Towards Automatic Bias Analysis in Multimedia Journalism

Reemt Hinrichs[1], Hauke Steffen[1], Hayastan Avetisyan[2], David Broneske[2] and Jörn Ostermann[1]

[1]Institut für Informationsverarbeitung, Leibniz Universität Hannover, Appelstr. 9a, Hannover, 30167, Germany.
[2] Infrastruktur und Methoden, Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH, Lange Laube 12, Hannover, 30159, Germany.

Corresponding authors: hinrichs@tnt.uni-hannover.de;

Abstract

This work investigates the application of machine learning for the analysis of video journalism to get insights into media bias in the German video journalism landscape. For this purpose, a custom dataset made up of subtitles from video data of major German news outlets ranging across the political spectrum was created. Media bias was assessed utilizing mention and sentiment analysis with respect to the major political parties in Germany. Sentiment analysis, performed using german-news-sentiment-bert, revealed significant differences in the reporting sentiment between media outlets. The German public broadcast outlet ARD was found to report with neutral sentiment less frequently than the mean, instead using negative sentiment significantly more often, especially while mentioning parties on the political edges. Mention analysis revealed that politicians get mentioned more often when in governing coalitions and, furthermore, it revealed a slight association between the assumed political ideology of media outlets and how frequently they report on political parties with a similar ideology, i.e., right-leaning outlets mention left-leaning parties and politicians less frequently and vice versa.

Keywords: media, journalism, bias, machine learning, mention analysis, sentiment analysis

1 Introduction

A key aspect of democratic states is the election of governing parties or individuals, directly or indirectly, by its citizens. To allow for a reasonable judgement in, e.g., the election choice, the citizens have to be informed about topics like policy decisions, relevance and backgrounds of social issues as well as questions of the day to day political establishment. In modern democracies, the idealized purpose of journalism is to provide this information for the citizens, as well as overseeing the government and preventing abuse of power [1]. However, for journalism to fulfill this idealized role in a democratic state, it is important to report in an unbiased manner, which encompasses, for instance, a neutral language, minimal distance to objective, measurable reality and distinguishing the relevant from the irrelevant [2].

The history of journalism is filled with doubts and concerns regarding its partiality. A recent example is the 2015/2016 refugee crisis in Germany [3], where many citizens placed at least some level of distrust in the German media, with some groups even going so far as to call parts of the media 'Lügenpresse' (lying press). One common source of concern are the ownerships of newspapers and broadcasting channels [4]. For example, in 2021, 54 % of the so-called opinion market in Germany were owned by just five companies [5]. Two of those are the German public broadcast companies Arbeitsgemeinschaft der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland (ARD) and Zweites Deutsches Fernsehen (ZDF), which are obliged by law, according to Section 26 (2) of the State Media Treaty [6], to comply with the principles of objectivity and impartiality in their reporting. Since every owner is assumed to have its own political agenda and economic interest [4], the unbiasedness of at least a subset of the journalistic production is a priori questioned. Another source for concern is the political ideology of journalists themselves, which likely have a tendency to be left-leaning [7]. Currently, applied analyses of media bias (e.g., [8, 9]) are mostly done manually by social scientists in the form of content analysis, frame analysis or meta analysis [4], requiring high amounts of expertise mostly done manually. Additionally, the domain tends to focus on the analysis of print media and partly television. In the past two decades, however, with increasing adaptation of the internet, journalism was published in a wide variety of formats: web audio journalism, web video journalism, web (print) journalism, classical print, television and radio journalism as well as journalism on social media platforms like Facebook and Twitter [10]. As such, manual analysis of the entire media sphere, even of a single country appears to be a rather challenging task. While automation sometimes is used, it is very rare to be used for non-print journalism and then only very basic tasks like term occurrences [8]. Consequently, the development of the media sphere forces a multimedia approach in the domain of media bias analysis. Since the space of multimedia journalism spans a wide range, and approaches from the social science require a high number of resources, automated methods from computer science would be desirable. Being able to automatically measure multimedia journalistic bias in a fast and reliable manner would make analyses much more robust and transparent, and would allow to (nearly) instantly objectively quantify media bias given a bias definition and characterization.

In this work, we are concerned with (semi-) automatic bias analysis of video journalism.

2 Background

2.1 Media Bias

The term "bias" in the context of journalistic products (text, audio, video) in general terms can be defined, with reference to Williams [4, 11], as an intentional, systematic and sustained deviation in the reporting or description from objective reality. To actually make use of this rather vague definition, further explanations, especially their introduced term "objective reality," are required.

Certainly, any journalistic product that goes beyond stating facts like the date an event occurred will contain segments where deviation from "objective reality" can be debated. Did a politician really make an annoyed comment or was this just the subjective impression of a journalist due to, e.g., his own upbringing, who is writing about a press conference? Perhaps no other journalist writing about this fictional press conference judged the comment as "annoyed". However, while absolute objective reality is for a large part of human interaction and communication a fiction, given a "reasonable observer", towards the extremes, objective descriptions and judgements can be made and are that of a "reasonable observer". Assume the same politician, but now assume he has a red face, raising his voice to extreme degrees and using very explicit language. Is this fictional politician annoyed? We argue, that any "reasonable observer" has to agree in this fictional case, due to the extremeness of the expression and the nature of human behavior making any other interpretation extremely unlikely.

In a similar manner advances in science are made, where scientific evidence at some point is sufficiently conclusive, that most researchers assume a theory to be valid.

In this sense, "objective reality" is not a point, but rather a domain in which journalistic products have to be contained in to be considered objective or accurate, or more precisely and lending from statistics, it cannot be rejected to be objective or accurate. Leaving this abstract domain yields biased journalism or media bias.

While the thoughts above suggest the general existence of media bias, it does not generally allow to construct such a "reasonable observer" to judge journalistic products with respect to existing media bias.

Take for example discussions and reports of Donald Trump's presidency including his candidacy. The reporting was generally overwhelmingly negative, the most extreme being the German news channel ARD with 98 % of news about Trump being negative, nonetheless, considerable differences in the reporting existed between individual news outlets [12]. Here, a "reasonable observer" was represented by the coders and the utilized codebook to judge reportings as positive, neutral or negative. However, we lack such a "reasonable observer" to judge, whether these 98 % of negative reports are an accurate representation of Trump's presidency or not. In general, to make such judgements, we have to equip this "reasonable observer" with some additional framework or idea of good and bad, right and wrong. Except for very extreme political stances, no such framework can generally be assumed, or it is not obvious how this should be done, and the "reasonable observer" cannot be used/assumed to judge observed tendencies in reporting objectively as (objectively) biased.

Due to this, analysis of media bias firstly has to reveal tendencies in characteristics of interest of journalistic products. These tendencies then have to be compared between, e.g., media outlets to observe differences in the descriptions and reportings.

The previous discussion focused on an attempt to objectify media bias generally. A related idea is that of the so-called hostile media phenomenon, which describes the perception of bias in media as a consequence of preexisting strong stances on certain issues in the observer [13]. The result is a tendency to perceive coverage, if not obviously in favor of ones own position, as antagonistic to ones position and thus as biased. However, this idea is basically the opposite of the previously described "reasonable observer". While certainly such cases exist, with the recent corona pandemic revealing quite a few instances of this phenomenon, this approach, to the conviction of the authors, should not be taken to the extreme, i.e., that media bias does or cannot exist and is a sole fabrication of the minds of the audience.

