Perceived risk, political polarization, and the willingness to follow COVID-19 mitigation guidelines	Negative binomial regression models and logit models to predict mitigation behavior	Natural language processing to identify posts advocating non-compliance	Public posts on Twitter and a nationally representative survey, administered by AARC	Links between political orientation and willingness to follow mitigation guidelines depend on perceived risk.	Exploring dependency of areal death rates and adherence/ compliance.
Kathie Treen et al, 2022	Discussion of Climate Change on Reddit: Polarized Discourse or Deliberative Debate?	Finding sub communities within subreddits using reply networks and measuring topics polarity	ClimateBiasScore (CBS) along with reply network to form communities	BigQuery, 2019a	No strong evidence of polarization except for one rogue community was found.	Finding what features of reddit contribute to the polarization and what can be done for the same
Jason Baumgartner et al, 2019	The Pushshift Reddit Dataset: impact on the Research work	Provide an API endpoint for research work to access and parse reddit data easily	PostgreSQL, ElasticSearch and Ingest engine	Public posts and data on Reddit	The data was very difficult to parse through the regular API	Adding more functionality and search terms and parameters
Emily Chen, Ashok Deb, Emilio	#Election2020: the first public Twitter dataset on the 2020 US Presidential election	Topic modeling techniques	Use of machine learning algorithms to identify relevant tweets and their sentiment.	The dataset includes tweets collected between August 1, 2020, and January 31, 2021, using	A comprehensive way to understand the use of Twitter during the 2020 US Presidential	Future research could explore the ethical implications of social media use in politics, particularly with regard to privacy, transparency, and

				the Twitter API.	election and how it influenced public opinion. The paper also provides a valuable dataset for further research in this area.	accountability.
Ferrara, 2022						
Aleksandra Urman, 2019	Context matters: political polarization on Twitter from a comparative perspective	Network-analytic audience duplication approach	The political Twitterspheres of the countries are classified as perfectly integrated, integrated, mixed, polarized and perfectly polarized	The paper does not provide details about a particular dataset.	Political polarization on Twitter is not uniform across countries Different issues generate polarization in different countries Network dynamics can reinforce polarization	The relationship between online and offline polarization needs further investigation The impact of social media algorithms on polarization is not yet clear
Anindita Borah & Sanasam Ranbir Singh , 2022	Investigating political polarization in India through the lens of Twitter	Combination of identity-based theory, framing theory, and social media analysis to understand political polarization	The investigation performs social network analysis and content analysis of the tweets posted by Indian politicians during some major events in India from 2019 to 2021.	A sample of tweets using Twitter's Streaming API, using keywords related to the major political parties and candidates in India. The dataset includes a total of 22,130 tweets collected from April 11, 2019 to May 19, 2019.	The study provides insights into the complex dynamics of political polarization on Twitter in India and highlights the importance of considering both identity-based factors and framing effects in understanding these dynamics.	Investigating the impact of interventions aimed at reducing polarization Exploring the role of offline factors in driving polarization on Twitter Examining the role of different social media platforms in shaping political discourse

3. Objective

The primary objective of this project is to identify the existence of echo chambers within Reddit and understand the characteristics that lead to their formation. The study will use a mix of quantitative and qualitative methods to investigate the prevalence and traits of these echo chambers on Reddit. The topic chosen for this study is 'Abortion Rights' due to its controversial nature. This study aims to identify and analyze pro-life as well as pro-choice echo chambers that may have formed on the Reddit forum.

4. Innovation

A large proportion of studies analyzing echo chambers usually work on Twitter APIs as the assessment and mapping of individual identities and beliefs across communities is relatively easier. This project, however, aims at looking at a relatively lesser analyzed platform, i.e., Reddit.

Moreover, most researchers create social graphs to identify the sub communities within a social media platform. This project's main objective is to calculate the similarity of ideology in the comments of a Reddit post and assign scores to indicate the polarity of the same.

This project assesses the communities of Pro-lifers and Pro-choicers and examines the spread and overlap of both ideologies on a social media platform, while most papers on this topic do not perform echo chamber analyses.

