KU LEUVEN

# Article-level Media Bias Classification

Stefan Liemawan Adji

Thesis submitted for the degree of Master of Science in Artificial Intelligence, option Speech and Language Technology

**Supervisor:**
Prof. Miryam de Lhoneux

**Assistant-supervisor:**
Timo Spinde and Tomáš Horych of Media Bias Group

Academic year 2023 – 2024

# Contents

# Abstract

In today's digital, information-rich society, media bias poses a significant challenge to the objectivity and credibility of news reporting. As someone living in our current society, one has inevitably encountered some form of bias in the media, either consciously or unconsciously. Media bias can shape our perceptions, influence our opinions, and affect our understanding on various issues. It is crucial to recognise and address this bias to ensure a well-informed and balanced perspective. By being aware of the inherent biases in media sources, individuals can critically evaluate the information they consume and seek out diverse viewpoints to form a more comprehensive understanding of the world. This thesis project aims to understand forms of media bias appearing in articles and build a system to detect and classify existence of media bias on the article-level.

# Chapter 1

# Introduction

## 1.1 Overview

This thesis project is part of an Advanced Master of Artificial Intelligence programme—Speech and Language Technology, in collaboration with the Media Bias Group [15], who provided the topic and additional guidance along the project.

## 1.2 Motivation

Media misinformation and manipulation are rampant in today's landscape. Despite being a major societal issue, there have been hardly enough resources and work dedicated to the realm of media bias and its broader context.

Considerable work has been done on fake news and its detection, media bias operates on a higher level and cannot be described the same simply fake news. Fake news involves intentionally spreading false information, whereas media bias refers to the distortion or manipulation of information by media outlets, which may or may not be intentional, to favour certain perspectives or agendas. Unlike fake news, media bias can influence public perception subtly through selective reporting, framing, or sourcing, making it a complex and challenging issue to address comprehensively.

To combat media bias effectively, it is crucial to develop robust methodologies and technologies for detecting and analyzing biased content across various media platforms. Additionally, raising awareness about media literacy and critical thinking skills can empower individuals to identify and navigate biased information effectively in today's media landscape. Ultimately, addressing media bias requires concerted efforts from researchers, policymakers, media organisations, and the public to promote transparency, accountability, and integrity in news reporting and consumption.

## 1.3 Goals

To aid in the work of media bias, the goals of this project are defined as follows:

1. Extend current media bias dataset to make it suitable for article-level classification.
2. Propose a method to effectively represent articles in a space vector.
3. Design a system that is able to effectively detect and classify bias on the article level of granularity.
4. Validate the resulting system and dataset.

## 1.4 About the Media Bias Group

The Media Bias Group [15] was established in mid-2020 by Timo Spinde during his pursuit of a Ph.D. in computer science, having been integrated into the topic since his undergraduate studies, with a vision to aid others perceive news in a more balanced and conscious manner. After a year of planning how a system could uncover bias on a vast scale encompassing millions of articles, he founded the group and forged connections with various partners, particularly those relevant to specific aspects of the project. In just one year, the project has garnered support from multiple other research groups, with around twenty students from seven countries joining to contribute to the system. Since 2021, the group has also begun offering its first Ph.D. positions.

The group is comprised of a collective of scholars across various fields such as Psychology, Linguistics, and Computer Science, with a shared goal to comprehend the factors influencing human perception of news content as biased or one-sided. Currently, the network includes six main researchers and coordinators, twenty-one professors and postdocs, as well as eight active students. Numerous publications related to media bias have been published through the network into major conferences such as EMNLP 2021 [40], along with dataset and benchmark creations.

# Chapter 2

# Media Bias

## 2.1 Definition

Allsides [19] defines media bias as "The tendency of news media to report in a way that reinforces a viewpoint, worldview, preference, political ideology, corporate or financial interests, moral framework, or policy inclination, instead of reporting in an objective way (simply describing the facts)". This phenomenon has existed and been researched since the 1950s [49], highlighting its enduring presence and impact on public perception. Media bias can manifest in various forms, including the selection of stories, framing of issues, and choice of language.

While biases may not always be intentional, they might cause significant consequences, possibly leading to inequalities and injustices. Some news outlets tend to use catchy headlines which trick readers into clicking, known as "clickbait", which are often ambiguous or misleading. Biased information can and has been used as a way to shape and influence public opinion [2]. A survey of journalists from the United States, Great Britain, Germany, Italy, and Sweden found evidence that journalists' personal beliefs substantially influence their news decisions, expressed within the stories they choose and the statements they write [34].

Media bias can manifest within different levels, categorised into 4 major types [39], as shown in Figure 2.1.

An example of a biased article can be seen from a recent article published by The Economist in Figure 2.2. The headline includes a negative framing of certain groups, which can influence readers' perceptions before they even read the article, suggesting that the existence of these groups is inherently problematic and harmful. Terms like "incels" (involuntary celibates) and "anti-feminists" carry strong negative connotations, which can evoke emotional responses and suggest a negative view of the groups mentioned. This is a strong example of framing bias.

This project will mainly focus to address text-level context bias: phrasing bias, spin bias, and statement bias [39]: statement bias refers to "members of the media interjecting their own opinions into the text", phrasing bias is characterised by inflammatory words, i.e., non-neutral language, spin bias describes a form of bias introduced either by leaving out the necessary information.

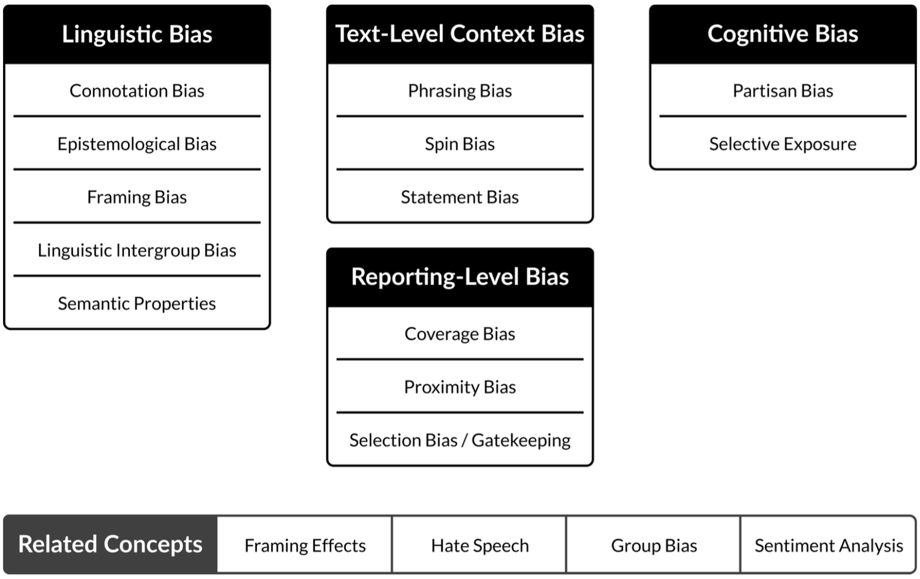
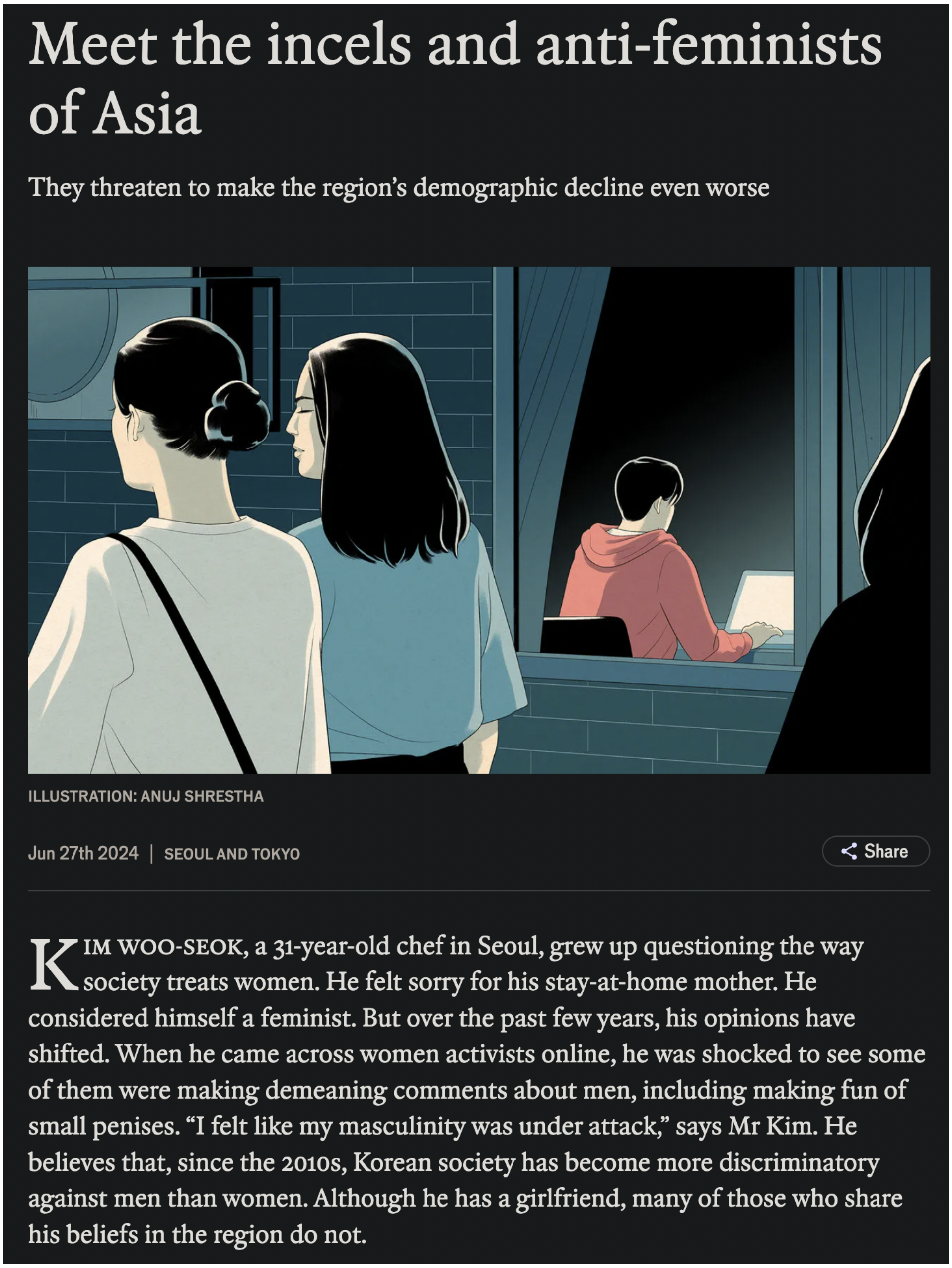