In the literature, motivated fundamentally by the aforementioned considerations, the term "bias" (as in media bias) usually is actually used to refer to tendencies (perhaps with respect to a certain characteristic) in the reporting with respect to, e.g., other outlets. In the following, the term "bias" will be used in this way.

2.2 Manifestations of Media Bias

The previous section discussed what media bias is or how it can be defined, but not how it practically manifests or what tendencies to look out for or measure. While several ways of how bias actually surfaces are proposed and used in the decades spanning literature on this topic – examples being spin and stance bias – a wide spread partitioning of the manifold ways in which bias can surface are gatekeeping bias, coverage bias and statement bias [2]. A possible reason, why these three types appear to be popular, might lie in the concreteness of these categories and more or less obvious ways to measure aspects of them, compared to vaguer concepts like spin bias, which can be described as mix of the three types of bias mentioned.

Gatekeeping bias refers to the selection of "newsworthy stories" out of the entirety of "newsworthy stories" a news outlet is reporting on. A typical example is the number of protests or rallies in a city that occur compared to the number being reported by some news outlets [14].

Coverage bias refers to the emphasis given to certain topics in the reporting within all topics covered by a news outlet. For example, a newspaper might publish many articles about one political party and only rarely cover rivaling parties.

Finally, **statement bias** refers to how a story is presented, i.e., favorably or unfavorably (or something in between). Statement or presentation bias, like gatekeeping and coverage bias, can surface in many different ways. A prime example is word choice, e.g., referring to "terrorists" rather than "freedom fighters" and similar decisions. Other examples could be the choice of pictures in an article, gestures and facial expressions of journalists in a video report or the tone of voice used. To get a complete picture of the bias of a media outlet, one would have to not only cover all three types of bias mentioned, but also consider all the different ways in which they can manifest, with the aforementioned examples, of course, being not exhaustive. This appears to be at least extremely difficult if not impossible and no published research effort attempted, let alone succeeded, in doing so. Due to this difficulty, research articles are usually concerned only with a smaller subset of possible types and/or manifestations of bias, e.g., only concern themselves with gatekeeping bias while ignoring how articles are phrased.

In the following, related research is described which is at least partly concerned with gatekeeping, coverage, or statement bias. Both, research from the social science and computer science are considered. A thorough review discussing ways in which the computer science can benefit from approaches in social science can be found in [4].

2.3 Related work

McCarthy et al. [14] investigated gatekeeping bias of several U.S. newspapers by comparing the reporting of local newspapers in Washington D.C. to nonlocal newspapers on demonstrations held in the city. Through obtaining police records of all rallies and demonstrations held in Washington in 1982 and 1991, a ground-truth was obtained and compared to the reporting of the news outlets. News articles and television broadcasts were manually assessed with the help of newspaper indices and broadcasting abstracts to reduce the workload. McCarthy et al. found, that the estimated demonstration size as well as the importance to the current media issue attention cycle predicted the coverage best.

Maurer et al. [15] analyzed the COVID-19 pandemic reporting of the German media landscape with respect to coverage and statement bias. A set of coders analyzed a selected number of articles and television broadcasts across a subset of the program covering approximately 15 months and several online newspapers as well as television channels. Due to the amount of data and their manual approach, only a subset of the totality of the media production was considered. The study finds no major bias with respect to, e.g., neutrality of reporting, but observed that a large proportion of actors mentioned in news reports about COVID-19 concerned politicians (especially one of the governing parties at the time, CDU), instead of more important actors, e.g., scientists or affected people.

Patterson [12] analyzed the coverage and statement bias of major media outlet's coverage of President Donald Trump within the first 100 days in office as well as during his candidacy. For the analysis, all mentions of Donald Trump covering more than five lines (text) or five seconds (television) were considered and the coders identified the news source, the topic, and the tone of the report. Patterson found that, while the media reported about President Trump significantly more than previous presidents, the coverage was also overwhelmingly negative, ranging from 52 % negative coverage for Fox News up to 98 % for the German broadcasting channel ARD.

Shultziner and Stukalin [16] analyzed how different media outlets framed the 2011 Occupy protest movement in Israel. They assessed statement and coverage bias, where the latter was split into topic, front-page and size bias. Front-page bias is the tendency to locate certain articles/topics prominently on the first page. Using three coders, six newspapers were analyzed daily for the entire period of the protest and an additional grace period. Statement bias was classified into positive, neutral, and negative according to so-called protest paradigm characteristics [17]. Topic bias was determined via classification into three different categories. The authors found that front-page headlines, articles sizes and additional visuals were strategically employed to align the reporting with the outlet's news ideology.

Grefenstette et al. [18] used sentiment analysis to measure how frequently positive words were used in vicinity of politician names in news reports, effectively measuring statement bias. For a given entity, articles were crawled and the 120 characters preceding and succeeding the entity were compared to an affect dictionary. An affect score was assigned to the entity by dividing the number of positive by the number of negative affects. The authors found that the extracted polarity correlated with public opinion of the specific medium, e.g., the conservative Washington Times mentioned the conservative president George Bush with positive sentiment more often than other media.

In 2018, Peng [19] investigated statement bias, in the sense of presentation bias, by analyzing how different news outlets produced ideological bias in their visual coverage. Artificial neural networks were used to assess automatically objective qualities like face angle, skin quality or number of background face in the media coverage of the former U.S. presidency candidates Hillary Clinton and Donald Trump used in news articles. Peng found that, e.g., Clinton was depicted more happily, while Trump was shown with a more angered expression. His work is one of the few that goes beyond text analysis, especially when considering automatic approaches.

Kim et al. [20] examined the dynamics of coverage bias of three U.S. television broadcasters by measuring the screen time of political actors across the period from 2010 to 2021. For this purpose, they used the Standford Cable

TV News Analyzer [21, 22], which contains the names and screen time of the individuals who appear on the screen. By comparing the channel's campaign finance to the amount of time an individual is on the screen, the authors derive a measure of media bias. They found that media bias was highly dynamic, even for shorter durations.

Congleton et al. [23], closely related to our work, investigated the political position of Dutch newspapers through, among others, party and politician mentions, i.e., they investigated coverage bias. They assessed normalized mention frequencies for parties and politicians in articles and evaluated coverage bias by comparing these normalized mentions to the party seats in the Dutch parliament. Using this metric, the authors observed underrepresentation of far-right parties in the news coverage of the considered newspapers. They also observed some coverage imbalance (too high or too low) for four politicians, including the far right.

Garz et al. [24] also analyzed German news outlets, however, they analyzed the social media content of the news outlets through automatic analysis of Facebook user engagement. They assess the outlet's slant through language similarity, namely, similarity to the language of the main political parties in Germany. The language analysis matched with the reputation of the media outlets, i.e., outlets usually assumed to be left-wing (right-wing) were found to use language more similar to left-wing (right-wing) parties' language.

Dewenter et al. [25] introduce the so called Political Coverage Index (PCI), which specifies the relative positioning of media within the political spectrum (left/right). The PCI is based on tonality of articles and newscasts on political parties and politicians. Using the PCI they assess the positioning of 38 media in Germany based on news items between 1998 and 2012 with respect to the formerly main political parties in Germany, SPD and CDU. Most importantly, they also consider television broadcasts, however, unlike our work, the data is hand-coded by Media Tenor International, a company specialized in media content analysis. Dewenter et al. find that media reports more negatively on governing parties.