5. Implementation

5.1 Software Requirements

- Python
- Reddit API
- PRAW module
- Numpy, pandas
- BERT framework
- Cardiffnlp/Twitter-roberta-base-stance-abortion model
- Pyplot
- Profanity (Python module)

5.2 Hardware Requirements

The code for this project was tested on a laptop with an Intel(R) Core (™) i7-8750H CPU @ 2.2 GHz with 6 cores. 16GB of RAM was available as memory.

5.3 Methodology

Echo chambers are communities where people are predominantly exposed to opinions and information that confirm their existing beliefs, while dissenting views are either absent or actively suppressed. Understanding the dynamics of echo chambers is important for promoting informed and diverse public discourse on controversial issues such as abortion. The purpose of this study is to investigate echo chambers in Reddit related to abortion rights.

5.3.1 Data Collection

Firstly, a list of all subreddits related to abortion rights was compiled. Each subreddit's relevance was manually verified. As a result, 32 subreddits and 10 individual posts were identified that were likely to represent a diverse range of opinions and perspectives on the issue.

Next, the Reddit API and PRAW (Python Reddit API Wrapper) module were used to extract the top 10 posts for each subreddit. Then the comment forest for each post were collected.

5.3.2 Data Filtering

To focus on comments that were specifically related to abortion rights, a list of relevant keywords (e.g., "abortion," "pro-choice," "reproductive," "women", "rights," etc.) was used to filter out comments that did not pertain to the topic. Comments that were too short or lacked any meaningful content were also excluded.

5.3.3 Data Analysis

To classify the remaining comments as pro-choice, pro-life, or neutral, the BERT machine learning framework and the Cardiffnlp/Twitter-roberta-base-stance-abortion model from Hugging Face was implemented. This model was trained on a large dataset of tweets related to abortion and is able to detect the stance of a given tweet as being pro-choice, pro-life, or neutral. The total score for each post was computed by calculating the sums of the scores of all classified comments. Finally, the distribution of these scores across subreddits and posts was analyzed to identify potential echo chambers. This was

done by setting a threshold of 80%, i.e., if 80% of the comments on a post are leaning towards a particular ideology, it is labeled as being an echo chamber for the same. Similarly, mixed posts were calculated by setting a threshold of 20% of pro-choice comments as well as 20% pro-life comments. It may be noteworthy to mention that the first 940 characters of the comments were utilized, as this limit represents the maximum number of tokens that the HuggingFace model can accommodate.

5.4 Novelty from Reference Project

The reference project, research paper "Discussion of Climate Change on Reddit: Polarized Discourse or Deliberative Debate?" [7], analyzes Climate Change polarized ideas within Reddit.

The topic of "Abortion Rights" is a highly debated topic amongst the youth, i.e., the main user demographic of Reddit, and thus might give us an insight into how communities can be allowed to grow on such platforms without any intervention.

The paper's novelty lies in using a model to assign a polarity on the spectrum to each post while focusing first on identifying the sub communities and then thinking about whether they are echo chambers or not.

5.5 Tools Used

- **Python**: Python is useful for Social Network Analysis and working with Reddit's API due to its flexibility and extensive libraries like NetworkX, PRAW, and pandas, which provide tools for data manipulation, analysis, and visualization.
- **Reddit API & PRAW**: The Reddit API is used to fetch posts in each subreddit as well as comment forest of each post while the PRAW module allows easy use with Python.
- **BERT & Cardiffnlp/Twitter-roberta-base-stance-abortion model:** The BERT framework along with the pre-trained model from Hugging Face is used to perform sentiment analysis on the comments and classify as pro-choice, pro-life or neutral.
- **Pyplot**: Pyplot is a Python matplotlib API which was used to visualize the data generated upon analysis. These visualizations were used to draw the study's conclusions.
- **Profanity:** This Python library is used to check for offensive language in comments and classify as offensive or non-offensive.