Figure 2.1: Media bias types as defined in [39]

## 2.2 Application

Another particularly dangerous example of media bias is its application in relation to elections. According to a survey done by [3], fake news propagated in social media played a pivotal role in the eventual election of President Trump during the 2016 election. Panagopoulos' study [31] revealed systemic biases leaning towards Democratic candidates during national and state levels pre-election polls conducted during the 2020 U.S. general election cycle. Rafail et al. [36] examined 201,678 media documents from Tea Party organizations, Fox News, MSNBC, and 785 newspapers, revealing significant differences in how the Tea Party frames itself compared to how other media sources frame the movement, MSNBC portrays it as the worst aspect of the Republican Party, while Fox News sees it as the best, sharply in contrasts with how activists frame the movement as conservative but not strictly Republican, often clashing with Republican Party goals. An analysis by Pew Research Center [20] found that miscalculations observed in the 2020 U.S. election polls would adjust public opinion on issues by an average of just under 1 percentage point. Although errors of such scale would not have produced substantial differences on the American public opinions, this shows the underlying bias within polls specifically and the failure of accurately representing surveys.

The phenomenon of media bias in media has not gone unnoticed, particularly by readers and consumers, lowering public trust in media outlets. In the United States particularly, trust in news media is at an all time low, falling consistently and significantly over the past 20 years [14, 13, 29]. Approximately half of Americans believe that the media is significantly responsible for the political divisions within



# Meet the incels and anti-feminists of Asia

They threaten to make the region's demographic decline even worse

ILLUSTRATION: ANUJ SHRESTHA

Jun 27th 2024 | SEOUL AND TOKYO

Share

KIM WOO-SEOK, a 31-year-old chef in Seoul, grew up questioning the way society treats women. He felt sorry for his stay-at-home mother. He considered himself a feminist. But over the past few years, his opinions have shifted. When he came across women activists online, he was shocked to see some of them were making demeaning comments about men, including making fun of small penises. "I felt like my masculinity was under attack," says Mr Kim. He believes that, since the 2010s, Korean society has become more discriminatory against men than women. Although he has a girlfriend, many of those who share his beliefs in the region do not.

FIGURE 2.2: A recent article from The Economist portraying gender discrimination with negative-framing headline

the United States, with a growing number of Americans losing faith in the media's objectivity and perceiving it as actively engaging in ideological wars [13]. Reuters Institute reported only less than half of their respondents (40%) generally trust the majority of news sources, with the US ranked on 29 out of 40 countries (32%) [29, 38], ranking far below developing countries such as Indonesia, Phillipines, and Turkey (full figure shown in Figure A.5).

Readers themselves are not exempt from bias, as they are known to prefer to pick, follow, and consume articles that align with their own beliefs and ideology, an issue known as filter bubble [18] or selective exposure [39]. This incident reinforces existing biases and limits exposure to diverse perspectives, creating echo chambers that hinder critical thinking and informed decision-making. The combination of media bias and filter bubbles can distort reality, perpetuate misinformation, and deepen societal divisions. It is essential for readers to seek out a variety of sources and viewpoints to gain a more balanced and comprehensive understanding of the issues at hand.

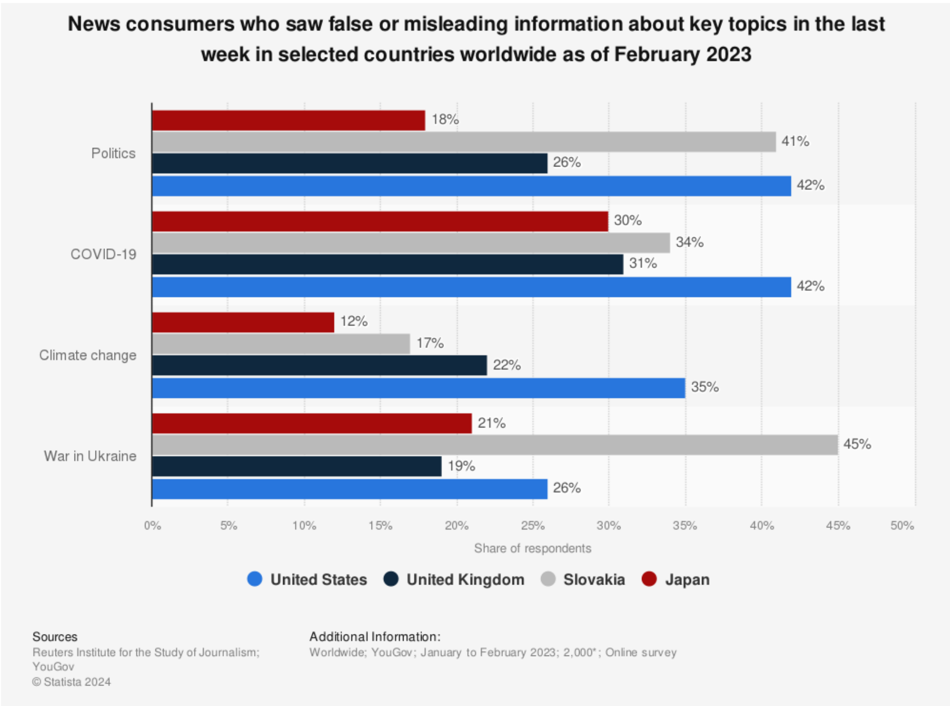



Figure 2.3: [37]

Based on surveys from Reuters and YouGov (shown in Figure 2.3 and Figure 2.4), the United States is shown to have the highest rate in politics, COVID-19, and climate change, with nearly half of the respondents claimed to have seen false or misleading information in the last week. Furthermore, most adults of every age group responded to seeing false or misleading information on a **daily** basis. These numbers are deeply concerning, highlighting the extent of the misinformation and

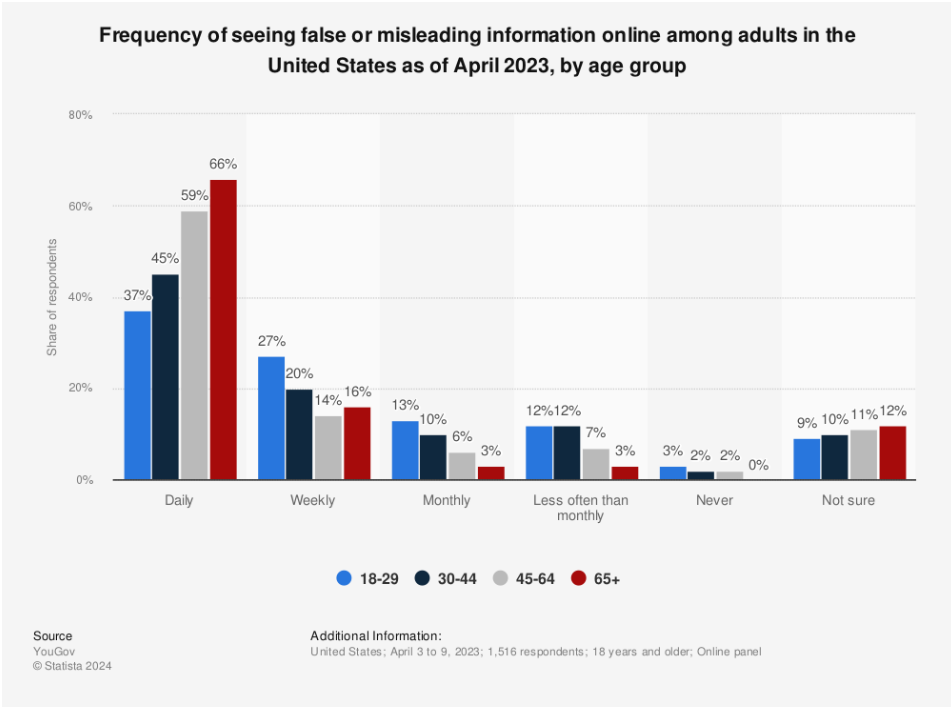



FIGURE 2.4: Rate of misleading information in the Using[51]

media bias problem in the US.

## 2.3 Goal

Ideally, unbiased media content that objectively and fairly represents multiple or a range of perspectives is desirable, news sources should remain neutral and let readers build their own opinions on the subject [30]. However, this is often unachievable due to human capabilities and resource limitations; journalists cannot possibly possess complete knowledge on every topic, be physically present everywhere, or interview every relevant individual on a significant subject [19]. Truth and journalism objectivity is a complex matter full of choices and dilemmas, where ultimately it falls on the journalists' own preferences and criteria [5].

Therefore, instead of eliminating media bias, our goal should be to draw attention to its existence, giving readers awareness of such content [39], ultimately building a tool to defend readers from media manipulation, to let readers know the quality of the article, and if they fall under a victim of political agenda or indoctrination.

Therefore, the primary objective of a media bias classifier should be to develop an automated system capable of classifying media bias at the article level. This system should be able to recognise and categorise bias in news articles from diverse sources, ensuring that readers are aware of what they are reading and therefore can make informed choices and opinions. The classification should encompass various dimensions of bias, including political orientation (e.g., left, right, center), sensationalism, and framing techniques.

# Chapter 3

# Literature Review

## 3.1 Dataset

Available datasets for article-level media bias classification are quite limited, often in different formats or covering different types of bias:

- The BASIL dataset [11] (300 articles) is the most commonly used bias dataset, containing 300 articles with both article-level and phrase-level annotations. This dataset focuses on informational bias in news articles as it appears more frequently than lexical bias. However, the obvious drawback of this dataset is the low amount of articles included as it is not nearly enough data to get a good working detection model.