Finally, Dallmann et al. [26], the work most similar to ours, assessed coverage and statement bias of four German online newspapers with respect to political parties of Germany. Coverage of parties was measured through counting the mentions of party names and politicians in all news articles within four years. Statement bias was measured through sentiment analysis and vocabulary similarity. Sentiment analysis was performed using SentiStrength [27] applied to each mention of a party together with the preceding and succeeding four words. The authors found statistically significant differences in the reporting of the newspapers depending on the party being reported on.

2.4 Contribution

From the cited literature, which represents not even remotely an exhaustive overview, it can be seen, that, while being more and more developed, almost all automatic approaches used in either social science or computer science (or their intersection) focus solely on textual journalistic products, be it actual newspaper articles or social media posts or similar things. Going beyond mere text, like Peng [19] or [22], is rare. To the best of our knowledge, no published research is dedicated to investigating media bias automatically from television audio. At best, abstracts or selected transcripts are used, e.g., in [16]. However, this represents a considerable gap in research and attempts should be made to close it. Otherwise, a large part of potential media bias will be missed by automatic (which currently do not consider television) and manual (which can only cover a small part of the entire television program) approaches. In this work, we take a first step towards closing this gap by investigating the applicability of (semi-)automatic mention and sentiment analysis techniques to video journalism for automatic bias analysis. More precisely, we assess aspects of coverage and statement bias in this work. We analyze subtitles extracted automatically from news videos of popular German TV channels, including the public broadcasting channels ARD and ZDF. Sentiment analysis is performed using german-news-sentiment-bert [28], which is a BERT-like model fine-tuned on migration-related German news articles. The aim is to investigate bias aspects of journalism with respect to the members of the German Bundestag as well as the main political parties in Germany [29], namely die Linke (engl. The Left), Bündnis 90/die Grünen (engl. Alliance 90/The Greens), Sozialdemokratische Partei Deutschlands (engl. Social Democratic Party of Germany, SPD), Christlich Demokratische Union Deutschlands (engl. Christian Democratic Union of Germany, CDU), Christlich-Soziale Union in Bayern (engl. Christian Social Union in Bavaria, CSU) and Alternative für Deutschland (engl. Alternative for Germany, AfD). The recently founded Bündnis Sahra Wagenknecht (engl. Sahra Wagenknecht Alliance, BSW) did not exist during the investigated time-frame and as such is not part of the analysis. Additionally, to assess the reliability of the sentiment and mention analysis, extensive manual verification is performed.

In summary, the overall contribution of our work is the following:

- Novel extension of mention and sentiment analysis to video journalism
- Automatic bias analysis of large parts of the german video journalism including the main public german broadcast
- Evaluation of german-news-sentiment-bert for subtitle-based sentiment analysis with respect to parties and politicians
- Evaluation of the reliability of Youtube's automatic subtitles for party and politician mention analysis

Our approach is closely related to the works of Dallmann et al. [26] and [23]. Both perform mention analysis very similar to us, and Congleton et al., like us, compare the normalized mentions to the seats of Dutch parties in the Dutch parliament. However, both focus on text journalism.

Table 1: Abbreviations and political ideology of the news outlets considered in the dataset.

(a) Left to center

Medium	Abbreviation	Political ideology
junge Welt	JW	left
NachDenkSeiten	NDS	left
taz	taz	left leaning
Süddeutsche Zeitung	SZ	left leaning
stern TV	St	left leaning
DER SPIEGEL	Sp	left leaning
ZEIT ONLINE	ZEIT	left leaning
Der Tagesspiegel	TS	center
ARD	ARD	center
ZDF	ZDF	center

(b) Center to right

Medium	Abbreviation	Political ideology
ZDFheute Nachrichten	ZDFh	center
Bayerischer Rundfunk	BR	center
ntv Nachrichten	ntv	center
RTL	RTL	right leaning
FOCUS Online	FOCUS	right leaning
faz	faz	right leaning
WELT	WELT	right leaning
BILD	BILD	right leaning
NZZ Neue Zürcher Zeitung	NZZ	right leaning
Junge Freiheit	JF	right
COMPACTTV	CTV	right

3 Methods and Materials

3.1 Dataset

Since no suitable dataset existed, a custom dataset was created for this work, consisting of subtitles derived from news videos scraped from the internet. The selection of news media outlets aimed to cover the most relevant news outlets in Germany as well as being balanced with respect to political ideology, i.e. to contain right and left leaning media as well as those with a more centered view. To determine the political ideology of the respective outlets, due to a lack of scientific literature regarding this question, several online resources [30–32] were combined to assign each media outlet with a political ideology. The news media outlets considered are given in Table 1 together with the respective abbreviation used throughout this manuscript and their assigned political ideology. Note that, aside of ARD, ZDF and Bayerischer Rundfunk, all news outlets considered are privately owned. In the first step of the dataset creation process, raw subtitle data was scraped from three different sources, namely YouTube [33] and the ARD and ZDF Mediathek [34, 35], which are the online streaming services of the public broadcasting companies of Germany that offer most programs of the previous year. All available subtitles of the news outlets named in Table 1 were downloaded from their respective YouTube channels. Most of the videos downloaded from YouTube contained auto-generated subtitles. Some, however, were manually created by the respective publisher. For a small portion of videos, less than 1 % and most stemming from the YouTube channel of Tagesschau, which has auto-subtitles turned off, there were no available subtitles at all, thus the corresponding videos were discarded. Since the German public broadcast did not provide much data on YouTube, we gathered more data from German public broadcast sources and used subtitles provided by the ARD and ZDF Mediathek, which were entirely manually created by the respective provider. This way, we scraped the past year of subtitle data for political talk shows Anne Will [36], hart aber fair [37], Maischberger [38], Markus Lanz [39], Maybrit Illner [40], as well as the political magazines Tagesthemen [41], frontal [42] and Monitor [43].

The Mediathek data was downloaded using the open-source software MediathekView [44]. To scrape subtitle data from YouTube, we used the python packages youtube-search-python [45] to find relevant videos published by the respective channels, as well as youtube-transcript-api [46] to download the subtitle data. The resulting subtitles were saved in a pandas dataframe containing the respective medium, video id, title, description, duration, publishing date and video category as columns. Since the two sources yielded data in different formats, we used regular expressions to transform the Mediathek data into the same format and subsequently combined the two dataframes. At this point in the process, each row of the dataframe contained information about one video. The duration of the videos in the dataset, however, showed considerable variance. While some videos were less than a minute long, some were up to twelve hours of live coverage. To combat this issue, we split each video into clips of one minute plus a residual clip with the remaining seconds of the video sequence.

While the dataset initially contained videos dating back up to 2009, upload statistics showed that until approximately 2017 media outlets published only a very limited amount of videos per day on their youtube channels. Ever since the amount of published videos per day has risen steadily.

Despite these earlier videos, the analyses presented in this work used only video material published in the 2017-2021 and 2021- legislatures.

3.2 Mention analysis

In the first step of the mention analysis, similar to Dallmann et al. [26], we created a list of every member of the German federal parliament and their respective political party in the current legislature (Bundestag 2021 [47]). Secondly, we created a list of search terms (see Appendix) for each party, e.g. "fdp" and "freien demokraten" (free democrats) for the Free Democratic Party FDP. For the AfD, we had to additionally consider the abbreviation "afg", which occurred rather frequently as an incorrect transcript in the subtitles of YouTube.