5.6 Screenshots of Demo along with Visualization

```python
for submission in reddit.subreddit(combined_list[0]).hot(limit=3):
    print("Post Title ", submission.title)

    top_level_comment_count = 0
    for top_level_comment in submission.comments:
        print("top level comment: ", top_level_comment.body[0:30], '...')

        for second_level_comment in top_level_comment.replies:
            print("\tsecond level comment: ", second_level_comment.body[0:30], '...')

            for third_level_comment in second_level_comment.replies:
                print("\t\tthird level comment: ", third_level_comment.body[0:30], '...')

        top_level_comment_count += 1
        if top_level_comment_count == 3:
            break
        print()

    print("-------------------------")
    print()
```

```
Post Title  Need Links/Phone Numbers/ Resources for crisis pregnancy centers and others akin
top level comment:  **International resource regis ...
        second level comment:  US pregnancy map should be rem ...

top level comment:  Irish pro-life pregnancy centr ...
        second level comment:  Gallimh abu ...

top level comment:  These sites routinely host fun ...
-------------------------

Post Title  Pro Life Weekly Chat!
top level comment:  So anyone here play super robo ...
        second level comment:  No, but that sounds really fun ...
                third level comment:  https://youtu.be/GxzMmMeusD4 ...
```

Fig 5.1: Fetching posts and comments using RedditAPI and PRAW

```python
tokenizer = AutoTokenizer.from_pretrained('cardiffnlp/twitter-roberta-base-stance-abortion')
model = AutoModelForSequenceClassification.from_pretrained('cardiffnlp/twitter-roberta-base-stance-abortion')
```

```python
def relevant_comment(text):
    keywords = ["pro-life", "pro-choice", "abortion", "fetus", "pregnancy", "reproductive rights", "women's rights", "roe v. wade
    new_keywords = ["heartbeat","brain","conception","ban","child", "access", "reproductive", "healthcare", "women's autonomy", "
    final_keywords=keywords+new_keywords
    for word in final_keywords:
        if word in text.lower():
            return True
    return False
```

```python
def sentiment_score(comment):
    tokens = tokenizer.encode(comment, return_tensors='pt')
    result = model(tokens)
    return int(torch.argmax(result.logits))+1
```

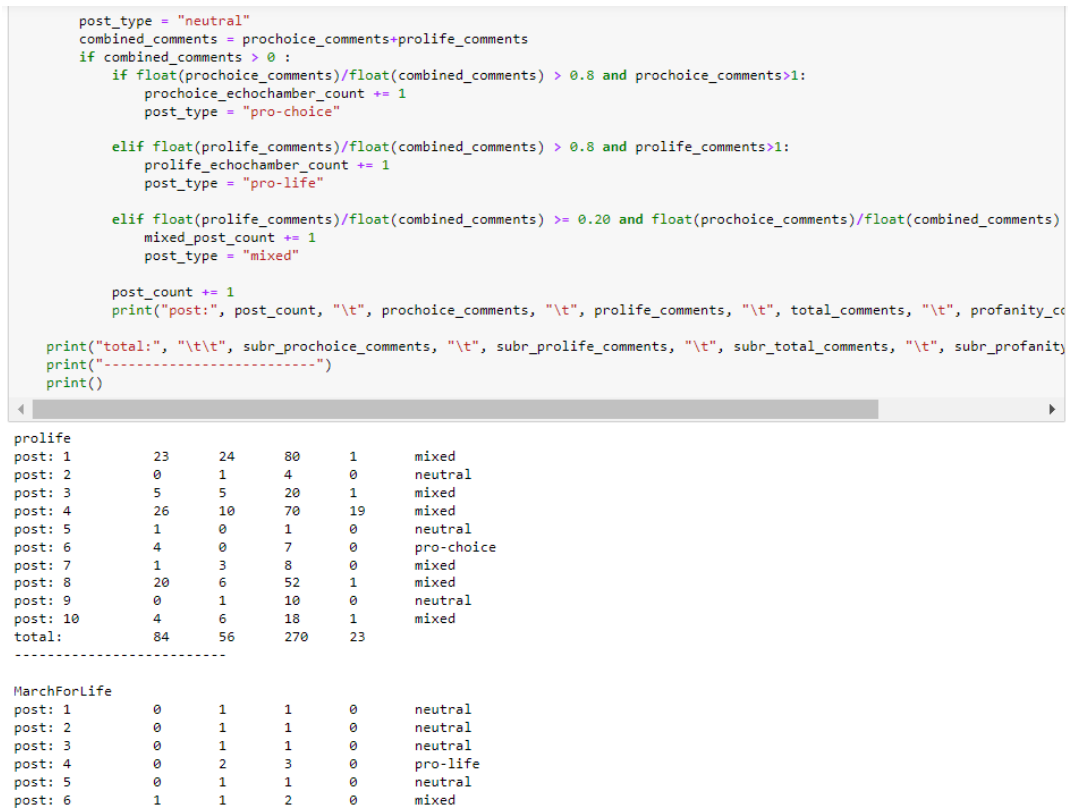*Fig 5.2: Filtering and Sentiment Analysis of comment*

```
        post_type = "neutral"
        combined_comments = prochoice_comments+prolife_comments
        if combined_comments > 0 :
            if float(prochoice_comments)/float(combined_comments) > 0.8 and prochoice_comments>1:
                prochoice_echochamber_count += 1
                post_type = "pro-choice"

            elif float(prolife_comments)/float(combined_comments) > 0.8 and prolife_comments>1:
                prolife_echochamber_count += 1
                post_type = "pro-life"

            elif float(prolife_comments)/float(combined_comments) >= 0.20 and float(prochoice_comments)/float(combined_comments)
                mixed_post_count += 1
                post_type = "mixed"

        post_count += 1
        print("post:", post_count, "\t", prochoice_comments, "\t", prolife_comments, "\t", total_comments, "\t", profanity_co

    print("total:", "\t\t", subr_prochoice_comments, "\t", subr_prolife_comments, "\t", subr_total_comments, "\t", subr_profanity
    print("-------------------------")
    print()
```

```
prolife
post: 1      23    24    80    1      mixed
post: 2      0     1     4     0      neutral
post: 3      5     5     20    1      mixed
post: 4      26    10    70    19     mixed
post: 5      1     0     1     0      neutral
post: 6      4     0     7     0      pro-choice
post: 7      1     3     8     0      mixed
post: 8      20    6     52    1      mixed
post: 9      0     1     10    0      neutral
post: 10     4     6     18    1      mixed
total:       84    56    270   23
-------------------------

MarchForLife
post: 1      0     1     1     0      neutral
post: 2      0     1     1     0      neutral
post: 3      0     1     1     0      neutral
post: 4      0     2     3     0      pro-life
post: 5      0     1     1     0      neutral
post: 6      1     1     2     0      mixed
```

