

A3: Political Bias

In this assignment, you will be trying to calculate the average political bias and reliability of posts in a subreddit.

The code you are starting with here already does a search on a subreddit and finds the reliability and political bias of url web addresses posted to subreddits. You will need to add loop variables to calculate these averages (see chapter 8 practice and demos).

After you get the averages to work, you will then try your code on other subreddits, and then you will answer some reflection questions.

First, we'll do our normal Reddit PRAW login steps (though we'll not use fake-praw for this assignment, only real Reddit).

Reddit PRAWSetup

```
In [2]: import praw
```

```
In [3]: %run reddit_keys.py
```

```
In [4]: # Give the praw code your reddit account info so
# it can perform reddit actions
reddit = praw.Reddit(
    username=username, password=password,
    client_id=client_id, client_secret=client_secret,
    user_agent="a custom python script for user /" + str(username)
)
```

Version 7.7.0 of praw is outdated. Version 7.7.1 was released Tuesday July 11, 2023.

Load Bias and Reliability Info

The code to load the bias and reliability info is in another file: `a3-supporting_code.ipynb`. You can look at that file if you are interested, but you are not required to.

The part you should know is that the measures of website bias and reliability are based on the [Media Bias Chart](#) (old version 9.0). We took their ratings of reliability and bias and simplified them into a scale with bias ranging from -4 (extremely liberal) to +4 (extremely conservative), and reliability from +2 (fact reporting) to -4 (fabricated or inaccurate). We then chose a few of the more common websites to let us look up info.

Media bias chart with grid, showing the range labels

```
In [5]: # load the pandas code library, which helps us use the csv file
import pandas as pd

# load the media_info.csv file
```

```
media_info_df = pd.read_csv('media_info.csv')

# get a list of the sites we have information for in the media_info.csv
media_sites = media_info_df['site']

def find_matching_site(url):
    for site in media_sites:
        if site in url:
            return site

# make a lookup dictionary, where the key is the site, and the value is the
media_bias_lookup = {m_info['site']: m_info['bias'] for i, m_info in media_in

# make a lookup dictionary, where the key is the site, and the value is the
media_reliability_lookup = {m_info['site']: m_info['reliability'] for i, m_in
```

Get a list of results from a subreddit

We will now get a list of results from reddit. To start with we get 20 posts from the "news" subreddit

In [6]:

```
# Look up the subreddit "news", then find the "hot" list, getting up to 100 s
submissions = reddit.subreddit("news").hot(limit=20)

# Turn the submission results into a Python List
submissions_list = list(submissions)
```

TODO: Modify the code below (Run Search)

The code below loops through each reddit submission, and if the submission was a website url, the program checks to see if we have reliability/bias info on the site. If we have that info we calculate the bias and reliability and display it.

Your job is to add loop variables to the code to calculate the number of urls we had info for (`number_matched_urls`) and then the total bias and total reliability for those urls. Then you can use that at the end to calculate the average bias and average reliability.

In [7]:

```
number_matched_urls = 0
total_url_bias = 0
total_url_reliability = 0

for submission in submissions_list:

    print(submission.url)

    if submission.url:

        matching_site = find_matching_site(submission.url)

        if matching_site:

            url_bias = media_bias_lookup.get(matching_site, 0)
            url_reliability = media_reliability_lookup.get(matching_site, 0)
            number_matched_urls += 1
            total_url_bias += url_bias
            total_url_reliability += url_reliability
```

```

        print("  Bias: " + str(url_bias))
        print("  Reliability: " + str(url_reliability))
    else:
        # We didn't have info on this site
        print("** Did not recognize site!")

```

```

print()

```

```

average_bias = total_url_bias / max(number_matched_urls, 1) # Avoid division
average_reliability = total_url_reliability / max(number_matched_urls, 1) #

```

```

print("-----")
print("Total number of URLs we could measure:", number_matched_urls)
print("Average bias:", average_bias)
print("Average reliability:", average_reliability)

```

<https://www.waff.com/2024/02/23/life-support-hospital-officials-warn-dozens-ural-alabama-hospitals-are-risk-closing-their-doors/>
 ** Did not recognize site!

https://www.fox23.com/news/owasso-police-release-body-cam-video-911-call-lead-ing-up-to-students-death-following-fight/article_6bc7bb00-d28a-11ee-a4ba-b762d43b180b.html
 ** Did not recognize site!

<https://globalnews.ca/news/10312716/mom-vacation-toddler-dead-home-guilty-murder-kristel-candelario/>
 ** Did not recognize site!

<https://www.bbc.co.uk/news/business-68379318>
 Bias: -1
 Reliability: 1

<https://abc7.com/rebecca-grossman-trial-boys-crash/14461388/>
 ** Did not recognize site!

https://abcnews.go.com/US/jury-finds-nra-liable-mismanagement-wayne-lapierre-violated/story?id=107269909&cid=social_twitter_abcn
 Bias: -1
 Reliability: 1

<https://apnews.com/article/mexico-president-published-reporters-phone-number-ef537c9c37b3506ff67ea18ebe40629e>
 Bias: -1
 Reliability: 2

<https://www.cbsnews.com/news/florida-measles-outbreak-unvaccinated-kids-school/>
 ** Did not recognize site!