We then searched the document corpus for instances of the party or politician mentions, respectively. The result are two matrices $\boldsymbol{E}_S$ with $S \in \{\text{party}, \text{politician}\}$ containing the different media outlets as rows and the political parties as columns. The cells contain the number of mentions $e_{S,mp}$ of party p by medium m, while P and M are the number of total parties and media outlets in the dataset, respectively.

Since there is much more data of media outlet WELT than, e.g., Stern TV, we normalize the matrices by row, i.e. divide each value by the sum of the respective row, yielding the matrix $\bar{\boldsymbol{E}}_S = (\bar{e}_{S,mp})_{m,p} \in \mathbb{R}^{M \times P}$ with $\bar{e}_{S,mp} = \frac{e_{S,mp}}{\sum_{p=1}^{P} e_{S,mp}}$. Lastly, to be able to better compare the different media outlets, we subtract the mean of each column of the matrices, yielding the matrix $\hat{\boldsymbol{E}}_S = (\hat{e}_{S,mp})_{m,p} \in \mathbb{R}^{M \times P}$ with $\hat{e}_{S,mp} = \bar{e}_{S,mp} - \frac{1}{M} \sum_{m=1}^{M} \bar{e}_{S,mp}$.

In the final analysis, both the $\bar{\boldsymbol{E}}_S$ and $\hat{\boldsymbol{E}}_S$ matrices are considered for party mentions as well as politician mentions. The raw matrices $\boldsymbol{E}_S$ are provided in the appendix.

In a further approach, we split party and politician mentions in the time frames of the 2017 and the 2021 legislature respectively, ignoring all older videos. We then normalize the mentions by the number of Bundestag seats that the respective party had won in the 2017 or 2021 federal elections, respectively, i.e. $\tilde{\boldsymbol{E}}_{S,y} = (\tilde{e}_{S,y,mp})_{m,p} \in \mathbb{R}^{M \times P}$ with $\tilde{e}_{S,y,mp} = \frac{e_{S,mp}}{n_{p,y}}$, where $n_{p,y}$ is the number of seats the party p had in the Bundestag parliaments of $y \in \{2017, 2021\}$. For better comparison, we then normalize each row of the matrix $\tilde{\boldsymbol{E}}_{S,y}$ such that the corresponding value of the AfD party equals to 1, i.e., each row was divided by the respective value of the AfD party.

3.3 Sentiment analysis

For sentiment analysis, we used german-news-sentiment-bert [28], which is a sentiment analysis model based on BERT [48] and fine-tuned on German news articles.

We extracted party mentions as described in the previous section and considered the ten preceding and ten following words alongside the party mention in contrast to [26], which only used the four preceding and following words. The sentiment of these texts was then classified by german-news-sentiment-bert, resulting in three matrices $\boldsymbol{S}_K$ containing the counts $s_{K,mp}$ of classifications with sentiment $K \in \{\text{pos, neu, neg}\}$ by media outlet m concerning political party p. Analogous to mention analysis, we then normalized the sentiment classification counts in each row by dividing each cell with the respective mention counts, yielding the matrix $\bar{\boldsymbol{S}}_K = (\bar{s}_{K,mp})_{m,p} \in \mathbb{R}^{M \times P}$ that contains the proportions $\bar{s}_{K,mp}$ of mentions with sentiment K of party p by medium m, i.e. $\bar{s}_{K,mp} = \frac{s_{K,mp}}{e_{mp}}$.

Finally, we subtract the mean of each column to compare between outlets. The components of the resulting matrices $\hat{\boldsymbol{S}}_K = (\hat{s}_{K,mp})_{m,p} \in \mathbb{R}^{M \times P}$ then contain the deviation from the mean of classification proportions with sentiment K of party p across all media, i.e. $\hat{s}_{K,mp} = \bar{s}_{K,mp} - \frac{1}{M} \sum_{m=1}^{M} \bar{s}_{K,mp}$.

In our analysis, we considered the matrices $\bar{\boldsymbol{S}}_K$ as well as $\hat{\boldsymbol{S}}_K$ and provide the raw data in the form of the matrices $\boldsymbol{S}_K$ in the appendix.

3.4 Manual verification

As the sentiment analysis performance of german-news-sentiment-bert [28] used in this work has only been assessed and compared to the sentiment perception of humans for migration related news articles (i.e., text), manual verification of this BERT model was performed.

For this purpose, to guarantee the found party and politician mentions really are concerned with the correct entities, 100 samples of party mentions and 100 samples of politician mentions per party, as previously described,

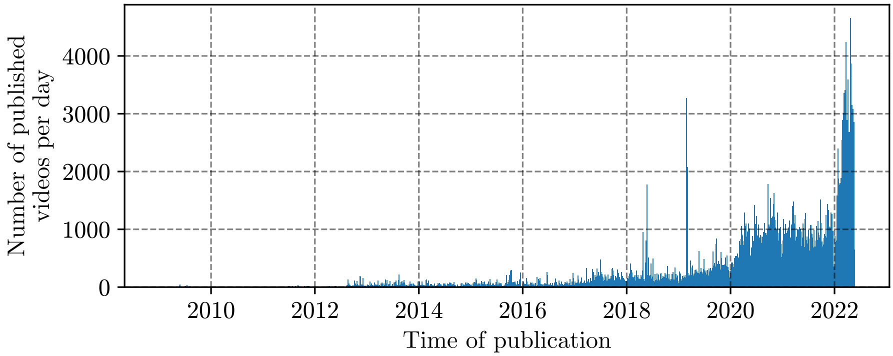


Fig. 1: Number of published videos per day over time in our dataset spanning approximately the years 2009 to mid 2022.

were randomly selected for each of the seven political parties. It was manually checked whether they actually referred to the party or politician of concern.

Additionally, 400 samples used in the mention and sentiment analyses, all of them consisting of a party mention and the ten preceding and following words as described in sentiment analysis subsection, were randomly selected and their sentiment - positive, neutral, or negative - was assessed by the two main authors without knowledge of the sentiment assigned by german-news-sentiment-bert. Then, the sentiment assigned by german-news-sentiment-bert and the sentiment assigned by the humans were compared.

4 Results

4.1 Dataset statistics

Figure 1 shows the number of published videos per day over time in our dataset. The oldest videos are from early 2009, the newest videos from May 2022.

Table 2: All considered media together with the total duration of the corresponding video material as well as its proportion to the entire dataset.

(a)

Medium	Duration [min]	Proportion
WELT	256716	28.0 %
BILD	191619	20.9 %
ZDFheute Nachrichten	89785	9.8 %
faz	68397	7.5 %
Bayerischer Rundfunk	66160	7.2 %
DER SPIEGEL	49539	5.4 %
ntv Nachrichten	46580	5.1 %
taz	26694	2.9 %
NachDenkSeiten	23686	2.6 %
ZDF	17418	1.9 %
COMPACTTV	16708	1.8 %

(b)

Medium	Duration [min]	Proportion
ARD	13753	1.5 %
FOCUS Online	12705	1.4 %
stern TV	6943	0.8 %
ZEIT ONLINE	5852	0.6 %
Junge Freiheit	5517	0.6 %
RTL	5440	0.6 %
Der Tagesspiegel	2961	0.3 %
NZZ Neue Zürcher Zeitung	2613	0.3 %
Süddeutsche Zeitung	2470	0.3 %
junge Welt	2139	0.2 %
ZDFinfo Dokus & Reportagen	1639	0.2 %
tagesschau	1362	0.1 %
Total	916696	100 %

Table 3: Proportion of correct mentions extracted out of 100 samples of party mentions and 100 samples of politician mentions for each party, randomly selected from our dataset.