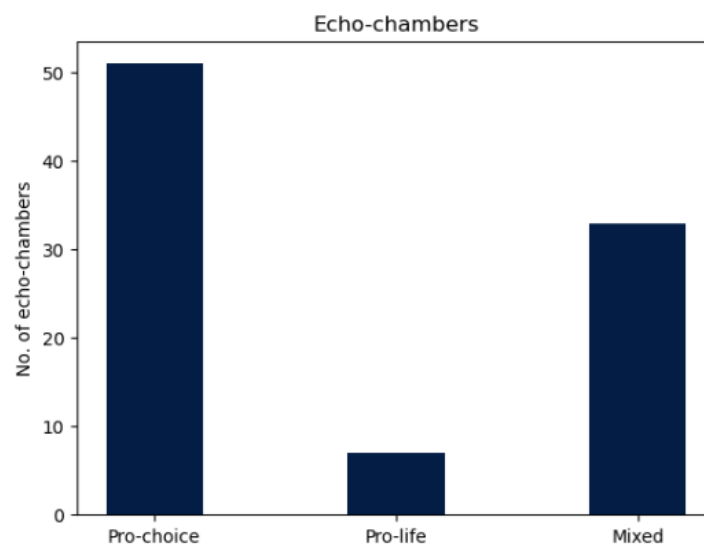*Fig 5.3: Classification of echo-chamber*
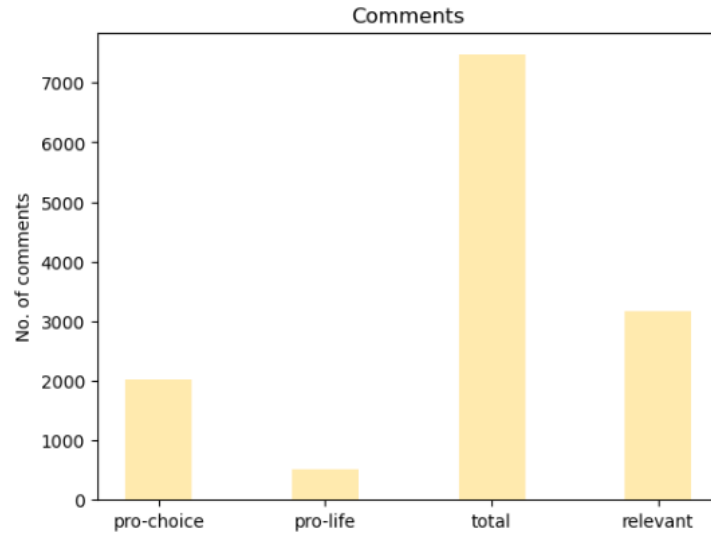