- NLPCSS [7] (6964 rows), annotated via Ad Fontes labels, contains article-level textual content with three bias labels (bias, neutral, or unknown). Focusing on political bias and unfairness, this can technically be a decent dataset for work for an article-level classification task.

- The BAT [41] dataset (6345 articles, 321 outlets, Ad Fontes labels) is the most suitable for article-level bias detection as it supplies both bias-score and reliability-score for the whole article. However, the dataset itself does not supply the textual content of the articles and therefore needs to be extended, something that I have been working on as well during my Master's Thesis this and last year.

Annotating bias is not a straightforward task. Traditionally, this is done by hiring experts and journalists to manually read and determine how biased the content is. As with [40], annotations are generally compiled and majority voted to achieve the final annotation given a particular text. It is important to use multiple annotators to minimise introducing another form of bias towards the dataset. Annotators' personal background moderately influenced their decisions and should be taken into consideration when building datasets, along with other factors such as topics, reading news habits, and honest mistakes [42]. Clearly, this is not a cheap procedure and can be a huge bottleneck in creating a reliable media bias dataset.

Alternatively, websites such as Allsides and AdFontes have their own experts and annotations which can be crawled and made use of. Several papers and datasets [41, 7, 21] have used this approach. However, since these datasets depend on manual labeling from a third-party organisation, the selection of articles likely also introduces bias into the dataset [41].

## 3.2 Classification

Current State-of-the-Art in media bias detection typically employs a transformer-based approach, by fine-tuning or exploiting Large Language Models such as BERT [9] or RoBERTa [25].

MAGPIE [17] is the first large-scale, RoBERTa-based, multi-task learning (MTL) model dedicated to bias-related tasks, a promising approach for media bias detection and can be used to enhance the accuracy and efficiency of existing models. Using MAGPIE's context representation instead of BERT for media detection can potentially improve performance. However, the model is only trained for sentence-level classification and also outputs binary results.

Many past works on automatic detection of media bias typically use the BASIL dataset, operating on a sentence level and output binary result (either biased or not) [27, 28, 16, 46, 22, 24, 23], complemented by the lack of appropriate and adequate datasets with article-level annotations [8]. One of the main problems with detecting on sentence level is that multiple sentences within the same article could have opposite or different biases. As media content is often delivered in the form of articles containing a number of paragraphs, detection at the article level is far more useful and desirable.

There are currently only a few works on media bias and article level classification. Chen et al. [6] utilised a Gaussian mixture model, incorporating probability distributions of frequency, positions, and sequential order of lexical and informational sentence-level bias to detect article-level bias. Kulkarni [21] utilised attention mechanism to model a network structure of an article to classify political ideology. Outside of media bias, Su et al. [43] used chunking methods combined with encoder-based '[CLS] Pooling' to extract representations on the document-level.

Furthermore, The traditional approach of assigning only a single frame label to news articles remains overly simplistic, given that a standard news story often incorporates multiple viewpoints, arguments, or facets, each potentially carrying distinct connotations or framing [45]. An integer score label would be a slightly better solution to represent bias from a text, although it is still hardly ideal.

Additionally, it would be good to consider the explainability of a model when classifying a bias, as we would need to not just get the result, but to understand why contents are considered bias. Existing media bias detection systems typically concentrate solely on predicting the likelihood of a certain text being biased, offering limited insights into the underlying reason behind the decision.

## 3.3 Working with article-length text

Article text length generally falls between medium to long sequences, as they are not as lengthy as legal documents or clinical studies that usually contain multiple pages of text. Most news articles stay between several hundred to several thousand words. There are several additional challenges when working with longer sequences as classification models that demonstrated significant results have been shown to perform poorly or even fail when they are tested on large documents [48].

The most straightforward approach of standard fine-tuning of an encoder-based language model is not necessarily effective for article-level processing as they are mostly only able to process a maximum of 512 tokens (BERT), prompting significant information loss. Several techniques have been attempted to address this issue such as the sliding window techniques, CLS techniques [43], where you exploit the 'CLS' token from a BERT embedding, and other more sophisticated models [21].

Additionally, there are other LLMs similar to BERT that are designed to handle longer sequences: Longformer [4], BigBird [53]. However, these models have a different architecture than the original BERT, often do not consistently surpass the baseline models in classification performance, and only performed notably better than the baseline models on only two datasets [32].

# Chapter 4

# Dataset Building

## 4.1 BAT dataset

### 4.1.1 Characteristics

The BAT dataset [41] is chosen instead of NLPCSS [7] for this project due to its article-level suitability and labels fluidity, as well additional metadata as it contains outlets information. It contains 6345 rows of manually labeled news articles from 255 English-speaking news outlets (US-based), originally scraped from Ad Fontes Media's website along with their respective **political bias** and **reliability scores**. Articles in the dataset encompassed a wide range of topics such as COVID-19, politics, and lifestyle. The political bias score measures the extent of political influence, ranging from -42 (most extreme left) to +42 (most extreme right). The reliability score reflects the article's truthfulness, with values ranging from 0 (least reliable, containing inaccurate or fabricated information) to 64 (most reliable, original fact reporting).

Both political bias and reliability scores on each article were rated using defined metrics and multiple sub-factors, performed by three randomly selected analysts from Ad Fontes Media's team of over 60 experts. The corresponding three scores were then averaged, producing the final article scores. Moreover, each group consists of analysts with different beliefs in the political spectrum i.e., left, center, and right.

The reliability score evaluates original fact reporting to analysis, opinion, propaganda, and inaccurate/fabricated information, with scores above 40 generally considered good and scores below 24 typically seen as problematic, scores between 24 and 40 suggest a variety of factors, including a strong presence of opinion and analysis or significant variability in reliability across different articles [12]. This metric is chosen as the main label in this project due to its correlation with textual-level bias: phrasing bias, spin bias, and statement bias described in Section 2.1

An example of a low-rated article can be seen in Table 4.1. The deceptive article contains many wrongful claims and blatantly false events that did not happen in real life. In contrast, Table 4.2 shows an example of a high-rated article. The content reports only facts regarding the event and statements from people related to the incident. Journalist opinions or political innuendos are non-existent.

Trump Win Validated by Quantum Blockchain System Recount of Votes

A recount of voting ballots nationwide was being done by elite units of the National Guard by early Sun. morning 8 Nov. To prevent fraud official ballots had been printed with an invisible, unbreakable code watermark and registered on a Quantum Blockchain System. As of this writing, in five states 14 million ballots had been put through a laser scanner – 78% of which failed because there was no watermark to verify the ballot. Of those that failed 100% had checked for Biden. An initial test showed that according to water marks on validated ballots fed into the Quantum Computer, Trump won re-election by over 80% of the legal ballot cast. The final validated vote tallied in that test: Trump 73.5 million votes to Biden's 25.9 million – and that didn't even account for Trump votes that people observed being tossed and never accounted for. Interesting enough, those figures corresponded with the two men's Twitter accounts: Trump had 88.8 million followers to Biden's 16.6 million. Using 'infrared' equipment that read which ballots were real, or fake the elite National Guardsmen had been deployed to the twelve targeted states of Alabama, Arizona, Pennsylvania, Colorado, Texas, Wisconsin, Tennessee, Washington, Virginia, Delaware, Illinois and Kentucky. In all nationwide, over 500 National Guardsmen were on guard over all ballot counting units. There was much more to the tests for fraudulent voting. In addition to the watermark these official ballots also contained ink made of corn, which created an electronic radiation circuit ID that could trace the location of that ballot through GPS transmission. In other words, they could trace if the ballot was filled out by the person named on the ballot. The Trump team would be filing a number of lawsuits onThey had been preparing for this for a long time under an election fraud investigation called Project Veritas. Judicial Watch:"Our new study shows 1.8M excess, or 'ghost' voters in 353 counties across 29 states. The data highlights the recklessness of mailing blindly ballots/ballot applications to voter registration lists,"@TomFitton Watch more: at http://judicialwatch.org Pennsylvania alone Trump's legal counsel Rudy Guliani had testimony of 50-60 poll watchers who claimed being deprived of an ability to inspect mail in ballots. Nationally, noted attorney Sydney Powell (rumored to be appointed the next FBI director) said, "Hammer and Scorecard – the NSA Security Software turned illegal Election Software – ran an algorithm that gave Biden a 3% vote advantage in Wisconsin, Michigan, Pennsylvania, Georgia, Nevada and Arizona."Rest assured, all legal issues would be accounted for by the time the Electoral College met on. By then real election results – post court battles – would determine all legally cast ballots. The joint session of Congress would make the election official on 3 Jan. 2021.

Table 4.1: Example of a biased article, reliability score: 4.67

Trenton police officer takes own life in Plainsboro parking lot, officials say

A veteran Trenton police officer took his own life in a parking lot Wednesday, officials said. Sgt. Daniel Pagnotta, a 21-year-veteran of the department, died this morning in Plainsboro, according to a city spokesman."Beloved by everyone in the Trenton Police Department, he was devoted to Trenton and police work," Mayor Reed Gusciora said in a statement. The statement described Pagnotta as a devoted husband and father of two who loved soccer and making people laugh. His father, also named Dan, is a retired Trenton police officer."Dan was proud to continue a legacy of law enforcement in his family," Gusciora said. "Dan and his family are on our minds and in our hearts. He will be dearly missed."

Table 4.2: Example of a non-biased article, reliability score: 57.67

### 4.1.2 Extension

The original BAT dataset only contains news titles and links (along with other metadata) and is missing the body content of articles. To overcome this, a Python script is written and executed, iteratively visiting each of the URLs from the dataset and scraping the news content. This was not an easy task as each website has its own unique structures and formats. Furthermore, the scraped text contains noises that are almost impossible to remove through the script. Some outlets such as The Nation, Chicago Tribune, and Truthout required manual intervention as the scraped text was duplicated over themselves. The first round of Extension resulted in 5270 rows of articles out of the original 6345 rows, mainly due to unavailable websites and missing articles.