Party	Linke	Grüne	SPD	FDP	CDU	CSU	AfD
Correct Party Mentions [%]	93	98	100	99	100	100	100
Correct Politician Mentions [%]	100	100	100	99	100	100	100

Perhaps unsurprisingly, we see a significant increase in the number of published videos over time, indicating an increasing adoption of YouTube as a platform for German news outlets. Furthermore, there is a recognizable peak after 2022, which is where most of the Mediathek data stems from.

Table 2 shows the amount of data our dataset contains for each news outlet and the corresponding proportion of the whole dataset. Since there was significant variance in the amount of data across media outlets and some outlets provided tiny amounts of subtitle data, we decided to set a cutoff of 10000 minutes and discard the data of any YouTube channel with less available material. Noteworthy is the large share of BILD and Welt, together amounting to close to 50 % of the entire video material. Both belong to the Axel Springer SE, which appears to see greater economic relevance in video journalism.

Additionally, to give a rough idea of the general content of the dataset, Table 7 in the appendix gives an overview of topics discovered in the dataset applying BERTopic, a Bert-like model which generates topic representations through class-based term-frequency procedures, to the subtitle data. The discovered topics generally agree well with the subjective impression of the authors, e.g., the top topic being the Russian-Ukrain war should be expected by anyone following German news. Similarly, even rarer topics like potential cannabis legalization have indeed been presented in the news from time to time.

4.2 Manual verification

4.2.1 Correctness of Party and Politician Mentions

Table 3 shows the number of incorrect party and politician mentions, respectively. The following paragraphs show some sentences our method extracted erroneously, along with the translation and a description of why the error occurred.

By far the most problematic search term was "die linke" (engl. the left) which yielded results that did not concern the party in as much as 7 out of the 100 samples. One common mistake, which occurred five times in our random sample of 100 mentions, was the mentioning of a left side of something, e.g. the left side of a soccer pitch. The other two errors were more intricate, with phrases like

> "seit jahren ist es zur mode der scheinheiligen rechten geworden die linke als buchstäblich korrupt darzustellen wie hier in recoleta haben eine" (engl. "for years

it has become the fashion of the hypocritical right to portray the left as literally corrupt as here in recoleta have a")

mentioning the political left as a whole, instead of the left as a party. Errors for Die Grünen were of similar nature, with the phrase

"ja ich weiß es nicht also entweder einer von reifen die grünen ideen habe ich würde ich dann will ich wirklich der" (engl. "yes I don't know so either one of mature the green ideas I have I would then I really want the")

mentioning the green ideas (e.g. "anti" nuclear power, importance of climate protection, ...) instead of the green party. The other incorrect party mention for Die Grüne, as well as the incorrect party mention for FDP, stem from a transcribing error made by YouTube. The only incorrect politician mention we found in our sample was for FDP politician Nicole Bauer, which instead concerned a cyclist with the same name.

4.2.2 Correctness of Sentiment Analysis

Figure 2 shows the results of our manual verification of the sentiment analysis. For our sample, the model has a weighted F1-Score of 0.578 and a balanced accuracy of 54 % (unbalanced 58 %), which is similar to related results for sentiment analysis of social media posts [49, 50]. Cohen's kappa is 0.32 corresponding roughly to "fair" observer agreement [51]. F1-Scores in sentiment analysis of news articles are usually at least slightly higher, but seeing that the investigated subtitles represent, at least in parts, ad-hoc speech, the results being similar to analyses done on social media posts is not quite surprising and more work has to be performed to improve performance for our research topic. It is apparent that predictions made by the model are most reliant on

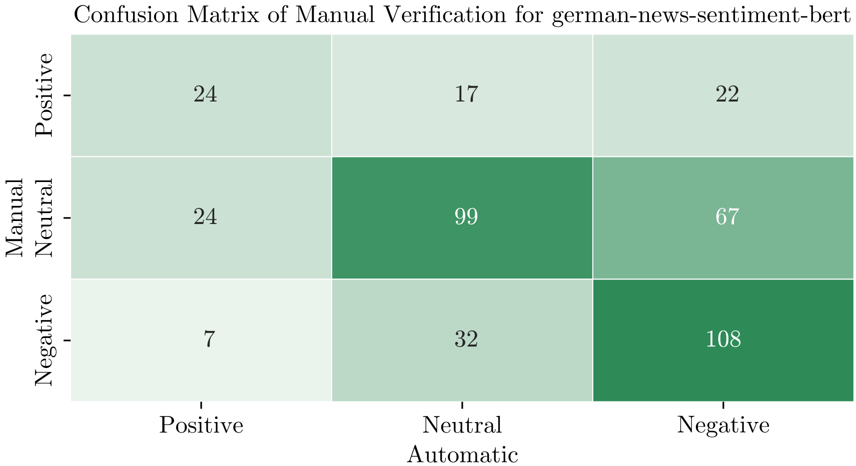


Fig. 2: Confusion matrix of manual verification of sentiment classifications of german-sentiment-news-bert.

Table 4: Sentiment classifications of german-news-sentiment-bert of mentions of Donald Trump in the first 100 days of his presidency and across our complete dataset by ARD broadcasts.

Data\Sentiment	Mentions	Positive	Negative
First 100 Days	200	18	75
Complete Dataset	13655	865	4109

sentences with neutral and negative sentiment. This is in line with the original technical report describing german-news-sentiment-bert [52], which found a tendency towards neutral and negative sentiment. The cause, also discussed in [52], is likely the dataset used for fine-tuning, which consisted of migration related news articles, which had a bias towards articles with neutral and negative sentiment. From this result, we must conclude that positive sentiments outputs should not be trusted, however, neutral and negative sentiment outputs are considerably more trustworthy. The discussion section picks up on this point.

To further verify the reliability of the model, we examined sentiment classifications of the former US President Donald Trump in our dataset (see Table 4). In former research [12], it has been found that in his first 100 days of presidency, German media mentioned Trump with overwhelmingly negative sentiment. For the ARD, a proportion of 98 % of the reporting on Trump was found to be negative. Since [12] used only positive and negative sentiment, we discarded the mentions that german-news-sentiment-bert classified as neutral and got a proportion of negative mentions of 80.6 % in the first 100 days of Trump's presidency and 82.6 % across our whole dataset, which qualitatively seems to agree well with previous research.

4.3 Mention analysis

Figures 3a and 3c show a heatmap plot of the matrices $\bar{\boldsymbol{E}}_{party}$ and $\bar{\boldsymbol{E}}_{politician}$ respectively. In the Figures 3b and 3d, we subtracted the mean of each column, showing the matrices $\hat{\boldsymbol{E}}_{party}$ and $\hat{\boldsymbol{E}}_{politician}$. In all Figures, media outlets are ordered from politically left on the top to politically right on the bottom and political parties are sorted by the seating arrangements in the German Bundestag, which correlates with the political ideology of the respective parties [53]. The ordering of the outlets is largely derived from [32] as mentioned in the dataset subsection. Table 5 shows the mean values of party and politician mentions by each party, respectively.