Fig 5.4: Number of echo chambers
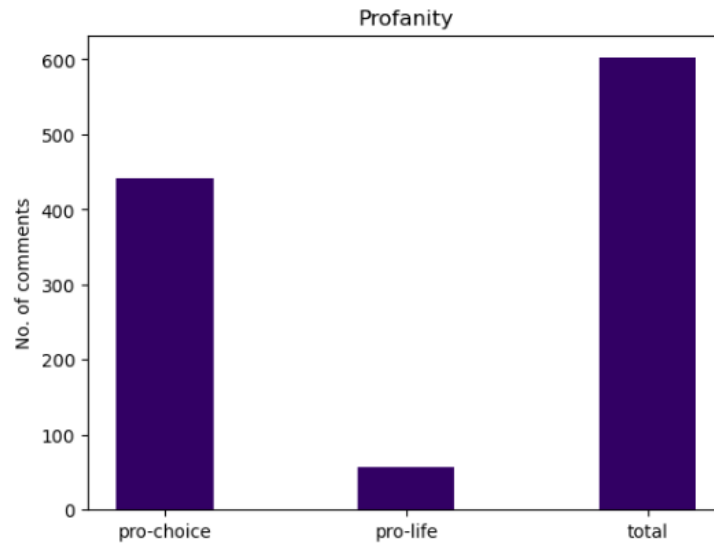
Fig 5.5: Number of comments analyzed

Fig 5.6: Number of comments with profanity

6. Results and Discussion

The initial assumption was that the pro-life community would be highly polarized and result in having a majority of echo chambers, as pro-lifers are considered to hold strong and often opposing beliefs about the morality of abortion, with some advocating for a complete ban while

others supporting certain exceptions. Additionally, the issue of abortion is often intertwined with political, religious, and ethical considerations, further contributing to the polarization of pro-life perspectives.

On investigating echo chambers within Reddit while focusing on the topic of abortion rights, the results uncovered evidence of highly polarized communities. It was observed that these communities mainly exhibited confirmation bias, where members share and promote information that reinforces their pre-existing beliefs while rejecting information that challenges them.

Moreover, an interesting trend identified is that this bias was especially clear when looking at echo chambers classified as Pro-Choice. The amount of Pro-Life echo chambers found were far less compared to pro-choice echo chambers. Even in posts/communities where members were supporting the Pro-Life movement, the discourse was mixed and couldn't be classified as an echo chamber. It was observed that pro-choice supporters were more active on the forum and voiced their opinions in larger numbers resulting in more comments as well as echo chambers being classified as pro-choice, which was completely against this study's initial assumption.

6.1 Polarization

The observed trend suggests that pro-choice users are more likely to engage with both pro-choice and pro-life communities, indicating a greater willingness to participate in dialogue and consider opposing viewpoints. In contrast, pro-life users appear to primarily communicate with those who share their beliefs, potentially contributing to greater polarization and a lack of understanding of alternative perspectives.

Contrary to intuition, more pro-choice echo chambers were found because the pro-choice community was debating and expressing views within the pro-choice community as well as with the pro-life community. Concurrently, pro-life communities resulted in lesser echo chambers and polarization even though their interaction was primarily with other pro-lifers.

6.2 Age

Reddit, as a social media platform, appeals to a younger audience. As a general trend, the youth tends to lean towards leftist and liberal political principals, which is in favor of abortion rights. Hence, a huge proportion of Reddit users speak in support of the pro-choice community, which may be the reason for the large number of pro-choice discussions as well as echo chambers as compared to pro-life discussions and echo chambers.