To remove noises from article content, the text are then pre-processed extensively. All the content of every article in the dataset were joined into one single list, split into words, and then matched with an English word list [10], outputting a list of words that do not exist in the dictionary, sorted by their occurences. Using this list, noise patterns were analysed and handled through a combination of string and regex methods, frequent noisy phrases and conjoined words were identified and fixed through giant JSON dictionaries as references. This process is repeated more than several times until the contents are valuable enough to work with. Note that at this point some noises still remain within the text as it will take an extensive amount of time and manual labour to completely clean the text. Figure 4.1 shows examples of conjoined words, which are highly prevalent inside scraped article contents. Figure 4.2 shows part of giant JSON files that are used to address these noises.

```
companiessuch,1
JuneteenthCopyright,1
chambersdecrying,1
billthenext,1
similarmovesto,1
itan,1
kyrstensinema,1
aholiday,1
holidayas,1
Republicanssuch,1
```

FIGURE 4.1: Example of conjoined words, over 60k of conjoined are expected to exist within the dataset article contents

An additional round of extension is done after the initial scraping by compiling the list of unavailable/unscraped websites and taking a closer look of the problem individually. This round involved searching for missing articles, fixing broken links, visiting the websites manually on the browser, and copy-pasting article contents into

```
"By submitting your email, you agree to ourTermsandPrivacy Noticeand to receive email correspondence from us.": "",
"Andrew WilliamsMatt BurgessScharon Harding, Ars TechnicaMedea Giordano": "",
"Rahmat Gul/APhide caption": ". ",
"Get the latest updates from the 2024 campaign trail, exclusive interviews and more Fox News politics content.": "",
"SubscribedYou've successfully subscribed to this newsletter!": "",
"AdvertisementSupported by": "",
"Michelle GoldbergByMichelle GoldbergOpinion Columnist": "",
"Paul KrugmanByPaul KrugmanOpinion Columnist": "",
"BuzzFeed News ReporterReporting From": "",
"ByVeronica CristinoByHannah Coates": "",
"This content can also be viewed on the site it originates from.": "",
```

```
"astatement": "a statement",
"whenshe": "when she",
"issomething ": "is something",
"andsocial ": "and social",
"Thestate": "The state",
"itoriginatesfrom": "it originates from",
"theNew": "the New",
"theWashington": "the Washington",
"theNational": "the National",
"PresidentDonald": "President Donald",
"PresidentJoe": "President Joe",
```

Figure 4.2: Dictionaries to fix conjoined words and noise phrases

a spreadsheet, resulting in additional 226 rows of articles. The final dataset consists of 5496 rows of articles.

### 4.1.3 Analysis

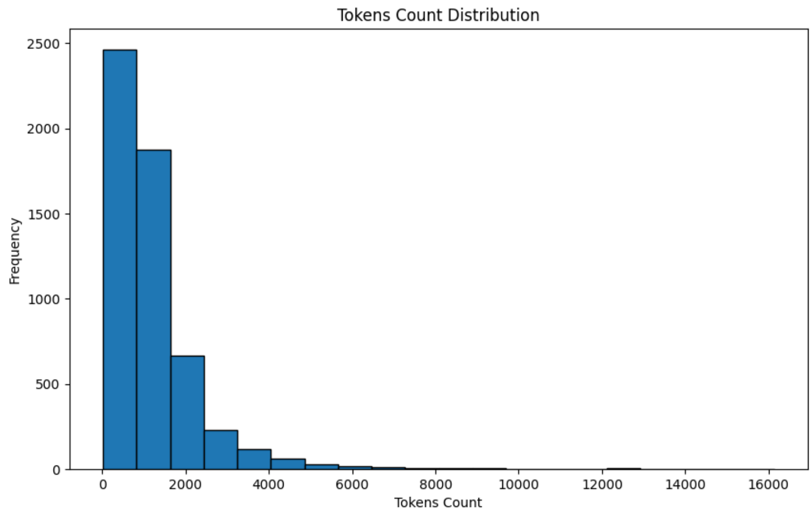



Figure 4.3: Articles token count distribution

The article content tokens length ranges between 17 tokens to 16139 tokens, with an average length of 1207.07 and a median value of 908 tokens. Only 9 articles have more than 10000 tokens, while there are 106 of articles with less than 100 tokens. Furthermore, only 1209 articles stay between 512 tokens, which is the limit for BERT input. The articles reliability score ranges from 1.0 to 58.67, the majority have a value between 20 - 50. Not a single articles were rated more than 60 despite the highest score being 64. Visualisations can be seen in both Figure 4.3 and Figure 4.4, as well as Figure A.1

Most articles are written and published within the last 6 years, with only 31 articles, a minuscule percentage, published before 2019, shown in Figure 4.5. From a personal analysis, these 31 articles generally contain similar topics to articles published

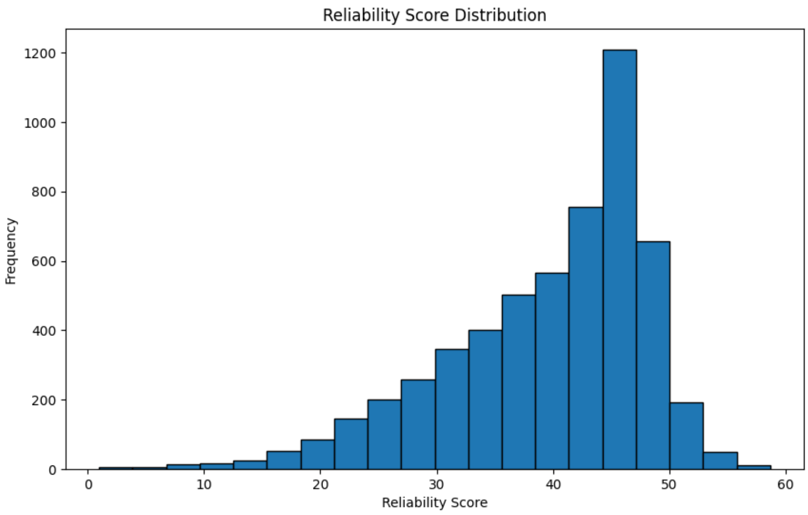



FIGURE 4.4: Reliability score distribution

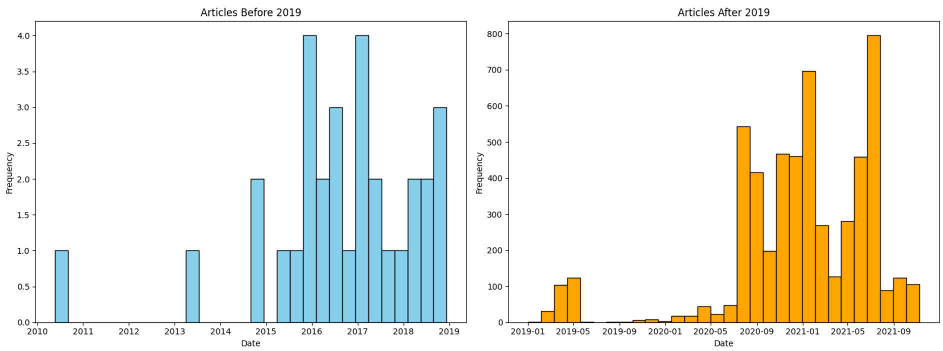



FIGURE 4.5: Article dates distribution

after 2019 and therefore should not hold consequental difference in behaviour and characteristics.

Figure A.3 shows that all classes seem to have similar token count, close to the overall average. Class 'Problematic' and 'Questionable', being the two most biased classes, seem to have lower average token count than the two other classes. However, further analysis (Figure 4.6) shows that there is virtually no linear relationship between token count and reliability score, with a Pearson correlation coefficient is 0.02. This proves that the length of an article has no significant impact on its reliability score. In other words, longer articles are not necessarily more or less reliable than shorter ones based on the provided data.

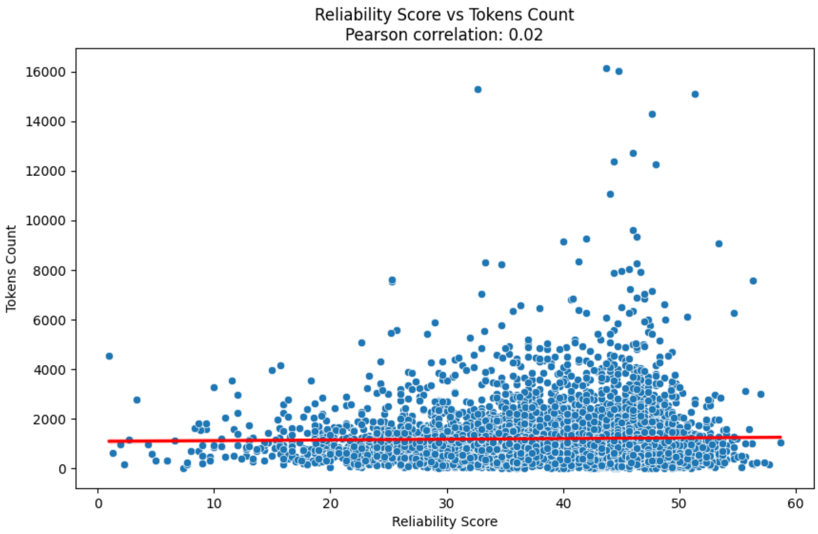



Figure 4.6: Pearson correlation between token count and reliability score

# Chapter 5

# Methodology

## 5.1 Pre-processing

### 5.1.1 Dataset split

The dataset reliability scores are grouped and split into 4 classes based on Ad Fontes split as previously described in Section 4.1.1:

1. Problematic —> scores between 0.00 and 24.00
2. Questionable —> scores between 24.01 and 32.00
3. Generally Reliable —> scores between 32.01 and 40.00
4. Reliable —> scores between 40.01 and 64

The dataset is then split into three sets of train, test, and validation with the following distribution:

- Train set: 4325 rows
  287 samples of class ’Problematic’, 611 samples of class ’Questionable’, 1033 samples of class ’Generally Reliable’, 2394 of class ’Reliable’
- Test set: 569 rows
  27 samples of class ’Problematic’, 54 samples of class ’Questionable’, 104 samples of class ’Generally Reliable’, 384 of class ’Reliable’
- Validation set: 603 rows
  34 samples of class ’Problematic’, 70 samples of class ’Questionable’, 128 samples of class ’Generally Reliable’, 371 of class ’Reliable’

The split is done in a way to ensure that articles from different outlets are distributed equally between the three sets. This is done by first grouping the articles based on their outlet and labels, then iterating over each group, splitting the rows equally, and distributing to the train, test, and validation set. Groups of less than 5 rows that are not enough to be split and therefore appended to the train set. To handle class imbalances, weighted loss is used when training the model, with weights in proportion to the distribution of each class.