Figures 3a and 3c show the proportion of party and politician mentions by medium and party, respectively. For example, 33.8 % of all party mentions of the rightmost medium CTV concern the rightmost party AfD in our dataset. Similarly, 18.8 % of politician mentions of the Bavarian media outlet Bayerischer Rundfunk (BR) concern the Bavarian party CSU.

As expected, the big parties CDU and SPD get mentioned the most in our dataset with respect to both party and politician mentions.

Fig. 3: (a) Percentage of party mentions by the respective medium. CDU and SPD are mentioned frequently, Die Grünen, CSU, rather less. (b) Same data but with columnwise zero mean. (c) Percentage of politician mentions by the respective medium. CDU and SPD are mentioned most frequently, while politicians of all other parties, especially Die Linke, CSU and AfD get mentioned less. (d) Same data but with columnwise zero mean. Media outlets appear to mention politicians of similar political ideology more often although the pattern is not as clear as with party mentions.

A one-way analysis of variance (ANOVA) found significant difference for both party mentions ($F(6, 70) = 18.3, p < 0.05$) and politician mentions ($F(6, 70) = 52.21, p < 0.05$). Furthermore, a Wilcoxon signed-rank test was performed to investigate pairs for party mentions as well as politician mentions ($m = 21$). After applying Bonferroni adjustment to the significance level of $p = 0.05$, the new threshold of significance was $p/m = 0.0024$.

The signed-rank test revealed the differences between CDU and SPD to be not significant for party mentions as well as politician mentions ($p > 0.0024$). Interestingly, the FDP appears to be mentioned more frequently as a party than Die Grünen, while for politician mentions, politicians of Die Grüne appear to be mentioned more frequently, albeit, after Bonferroni adjustment, only politician mentions remain statistically significant ($p < 0.0024$). For additional insight, before adjustment, both comparisons are significant with respect to a threshold of significance of 0.05.

Aside from getting significantly less politician mentions compared to Die Grüne, SPD and CDU ($p < 0.0024$), Die Linke also gets significantly less party mentions than all other parties ($p < 0.0024$) with the exception of CSU and Die Grüne. Considering around 7 % of our party mentions for Die Linke could be false positives of our mention extraction process, the real amount of party mentions could even be lower. Furthermore, politicians of CSU and AfD rarely get mentioned, with no significant differences between the two parties ($p > 0.0024$). The CSU is a local party of the state of Bavaria, but forms a traditional coalition with the CDU in the German Bundestag, and one possible explanation for the low amount of politician mentions of Die Linke and AfD is that they never were governing parties in the time frame of our analysis [54, 55]. A Wilcoxon signed-rank test between politician mentions of Die Linke and AfD found the differences to be not significant ($p > 0.0024$). Regarding the CSU, for both party and politician mentions, Bayerischer Rundfunk (BR) shows much higher values than other outlets, which is due to BR being a local medium in Bavaria, where the CSU has been the ruling party for decades.

Figures 3b and 3d also show proportions of party and politician mentions respectively, this time, however, the mean of each party has been subtracted, allowing for better comparison between news outlets. In both plots, but especially party mentions, we can see the pattern of media outlets mentioning parties with the same political ideology more often, resulting in higher values near the diagonal of the plots. For example, NDS, a medium on the political left, mentions Die Linke, a party on the political left, 5.5 % points more often than the mean. CTV, a medium on the political right, mentions AfD, a party on the political right, 20.8 % points more often than the mean. For politician mentions, the same pattern emerges, however more outliers are showing, e.g., CTV mentioning politicians of Die Linke 11.8 % more often than the mean. In the appendix the same evaluation is given, however, based on unique mentions, i.e., multiple mentions within a one-minute snippet of a political entity is counted as a single mention. While unique mentions are still of interest, we believe that multiple mentions are done for a reason, i.e., they convey an additional, underlying, importance, and as such are presented in the main body of this manuscript.

Figures 4a and 4b show the matrices $\tilde{\boldsymbol{E}}_{S,y}$ with $S \in \{\text{party, politician}\}$ and $y = 2017$ as previously described. This means, they show normalized party and politician mentions, where the mentions are normalized to the number of seats in the German parliament for the 2017-2021 legislature, and for ease of interpretation, the results were then normalized to the respective values of the AfD. If mention frequency of a party or politician corresponds well to the share of seats in the German parliament, a value close to 1 should be observed.

A one-way ANOVA found the differences between parties to be insignificant for party mentions per Bundestag seat ($F(5, 60) = 1.69, p = 0.15$), however an analysis of politician mentions per Bundestag seat ($F(5, 60) =$

Table 5: Mean values of party and politician mentions by party, respectively.

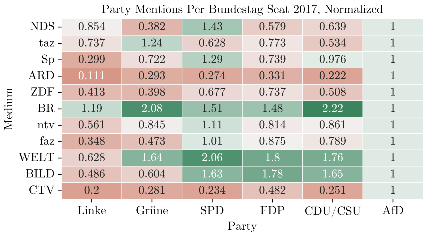
Type / Party	Linke	Grüne	SPD	FDP	CDU	CSU	AfD
Party Mentions	5.3 %	9.5 %	25.3 %	14.9 %	21.7 %	10.3 %	12.9 %
Politician Mentions	6.1 %	18.6 %	33.1 %	7.2 %	27.7 %	4.1 %	3.1 %

$3.06, p = 0.0159$) revealed significant differences. A Wilcoxon signed-rank test was performed for all possible pairs of corresponding parties. After a Bonferroni adjustment with $m = 15$, the significance level of $p < 0.0033$ was used. After adjustment, AfD politicians are significantly less mentioned than politicians of all other parties ($p < 0.0033$) with the exception of Die Linke ($p > 0.0033$). No other comparison is significant after Bonferroni adjustment. For additional insight, before adjustment, with a threshold of significance of 0.05, politicians of Die Linke get significantly less mentioned than Die Grüne, SPD and CDU/CSU ($p < 0.05$)

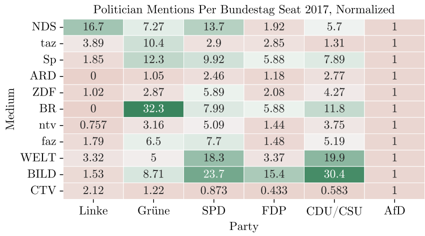
While no significant difference for the entirety of the considered media was observed, it is apparent that in the 2017 legislature, both ARD and ZDF mention parties other than AfD relatively less per Bundestag seat. For Die Linke, the difference is close to a factor of 10 compared to the AfD. On the contrary, BR, WELT and BILD mention center and left parties more often than AfD. The rightmost medium COMPACTTV mentions the AfD as much as five times as often as other parties with respect to the number of Bundestag seats. Figure 4b shows that even when correcting for the number of Bundestag seats in the given legislature, politicians of big parties get mentioned overwhelmingly more often, with media outlet WELT mentioning politicians of SPD and CDU/CSU 18.3 or 19.9 times more often than AfD politicians with respect to the number of Bundestag seats. The left media outlet NachDenkSeiten mentions politicians of parties other than AfD disproportionately more often, especially Die Linke gets mentioned 16.7 times more compared to their Bundestag seats. An exception is COMPACTTV, which as the sole exception underrepresents SPD, FDP and CDU/CSU compared to the AfD. Interestingly, COMPACTTV slightly overrepresents Die Linke and Die Grünen. An obvious explanation are attacks against their left wing politics.