7. References

[1] Conover, M., Ratkiewicz, J., Francisco, M., Goncalves, B., Menczer, F., & Flammini, A. (2021), "Political Polarization on Twitter", Proceedings of the International AAAI Conference on Web and Social Media, *5*(1), 89-96

[2] Rao, Ashwin; Rong-Ching Chang; Qiankun Zhong; Magdalena Wojcieszak; Kristina Lerman, 2023, "#RoeOverturned: Twitter Dataset on the Abortion Rights Controversy", Harvard Dataverse

[3] Lorentzen, David. (2014). "Polarization in political Twitter conversations". Aslib Journal of Information Management. 66. 10.1108/AJIM-09-2013-0086

[4] Kobellarz, J.K., Broćić, M., Graeml, A.R. et al. "Reaching the bubble may not be enough: news media role in online political polarization". EPJ Data Sci. 11, 47 (2022)

[5] Suarez Estrada, M., Juarez, Y., & Piña-García, C. A. (2022). "Toxic Social Media: Affective Polarization After Feminist Protests". Social Media + Society, 8(2)

[6] Block R Jr, Burnham M, Kahn K, Peng R, Seeman J, Seto C. "Perceived risk, political polarization, and the willingness to follow COVID-19 mitigation guidelines". Soc Sci Med. 2022 Jul; 305:115091

[7] Kathie Treen, Hywel Williams, Saffron O'Neill & Travis G. Coan (2022) "Discussion of Climate Change on Reddit: Polarized Discourse or Deliberative Debate?", Environmental Communication, 16:5, 680-698, DOI: 10.1080/17524032.2022.2050776

[8] Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020). "The Pushshift Reddit Dataset". Proceedings of the International AAAI Conference on Web and Social Media, 14(1), 830-839

[9] Chen, E., Deb, A. & Ferrara, E. "#Election2020: the first public Twitter dataset on the 2020 US Presidential election". J Comput Soc Sc 5, 1–18 (2022)

[10] Urman, A. (2020). "Context matters: political polarization on Twitter from a comparative perspective". Media, Culture & Society, 42(6), 857–879

[11] Borah, A., Singh, S.R. "Investigating political polarization in India through the lens of Twitter". Soc. Netw. Anal. Min. 12, 97 (2022)

Appendix A : Data Analysis Output

```
ProLifeVegan
post: 1          0      2      15      0      pro-life
post: 2          0      1      1       0      neutral
post: 3          0      1      2       0      neutral
post: 4          0      1      1       0      neutral
post: 5          0      1      1       0      neutral
total:           0      6      31      0
-------------------------

prochoice
post: 1          9      4      33      3      mixed
post: 2          1      0      3       0      neutral
post: 3          3      0      7       2      pro-choice
post: 4          6      0      11      0      pro-choice
post: 5          4      4      11      1      mixed
post: 6          4      0      17      1      pro-choice
post: 7          20     1      48      10     pro-choice
post: 8          1      0      1       0      neutral
post: 9          1      0      3       0      neutral
total:           49     9      136     17
-------------------------

abortion
post: 1          3      0      7       0      pro-choice
post: 2          2      0      4       0      pro-choice
post: 3          2      0      3       0      pro-choice
post: 4          2      0      5       0      pro-choice
post: 5          3      0      4       0      pro-choice
post: 6          4      0      13      0      pro-choice
post: 7          1      0      48      0      neutral
total:           17     0      94      0
```

Fig A.1: Individual Post Analysis Output

```
"My body my choi ...    186    73     470    50     mixed
TIL that saying  ...    11     0      20     1      pro-choice
Women charged po ...    63     13     440    23     pro-choice
The Pro-Choice M ...    22     10     57     10     mixed
In a few hours t ...    6      1      28     1      pro-choice
CMV: Abortions a ...    56     19     102    10     mixed
CMV: The pro-lif ...    24     14     60     14     mixed
Why is healthcar ...    56     28     143    19     mixed
Federal judge ha ...    92     22     498    34     pro-choice
And there it is, ...    87     20     501    42     pro-choice
```

Fig A.2: Final output

```
total post_count 287
prochoice echochamber_count 51
prochoice echochamber_count 7
mixed post 33

total_comment_count 7468
total_relevant_comment_count 3163
prochoice_comment_count 2020
prolife_comment_count 523

total_profanity_comment_count 602
prochoice_profanity_comment_count 442
prolife_profanity_comment_count 57
```

Fig A.3: Final output