A major drawback of this 'balance' splitting is that there is no unseen outlet in the test set and validation set. This can influence the final test metrics and may hinder the model's ability to generalise to new, unseen articles from unseen outlets. However, considering that new outlets are rarely introduced in the real life, it might be beneficial to slightly overfit on the patterns of existing outlets.

## 5.2 Features and baselines

The primary features will include the title and content of the articles. Ideally, a reliable article-level bias classifier should be able to generalise solely or mainly from the content of the articles, capturing the context of the article will be the key element of reliable performance. However, outlet information is also experimented and compared.

As baselines, traditional encoding methods such as Bag-of-Words and TF-IDF are implemented, combined with a simple logistic regression as a classifier. In these cases, the validation set are concatenated into the training set. An outlet majority votes is also implemented as a comparison to show the influence of outlet information in this particular task. This method works by simply taking the majority vote over classes for every outlet and use it as a classifier: *an article is from outlet A, majority of articles from outlet A is classified as class X, therefore, article A has a class X.*

## 5.3 Proposed methods

### 5.3.1 BoW + MLP

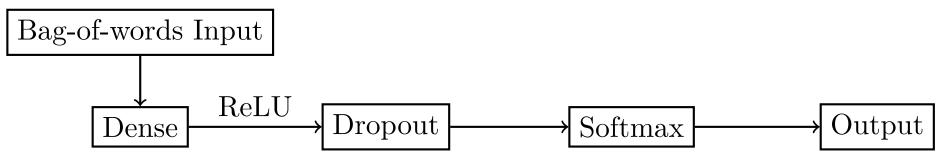



Figure 5.1: BoW + MLP architecture. The input article is encoded as a bag of words, then feed into a single linear multilayer perceptron before softmax operation.

Figure 5.1 shows the simple diagram of the BoW + MLP method, combining traditional encoding methods with a simple neural network training. This approach allows for a simple and computationally-cheap implementation, yet effective classifier. The model is trained on 10 epochs with a learning rate of 2e-5. Hidden size for the dense layer is 128 with a dropout probability of 0.2, no warmup steps are applied.

### 5.3.2 BERT fine-tuning

### 5.3.3 BERT sliding window fine-tuning

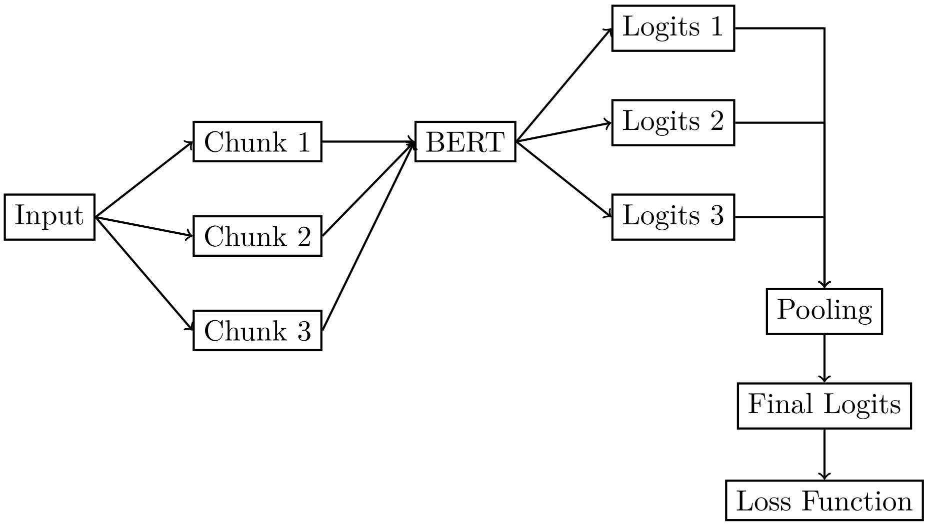



FIGURE 5.2: BERT Sliding window fine-tuning architecture. The input article is split into chunks, each chunk is processed by the model as a mini batch, and the resulting logits are pooled before applying the loss function.

Figure 5.2 illustrates the architecture of the sliding window fine-tuning method. Input texts are segmented into chunks, which are then processed as mini-batches by the model. The logits (output scores) produced by each chunk are then pooled together to produce the final logits, to which the loss function is applied. This method bypasses the sequence length limitation of BERT models (512 tokens), allowing for full representation and processing of text input without any loss of information.

A window size of 512 is used in the implementation with no stride, fine-tuned on 4 epochs with weighted loss.

### 5.3.4 CLS method

Figure 5.3 shows the full architecture of the CLS method. This method begins by similarly segmenting the input text into smaller chunks. These chunks are encoded into a higher dimensional space features by feeding them into a pre-trained Large Language Model (LLM) and extracting the last hidden state, as described in [44]. Subsequently, the chunks are then passed into $n$ transformer layers to enrich their contextual understanding. For each chunk, only the representation of the CLS token (the first token) is retained (CLS Pooling, as in [43]), serving as a concise summary of the entire chunk sequence. This summary representation is then processed through a Multi-Layer Perceptron (MLP). Finally, a softmax operation will be applied to the output of the MLP layer to determine the final output class. Using a LLM as an encoder and taking its last hidden state allows for a good contextual representation

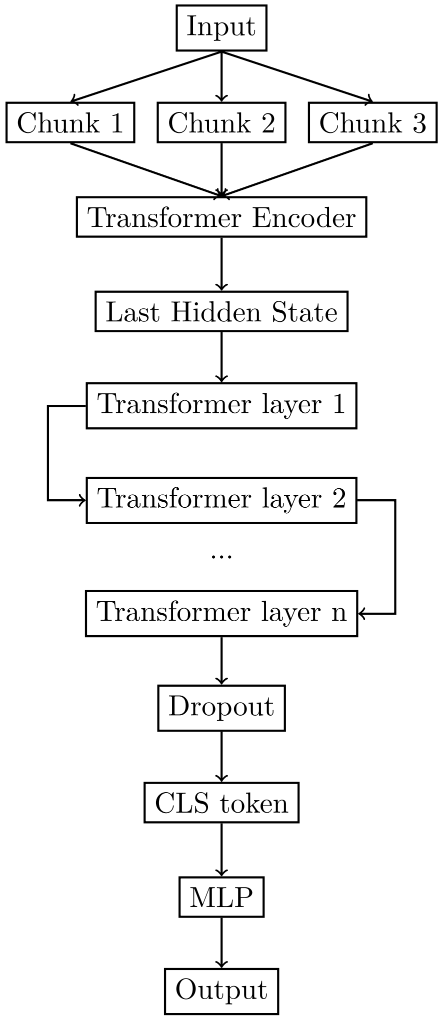



Figure 5.3: CLS method architecture

of each chunk. By only using the CLS token representation instead of the whole chunk, this method allows for a more effective, yet simpler approach compared to other chunk pooling methods.

Two different language models can be used as the transformer encoder in this method: BERT [9] or MAGPIE [17]. The MLP consists of two linear layers with ReLu activation function [1] and dropouts. Each chunk is set to contain 512 tokens, only 2 transformer layers are used, and 0.2 probability is applied to the dropout layer. Training is done over 3 epochs with weighted loss.

## 5.4 Implementation details

All methods are implemented in Python 3.12.0 [47] using the PyTorch [33] and transformer [50] package from HuggingFace. The BERT model utilised in these methods is the 'bert-base-cased' instead of the 'bert-base-uncased' to account for distinctions in capitalised words, which may be crucial [9]. AdamW [26] optimiser and Cross Entropy Loss is used on all cases.

### 5.4.1 Weighted loss

Weighted loss addresses the problem of training models on an imbalanced dataset by assigning higher weights to classes with fewer instances and lower weights to classes with more instances. This adjustment ensures that the model pays more attention to correctly predicting the minority class, thereby improving overall performance metrics. This method is applied through the scikit-learn library [35].

# Chapter 6

# Evaluation

For every method, precision, recall, and F1 score is evaluated on both overall and per class performance.

Table 6.1 shows the evaluation of all methods given only title and content of articles as features. It can be seen that frequency-based approaches such as BoW and TF-IDF performed generally decent. However, looking at per-class metrics, it is evident that the overall scores are heavily influenced by the performance of the 'Reliable' class due to its large support. On all of the other less represented classes, the models' performance suffered with much lower F1 scores with the TF-IDF + LR model having the worst F1 score on the 'Problematic' class. The BoW + MLP model achieved slight improvement in almost all instances compared to its predecessor due to the use of neural network. A typical BERT fine-tuning outperformed all frequency-based methods, though only with a close margin compared to the BoW + MLP model. The BERT sliding window fine-tuning (BERT SW FT) method outperformed all BoW and TF-IDF methods, also performing slightly better to a standard BERT fine-tuning of the first 512 tokens. Similarly, the CLS methods mostly outperformed traditional methods with best F1 score on the 'Problematic' class, able to generalise on the most biased class. While BERT CLS achieved slightly worse performance compared to the two fine-tuning methods, MAGPIE CLS performed best among all other methods with the highest overall precision, recall, and F1 score.