For completeness, Fig. 11b and Fig. 11a in the appendix depicts the same data for the 2021 legislature. Due to more limited amount of data, no detailed statistical analysis was performed for this period.

Figures 5a and 5b show the mean value of each party in the matrices $\tilde{\boldsymbol{E}}_{S,y}$ with $S \in \{\text{party, politician}\}$ and $y \in \{2017, 2021\}$. To account for outliers, we removed the extreme media outlets NDS and CTV and the local medium BR from this calculation. The party mentions show that in the 2017 legislature, across the German media landscape, most parties get mentioned relatively evenly when correcting for the number of Bundestag seats, with the exceptions of Die Grünen and especially Die Linke. As mentioned before, in the 2017 legislature, politicians seem to get mentioned overwhelmingly more often in the big parties, which could be due to the fact that in the 2017 legislature, CDU and SPD formed a governing coalition. The same pattern emerges in the 2021 legislature, where politicians of Die Grünen, SPD and FDP, which formed a governing coalition, get mentioned more often than other parties, especially AfD, with SPD getting close to 50 times the amount of politician mentions that AfD got in our dataset. Additionally, our dataset only covers a small fraction, less than one year of the four year legislature, of the 2021



(a)



(b)

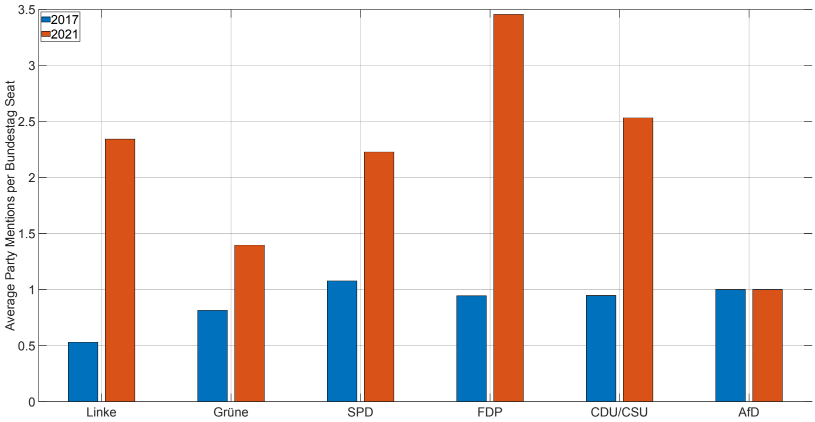
Fig. 4: Party and politician mentions per Bundestag seat by party and medium in the 2017 legislature. Each row is normalized to the value of the AfD. (a) ARD and ZDF mention parties other than AfD relatively less per Bundestag seat. BR, WELT and BILD mention centered parties relatively more with respect to their number of Bundestag seats. For instance, WELT mentions SPD more than two times as much per Bundestag seat than AfD. (b) CDU and SPD politicians get mentioned most across all media. For the left medium NDS, politicians in left leaning parties get mentioned significantly more than politicians on the right, with respect to the number of Bundestag seats.

onward legislature. Due to this, early media attention, which can be greatly affected by government changes, might bias the results. Extending the period of observation potentially evens these apparent extreme values out.

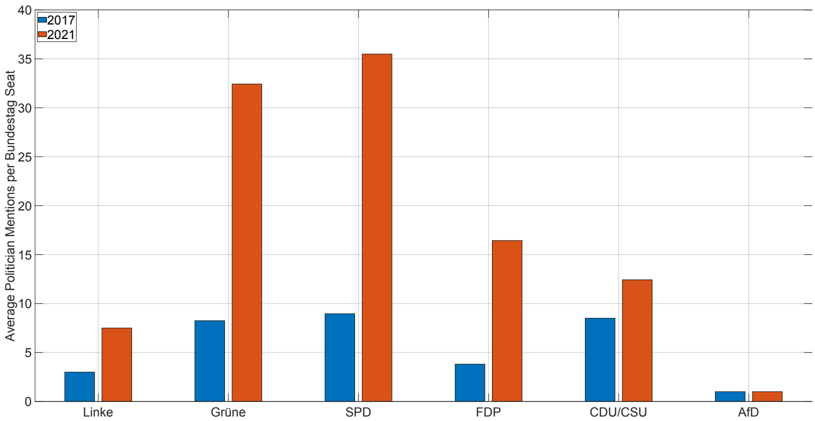
4.4 Sentiment analysis

Figures 6, 7 and 8 show the results of our sentiment analysis in form of the matrices $\bar{\boldsymbol{S}}_K$ and $\hat{\boldsymbol{S}}_K$ with $K \in \{pos, neu, neg\}$, respectively. Figure 9 shows sentiment analyses of different formats and subsets of the German public broadcast. Table 6 shows the mean values across all sentiments and media outlets. Figure 6a shows the proportion of party mentions of the respective medium that were classified as positive sentiment by german-news-sentiment-bert. A one-way ANOVA found differences between parties to be significant, with $F(6, 70) = 2.25$ and $p = 0.0486$. After Bonferroni adjustment with $m = 21$ however, a Wilcoxon signed-rank test found no significant differences between any parties ($p > 0.0024$). For additional insight, before adjustment, with a threshold of significance of $p = 0.05$, AfD mentions are significantly less frequently positive than mentions of Die Grüne, SPD and FDP ($p < 0.05$).

Across all media outlets in our dataset, AfD gets mentioned positively least frequently with a mean value of 9.5 % of party mentions that get classified as positive. Die Grünen get mentioned with positive sentiment most often and, with a mean of 14.4 %, have a higher proportion of positive mentions than the AfD. Figure 6b shows the same data but with column means subtracted to better compare between outlets. It is apparent that the most extreme outlets in our dataset, NDS and CTV, both show relatively less mentions with



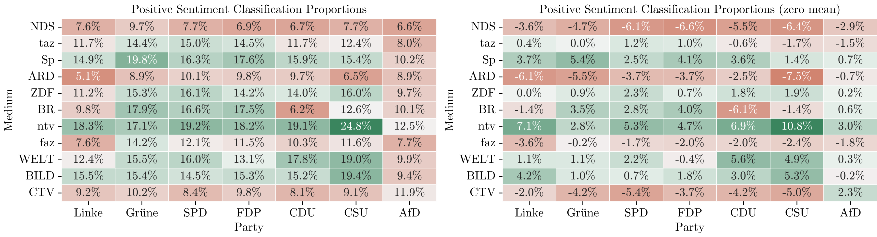
(a)



(b)

Fig. 5: Mean value of party mentions per Bundestag seat by party and legislature across all media outlets except NDS, CTV and BR. Normalized on the value of AfD. (a) SPD, FDP, CDU/CSU and AfD get mentioned relatively evenly in the 2017 legislature, while Die Grünen and especially Die Linke get mentioned relatively less when correcting for the number of Bundestag seats. (b) In the 2017 legislature, politicians of the big parties, especially SPD and CDU/CSU, get mentioned overwhelmingly more often than politicians of smaller parties, especially Die Linke and AfD.