<https://www.nbcnews.com/news/us-news/man-found-guilty-alaska-native-womans-killing-was-recorded-video-rcna140145>
 Bias: -2
 Reliability: 1

<https://www.cnbc.com/2024/02/23/united-airlines-raises-checked-bag-fee-following-american.html>
 Bias: -2
 Reliability: 1

<https://www.theguardian.com/us-news/2024/feb/23/trump-fundraising-wisconsin-felony-charges>
 Bias: -2
 Reliability: 1

<https://www.cnn.com/2024/02/23/us/uga-augusta-university-student-death/index.html>

Bias: -2

Reliability: 1

<https://www.cbsnews.com/news/military-tracking-balloon-western-us-military/>
** Did not recognize site!

<https://scrippsnews.com/stories/white-house-preparing-major-russian-sanctions-after-navalny-s-death/>
** Did not recognize site!

<https://www.washingtonpost.com/weather/2024/02/23/record-warm-february-winter-midwest-east/>
Bias: -2
Reliability: 1

<https://www.wect.com/2024/02/23/troopers-duplin-county-deputy-hits-vehicle-other-driver-killed/>
** Did not recognize site!

<https://www.reuters.com/world/us/us-detects-balloon-over-western-united-states-not-threat-us-official-2024-02-23>
Bias: -1
Reliability: 2

<https://www.cbsnews.com/news/taxes-h-r-block-ftc-intuit/>
** Did not recognize site!

<https://www.theguardian.com/uk-news/2024/feb/23/shamima-begum-loses-appeal-against-removal-of-british-citizenship>
Bias: -2
Reliability: 1

<https://www.bbc.co.uk/news/world-europe-68383608>
Bias: -1
Reliability: 1

Total number of URLs we could measure: 11
Average bias: -1.5454545454545454
Average reliability: 1.1818181818181819

Reflection tasks

Once you get the code above working and finding an average bias and reliability, modify the search to try at least three more subreddits (and get more posts at a time, like 100). Open up the subreddit separately and look at your results, then answer the questions below.

Note: For searches, you can search for different subreddits that might have different views and post links to news articles, like: "news", "science", "politics", "liberal", "conservative", "tech", "BlackLivesMatter", etc.

1. What additional searches did you run (at least 3)?

"science", "liberal", "conservative"

1. When doing those searches, what were your observations about the calculations of media bias and reliability? (For example: were there a lot of urls that you didn't

measure? Do you feel like the final calculated bias and reliability match the search results?). Answer with at least 3 sentences

When conducting these searches, I noticed that the calculations of media bias and reliability varied depending on the subreddit and the types of URLs shared within them. In some subreddits, there were more URLs with recognized sources, leading to more accurate bias and reliability calculations. However, in others, there were a significant number of URLs for which we didn't have information, which affected the overall accuracy of the calculations. Additionally, the final calculated bias and reliability sometimes didn't fully align with the perceived bias of the subreddit, indicating that the methodology or data sources might need refinement.

1. If you could redesign the Media Bias Chart, what would you want to do (e.g., add some other dimension besides just bias/responsibility, change how it is evaluated, add more news sources, consider different countries)? Answer with at least 3 sentences.

If I were to redesign the Media Bias Chart, I would consider adding more dimensions besides just bias/responsibility. This could include factors such as transparency, editorial standards, fact-checking rigor, and political affiliation of the news source. Additionally, I would expand the evaluation to include a broader range of news sources from different countries and cultural contexts to provide a more comprehensive view of media bias globally.

1. What might a social media companies or advertizers (including political campaigns) want to do with information on a users' political views and susceptibility to conspiracy theories? Answer with at least 3 sentences.

Social media companies might want to leverage information on users' political views and susceptibility to conspiracy theories for targeted advertising or messaging campaigns. By analyzing users' behavior, engagement patterns, and content preferences, they can tailor their content and advertisements to resonate more effectively with specific audience segments. This information could also be used to identify potential influencers or opinion leaders within certain communities to amplify messaging and influence public opinion.

1. Choose two ethics frameworks and use the frameworks to consider the different uses of the media bias and reliability information. Answer with at least 6 sentences total (e.g., 3 per framework).

From a utilitarian perspective, media bias and reliability information could be used to maximize overall societal welfare by promoting access to accurate and diverse news sources. By providing users with transparent information about the bias and reliability of different news outlets, individuals can make more informed decisions about the information they consume, leading to a more educated and engaged citizenry. However, from a deontological standpoint, the collection and utilization of user data for profiling and targeting purposes may raise concerns about privacy and autonomy. Ethical considerations must be taken into account to ensure that users' rights are respected and protected, even as we strive to improve media literacy and combat misinformation.