Table 6.2 shows the evaluation of methods, given the article's outlet information as additional features to the title and content. As a comparison, **using outlet information as the sole feature without any other information (with majority votes) outperformed all methods on overall recall and f1 scores**. This result signified how influential outlet information can be in terms of classifying media bias, with regards to this specific dataset and circumstances. Out of all the other methods, BERT fine-tuning reigned superior with the best overall precision score and a particularly strong f1 score on the 'Problematic' class, performing slightly better than the sliding window implementation (BERT SW FT). Interestingly, both BERT CLS and MAGPIE CLS methods were not able to capitalise on the outlet information as well as BERT FT and BERT SW FT as both CLS methods only saw little improvement compared to when outlet features were unavailable.

| Method | Class | Precision | Recall | F1 |
|---|---|---|---|---|
| BoW + LR | Problematic | 0.38 | 0.41 | 0.39 |
| | Questionable | 0.34 | 0.31 | 0.33 |
| | Generally Reliable | 0.36 | 0.40 | 0.38 |
| | Reliable | 0.85 | 0.83 | 0.84 |
| | Overall | 0.6922 | 0.6818 | 0.6865 |
| TF-IDF + LR | Problematic | 1.00 | 0.11 | 0.20 |
| | Questionable | 0.47 | 0.17 | 0.25 |
| | Generally Reliable | 0.36 | 0.31 | 0.33 |
| | Reliable | 0.79 | 0.94 | 0.86 |
| | Overall | 0.6904 | 0.7117 | 0.6725 |
| BoW + MLP | Problematic | 0.38 | 0.44 | 0.41 |
| | Questionable | 0.33 | 0.35 | 0.34 |
| | Generally Reliable | 0.41 | 0.42 | 0.42 |
| | Reliable | 0.88 | 0.85 | 0.86 |
| | Overall | 0.7162 | 0.7065 | 0.7110 |
| BERT FT | Problematic | 0.43 | 0.44 | 0.44 |
| | Questionable | 0.29 | 0.39 | 0.33 |
| | Generally Reliable | 0.44 | 0.48 | 0.46 |
| | Reliable | 0.90 | 0.83 | 0.86 |
| | Overall | 0.7359 | 0.7065 | 0.7193 |
| BERT SW FT | Problematic | 0.45 | 0.48 | 0.46 |
| | Questionable | 0.39 | 0.52 | 0.45 |
| | Generally Reliable | 0.42 | 0.50 | 0.46 |
| | Reliable | 0.91 | 0.81 | 0.86 |
| | Overall | 0.7472 | 0.7112 | 0.7257 |
| BERT CLS | Problematic | 0.44 | 0.67 | 0.53 |
| | Questionable | 0.41 | 0.48 | 0.44 |
| | Generally Reliable | 0.39 | 0.50 | 0.44 |
| | Reliable | 0.91 | 0.79 | 0.84 |
| | Overall | 0.7440 | 0.6994 | 0.7163 |
| MAGPIE CLS | Problematic | 0.46 | 0.63 | 0.53 |
| | Questionable | 0.36 | 0.41 | 0.38 |
| | Generally Reliable | 0.41 | 0.55 | 0.47 |
| | Reliable | 0.93 | 0.80 | 0.86 |
| | Overall | **0.7577** | **0.7117** | **0.7293** |

Table 6.1: Evaluation table, with features only including title and content of articles

| Method | Class | Precision | Recall | F1 |
|---|---|---|---|---|
| Outlet majority | Problematic | 0.56 | 0.70 | 0.62 |
| | Questionable | 0.58 | 0.46 | 0.52 |
| | Generally Reliable | 0.56 | 0.53 | 0.54 |
| | Reliable | 0.91 | 0.93 | 0.92 |
| | Overall | 0.7945 | **0.7996** | **0.7959** |
| BERT FT | Problematic | 0.65 | 0.56 | 0.60 |
| | Questionable | 0.45 | 0.50 | 0.47 |
| | Generally Reliable | 0.44 | 0.62 | 0.52 |
| | Reliable | 0.94 | 0.84 | 0.88 |
| | Overall | **0.8186** | 0.7883 | 0.7504 |
| BERT SW FT | Problematic | 0.46 | 0.41 | 0.43 |
| | Questionable | 0.41 | 0.48 | 0.44 |
| | Generally Reliable | 0.46 | 0.56 | 0.51 |
| | Reliable | 0.91 | 0.84 | 0.87 |
| | Overall | 0.7585 | 0.7359 | 0.7451 |
| BERT CLS | Problematic | 0.37 | 0.59 | 0.46 |
| | Questionable | 0.32 | 0.43 | 0.37 |
| | Generally Reliable | 0.40 | 0.47 | 0.43 |
| | Reliable | 0.92 | 0.79 | 0.85 |
| | Overall | 0.7384 | 0.6871 | 0.7071 |
| MAGPIE CLS | Problematic | 0.42 | 0.52 | 0.47 |
| | Questionable | 0.35 | 0.43 | 0.38 |
| | Generally Reliable | 0.43 | 0.55 | 0.48 |
| | Reliable | 0.93 | 0.82 | 0.87 |
| | Overall | 0.76034 | 0.7170 | 0.7342 |

Table 6.2: Evaluation table, with outlet information as additional features (outlet + title + content)

# Chapter 7

# Conclusion

### 7.0.1 Remarks

Simpler, frequency-based methods such as BoW and TF-IDF with logistic regression already provides decent overall result. Transfomer-based approach

Given that the dataset imbalance is solved, it is assumed that BERT and MAGPIE methods should

While we can seemingly just take the outlet and use this information to classify media bias...

Overall, including outlet information increase the performance in all methods, some more significant than others.

Generally, I argue that it is better to have a model that generalise well over article content rather than relying on outlet and metadata information to classify media bias in articles.

Furthermore, we have not tested the model with unseen data containing unseen outlets...

The whole dataset is build by Ad Fontes, articles are picked by Ad Fontes, annotations are done by people from Ad Fontes, therefore there is an underlying bias that exist original to the dataset, affecting models trained with this dataset.

It is really difficult to compare this result with other works on media bias classification as there is no comparable dataset particularly for article-level media bias classification.

### 7.0.2 Future Work

The methods need to be reevaluated with a bigger dataset and particularly much more examples of biased articles, as the current dataset is highly-imbalanced.

Classifying media bias right now, in this way, does not provide much explanaibility. Ultimately, we would want a model to also output a reasoning behind the classification.

Implement a more global methods using graph-based approach to encode multiple articles and define relationships between them. This would allow for a bigger and more detailed representation of articles, allowing the model to generalise according to combined features of the corresponding articles.

Carefully tune the test set to enable more representational articles and pattern, include unseen outlets.

# Appendices

# Appendix A

# The First Appendix

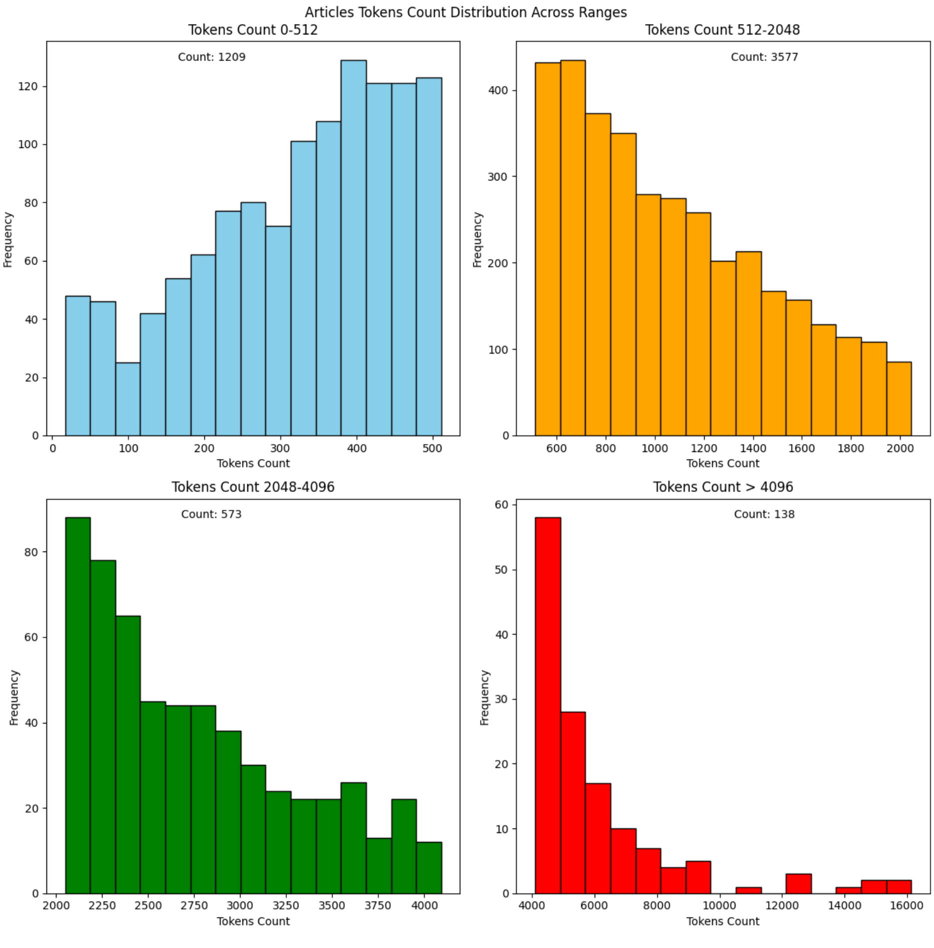



FIGURE A.1: Articles tokens count distribution

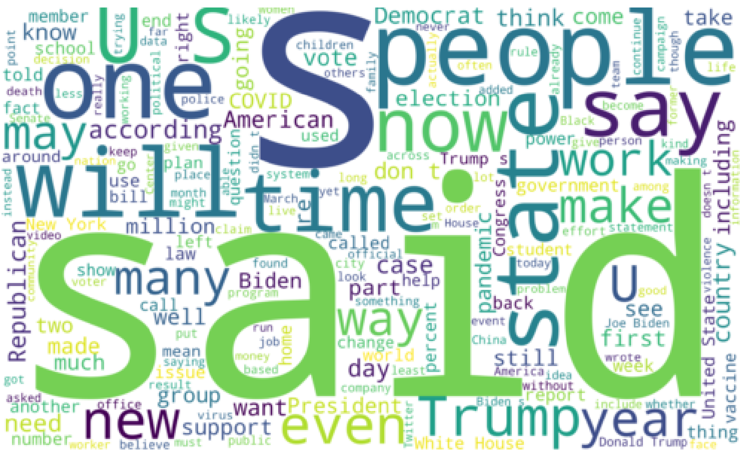



Figure A.2: Wordcloud

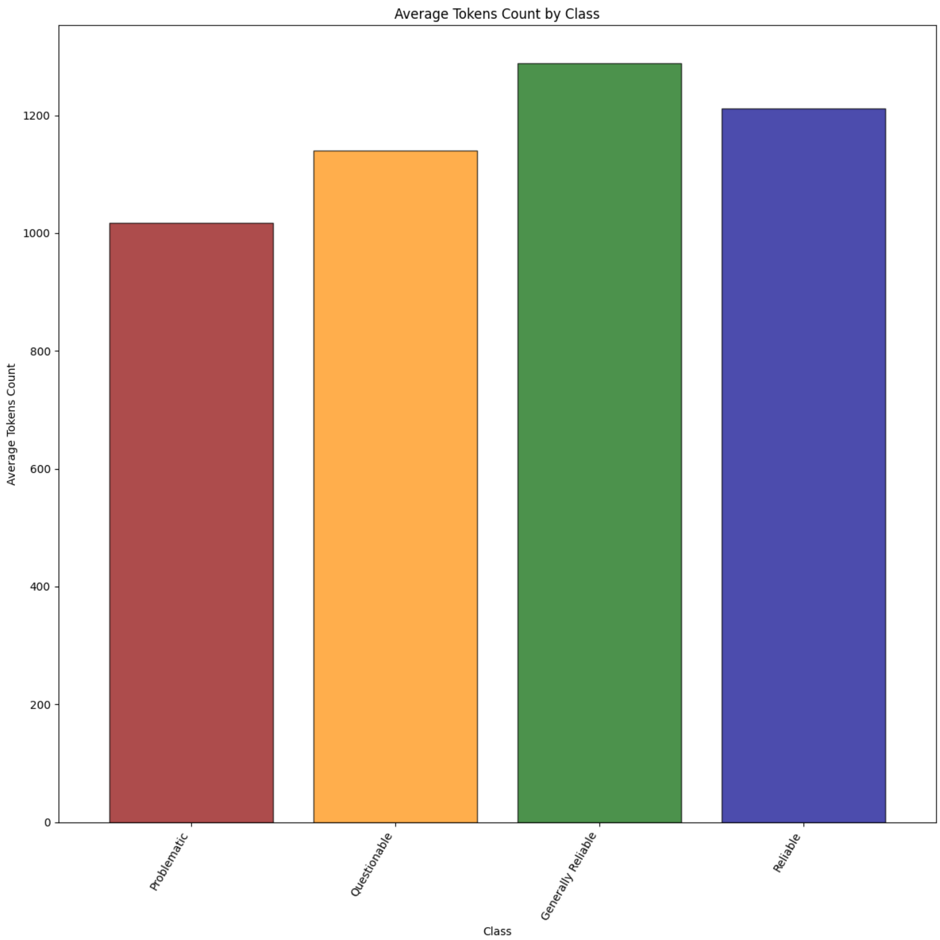



Figure A.3: Average tokens count per class

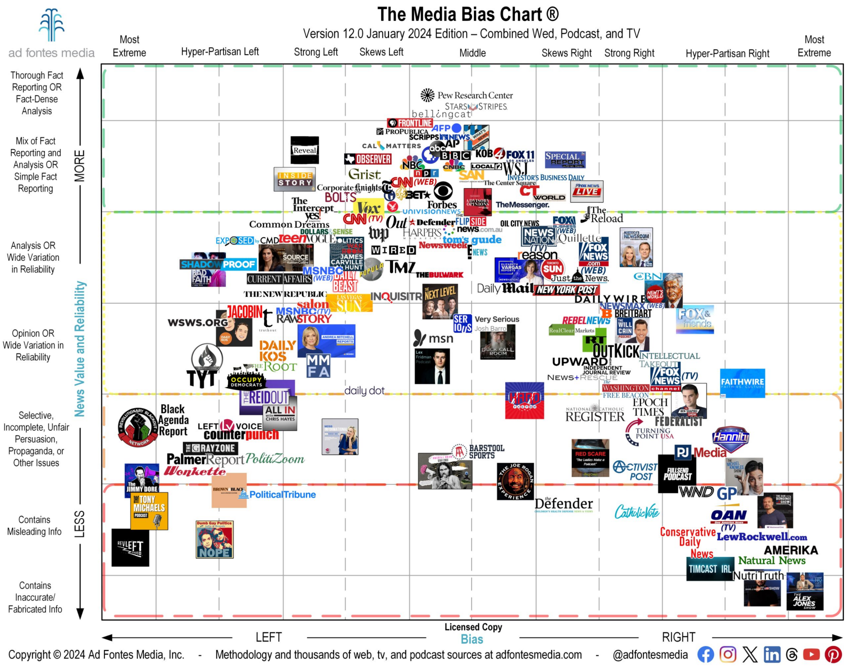



Figure A.4: Ad Fontes media bias chart

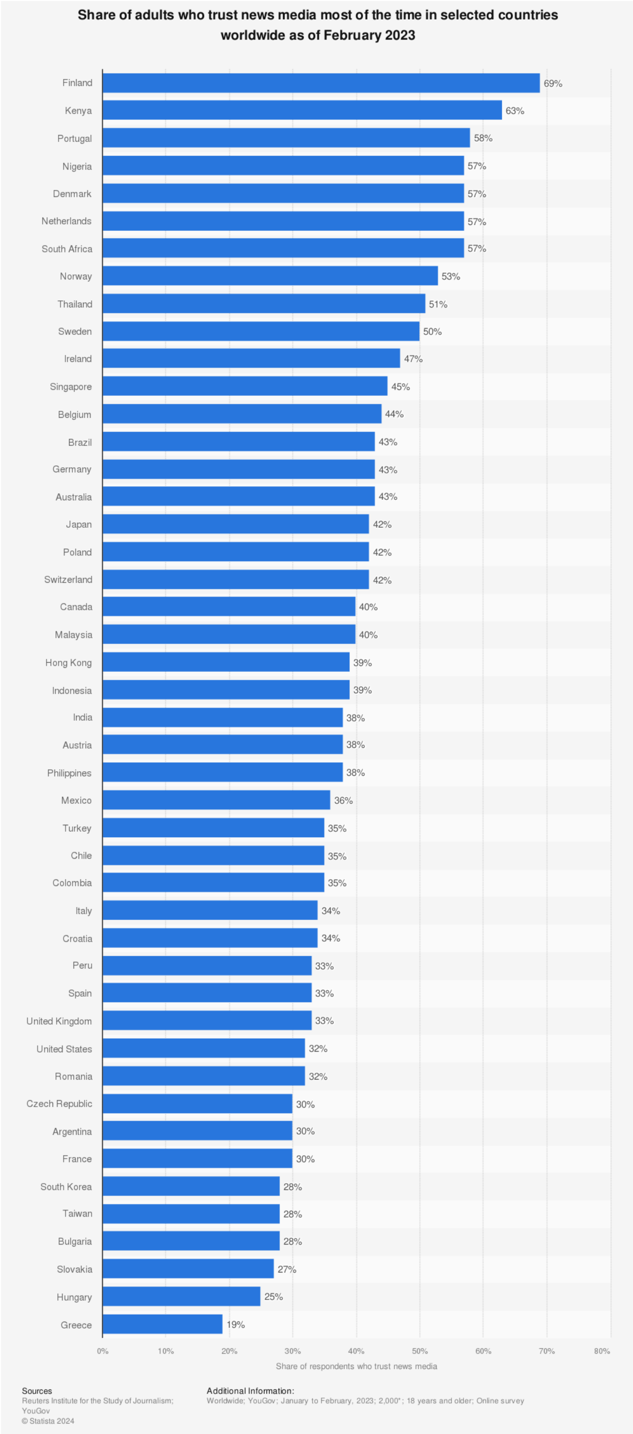



FIGURE A.5: Trustworthiness of news media worldwide, as of February 2023 [38]

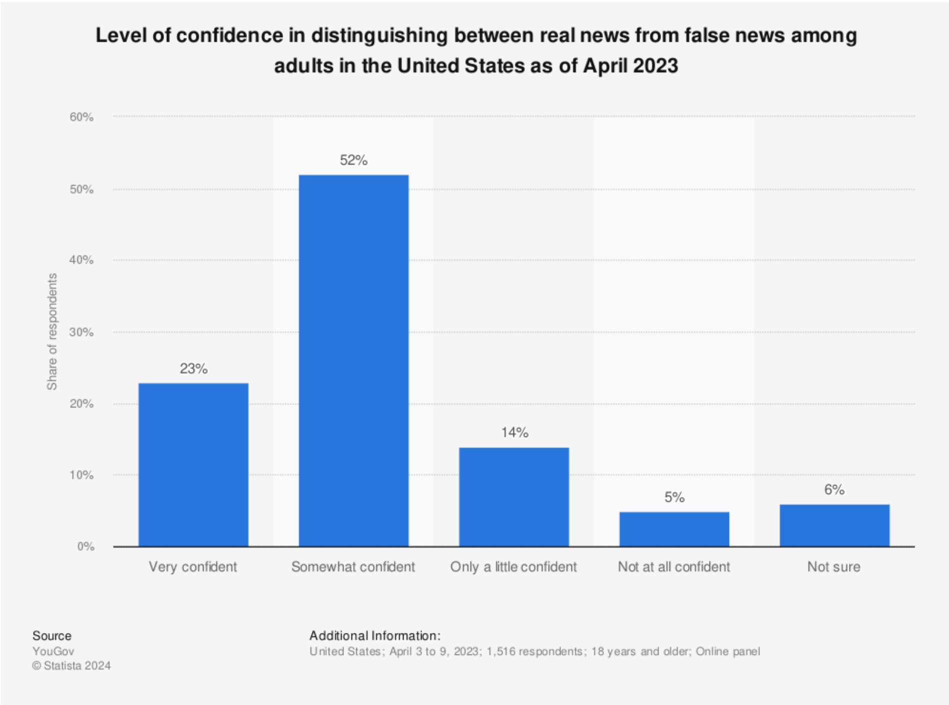



Figure A.6: Ability to recognise false information in the US [52]

# Appendix B

# The Last Appendix

# Bibliography


[1] A. F. Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018.

[2] V. P. Aires, J. Freire, F. G. Nakamura, A. S. da Silva, and E. F. Nakamura. An information theory approach to detect media bias in news websites. In *WISDOM 2020*, 2020.

[3] H. Allcott and M. Gentzkow. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–36, 2017.

[4] I. Beltagy, M. E. Peters, and A. Cohan. Longformer: The long-document transformer, 2020.

[5] S. Boudana. A definition of journalistic objectivity as a performance. *Media, Culture & Society*, 33(3):385–398, 2011.

[6] W.-F. Chen, K. Al Khatib, B. Stein, and H. Wachsmuth. Detecting media bias in news articles using Gaussian bias distributions. In T. Cohn, Y. He, and Y. Liu, editors, *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4290–4300, Online, Nov. 2020. Association for Computational Linguistics.

[7] W.-F. Chen, K. Al Khatib, H. Wachsmuth, and B. Stein. Analyzing political bias and unfairness in news articles at different levels of granularity. In D. Bamman, D. Hovy, D. Jurgens, B. O'Connor, and S. Volkova, editors, *Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science*, pages 149–154, Online, Nov. 2020. Association for Computational Linguistics.

[8] D. Demidov. Political bias of news content: Classification based on individual articles and media, 03 2023.

[9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, and T. Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages

4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[10] dwyl. english-words. `https://github.com/dwyl/english-words`, 2022.

[11] L. Fan, M. White, E. Sharma, R. Su, P. K. Choubey, R. Huang, and L. Wang. In plain sight: Media bias through the lens of factual reporting. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2019.

[12] A. Fontes. Home - ad fontes media. `https://adfontesmedia.com//`. Accessed: 2024-05-31.

[13] K. Foundation. American views 2020: Trust, media and democracy. Technical report, Gallup and Knight Foundation, 2020.

[14] J. Gottfried and J. Liedke. Partisan divides in media trust widen, driven by a decline among republicans. Technical report, Pew Research Center, 2021. [Online; accessed 29-April-2024].

[15] M. B. Group. Home - media bias group. `https://media-bias-research.org/`. Accessed: 2024-05-29.

[16] S. Guo and K. Q. Zhu. Modeling multi-level context for informational bias detection by contrastive learning and sentential graph network, 2022.

[17] T. Horych, M. Wessel, J. P. Wahle, T. Ruas, J. Waßmuth, A. Greiner-Petter, A. Aizawa, B. Gipp, and T. Spinde. Magpie: Multi-task media-bias analysis generalization for pre-trained identification of expressions, 2024.

[18] S. jeong Lim, A. Jatowt, and M. Yoshikawa. Understanding characteristics of biased sentences in news articles. In *CIKM Workshops*, 2018.

[19] Julie Mastrine. What is media bias? `https://www.allsides.com/blog/what-media-bias`, 2022. [Online; accessed 28-April-2024].

[20] S. Keeter, N. Hatley, A. Lau, and C. Kennedy. What 2020's election poll errors tell us about the accuracy of issue polling. Technical report, Pew Research Center, 2021. [Online; accessed 30-April-2024].

[21] V. Kulkarni, J. Ye, S. Skiena, and W. Y. Wang. Multi-view models for political ideology detection of news articles. In E. Riloff, D. Chiang, J. Hockenmaier, and J. Tsujii, editors, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3518–3527, Brussels, Belgium, 7 2018. Association for Computational Linguistics.

[22] N. Lee, B. Z. Li, S. Wang, P. Fung, H. Ma, W.-t. Yih, and M. Khabsa. On unifying misinformation detection. In K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tur, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty,

and Y. Zhou, editors, *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5479–5485, Online, June 2021. Association for Computational Linguistics.

[23] Y. Lei and R. Huang. Sentence-level media bias analysis with event relation graph, 2024.

[24] Y. Lei, R. Huang, L. Wang, and N. Beauchamp. Sentence-level media bias analysis informed by discourse structures. In Y. Goldberg, Z. Kozareva, and Y. Zhang, editors, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 10040–10050, Abu Dhabi, United Arab Emirates, Dec. 2022. Association for Computational Linguistics.

[25] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. Roberta: A robustly optimized bert pretraining approach, 2019.

[26] I. Loshchilov and F. Hutter. Decoupled weight decay regularization, 2019.

[27] I. Maab, E. Marrese-Taylor, and Y. Matsuo. An effective approach for informational and lexical bias detection. In M. Akhtar, R. Aly, C. Christodoulopoulos, O. Cocarascu, Z. Guo, A. Mittal, M. Schlichtkrull, J. Thorne, and A. Vlachos, editors, *Proceedings of the Sixth Fact Extraction and VERification Workshop (FEVER)*, pages 66–77, Dubrovnik, Croatia, May 2023. Association for Computational Linguistics.

[28] I. Maab, E. Marrese-Taylor, and Y. Matsuo. Target-aware contextual political bias detection in news, 2023.

[29] N. Newman, R. Fletcher, K. Eddy, C. T. Robertson, and R. K. Nielsen. Reuters institute digital news report 2023. Technical report, Reuters Institute for the Study of Journalism, 2023. [Online; accessed 30-April-2024].

[30] N. Newman, R. Fletcher, A. Schulz, S. Andı, C. T. Robertson, and R. K. Nielsen. Reuters institute digital news report 2023. Technical report, Reuters Institute for the Study of Journalism, 2021. [Online; accessed 29-April-2024].

[31] C. Panagopoulos. Polls and elections accuracy and bias in the 2020 u.s. general election polls. *Presidential Studies Quarterly*, 51(1):214–227, 2021.

[32] H. H. Park, Y. Vyas, and K. Shah. Efficient classification of long documents using transformers, 2022.

[33] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in pytorch. In *NIPS-W*, 2017.

[34] T. E. Patterson and W. Donsbach. News decisions: Journalists as partisan actors. *Political Communication*, 13(4):455–468, 1996.

[35] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[36] P. Rafail and J. D. McCarthy. Making the tea party republican: Media bias and framing in newspapers and cable news. *Social Currents*, 5(5):421–437, 2018.

[37] Reuters Institute for the Study of Journalism. News consumers who saw false or misleading information about key topics in the last week in selected countries worldwide as of february 2023 [graph], June 14 2023. In Statista. Retrieved July 01, 2024, from `https://www.statista.com/statistics/1317019/false-information-topics-worldwide/`.

[38] Reuters Institute for the Study of Journalism. Share of adults who trust news media most of the time in selected countries worldwide as of february 2023 [graph], June 14 2023. In Statista. Retrieved July 01, 2024, from `https://www.statista.com/statistics/308468/importance-brand-journalist-creating-trust-news/`.

[39] T. Spinde, S. Hinterreiter, F. Haak, T. Ruas, H. Giese, N. Meuschke, and B. Gipp. The media bias taxonomy: A systematic literature review on the forms and automated detection of media bias, 2024.

[40] T. Spinde, M. Plank, J.-D. Krieger, T. Ruas, B. Gipp, and A. Aizawa. Neural media bias detection using distant supervision with BABE - bias annotations by experts. In M.-F. Moens, X. Huang, L. Specia, and S. W.-t. Yih, editors, *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 1166–1177, Punta Cana, Dominican Republic, Nov. 2021. Association for Computational Linguistics.

[41] T. Spinde, E. Richter, M. Wessel, J. Kulshrestha, and K. Donnay. What do twitter comments tell about news article bias? assessing the impact of news article bias on its perception on twitter. *Online Social Networks and Media*, 37-38:100264, 2023.

[42] T. Spinde, L. Rudnitckaia, J. Mitrović, F. Hamborg, M. Granitzer, B. Gipp, and K. Donnay. Automated identification of bias inducing words in news articles using linguistic and context-oriented features. *Information Processing & Management*, 58(3):102505, 2021.

[43] X. Su, T. Miller, X. Ding, M. Afshar, and D. Dligach. Classifying long clinical documents with pre-trained transformers, 2021.

[44] C. Sun, X. Qiu, Y. Xu, and X. Huang. How to fine-tune bert for text classification?, 2020.

[45] G. Vallejo, T. Baldwin, and L. Frermann. Connecting the dots in news analysis: A cross-disciplinary survey of media bias and framing, 2023.

[46] E. van den Berg and K. Markert. Context in informational bias detection. In D. Scott, N. Bel, and C. Zong, editors, *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6315–6326, Barcelona, Spain (Online), Dec. 2020. International Committee on Computational Linguistics.

[47] G. Van Rossum and F. L. Drake Jr. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.

[48] L. Wan, G. Papageorgiou, M. Seddon, and M. Bernardoni. Long-length legal document classification, 2019.

[49] D. M. White. The "gate keeper": A case study in the selection of news. *Journalism Quarterly*, 27(4):383–390, 1950.

[50] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. L. Scao, S. Gugger, M. Drame, Q. Lhoest, and A. M. Rush. Huggingface's transformers: State-of-the-art natural language processing, 2020.

[51] YouGov. Frequency of seeing false or misleading information online among adults in the united states as of april 2023, by age group [graph], May 8 2023. In Statista. Retrieved July 01, 2024, from `https://www.statista.com/statistics/1462057/false-news-consumption-frequency-us-by-age/`.

[52] YouGov. Level of confidence in distinguishing between real news from false news among adults in the united states as of april 2023 [graph], May 8 2023. In Statista. Retrieved July 01, 2024, from `https://www.statista.com/statistics/657090/fake-news-recogition-confidence/`.

[53] M. Zaheer, G. Guruganesh, A. Dubey, J. Ainslie, C. Alberti, S. Ontanon, P. Pham, A. Ravula, Q. Wang, L. Yang, and A. Ahmed. Big bird: Transformers for longer sequences, 2021.