positive sentiment than other outlets, regardless of the political party mentioned. The only exception is the AfD that gets mentioned positively by the right medium CTV more often than the mean outlet. The public broadcasting outlet ARD also shows the pattern of mentioning all political parties with positive sentiment less frequently. In contrast, outlets like Spiegel and ntv mention every party positively more often than the mean. Figure 7a shows the proportion of party mentions by medium and party that german-news-sentiment-bert



(a) (b)

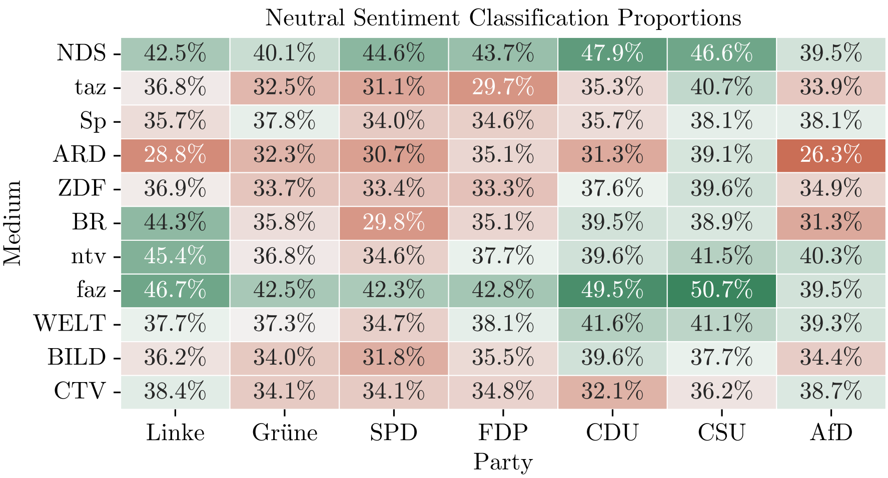
Fig. 6: Proportion of party mentions of the respective medium that were classified as positive. (a) The AfD is mentioned with positive sentiment least frequently across almost all media in our dataset. The largest proportions of positive sentiment classifications are attributed to Die Grünen, SPD and FDP. (b) Same data but with columnwise zero mean.

classified as neutral. Although all parties seem to have similar mean values of neutral sentiment classification proportions (see Table 6), Die Linke and CDU show much higher variance than other parties, especially Die Grünen. A one-way ANOVA found significant differences between the respective parties with $F(6, 70) = 2.77$ and $p = 0.02$. However, after Bonferroni adjustment with $m = 21$, a Wilcoxon signed-rank test found no difference to be signficant ($p > 0.0024$). For additional insight, before adjustment, with a threshold of significance of $p = 0.05$, differences are signficiant for AfD and CDU, SPD and Die Linke as well as CDU and SPD ($p < 0.05$). Figure 7b shows the same data but with columnwise zero mean. While media outlets like NDS and faz show a higher proportion of neutrally classified sentiments than the mean outlet, the classification proportions of neutral sentiment by public broadcasting outlets ARD and ZDF are lower than the mean across all parties, with ARD deviating as much as -10.2 % points for Die Linke and -9.7 % points for AfD. Figure 8a shows the proportions of party mentions that german-news-sentiment-bert classified as negative by medium and party. Across all outlets, AfD gets mentioned with negative sentiment most frequently (mean value: 54.4 %), while the Bavarian CSU gets mentioned negatively least frequently (mean value: 45.0 %). A one-way ANOVA found the differences between parties to be significant ($F(6, 70) = 2.77, p = 0.018$). In all Wilcoxon signed-rank tests performed,

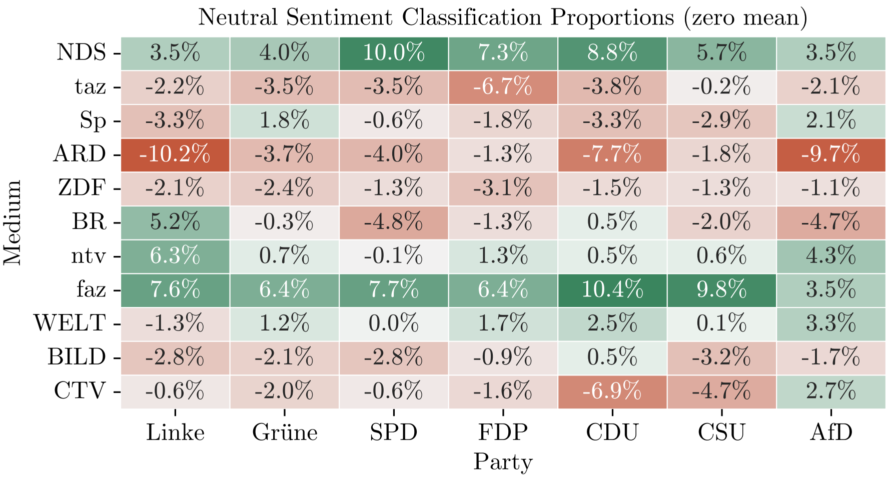
Table 6: Mean value per party of the proportion of party mentions classified as positive, neutral and negative in the data set. For example, on average 49.5 % of all party mentions of Die Grünen across all investigated media are negative.

Sentiment\Party	Linke	Grüne	SPD	FDP	CDU	CSU	AfD
Positive	11.2 %	14.4 %	13.8 %	13.5 %	12.2 %	14.0 %	9.5 %
Neutral	39.0 %	36.1 %	34.6 %	36.4 %	39.1 %	40.9 %	36.0,%
Negative	49.8 %	49.5 %	51.5 %	50.1 %	48.7 %	45.0 %	54.4 %

after a Bonferroni correction with $m = 21$, only AfD and CSU show a significant difference ($p < 0.0024$). For additional insight, before adjustment, with a threshold of significance of $p = 0.05$, differences between AfD and Die Linke, AfD and Die Grüne and AfD and FDP are significant ($p < 0.05$). Even when ignoring adjustment, AfD and SPD do not show a significant difference in negative sentiments ($p > 0.05$).

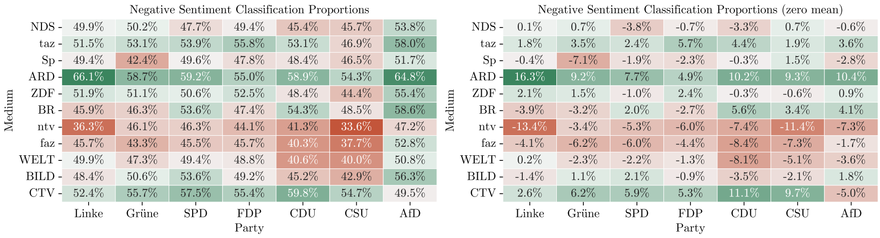


(a)



(b)

Fig. 7: Proportion of party mentions in the respective medium that were classified as neutral. (a) Linke, Grüne and CSU have a high proportion of neutral mentions, whereas the SPD and FDP are reported on with neutral sentiment least frequently. (b) Same as (a) but with columnwise zero mean.



(a) (b)

Fig. 8: Proportion of party mentions of the respective medium that were classified as negative. (a) The highest proportions of negative sentiment classifications were observed for the SPD, the FDP, the CSU and the AfD. All other parties, especially the Grüne, were comparatively rarely mentioned with negative sentiment. (b) Same as (a), but with with columnwise zero mean.

In Figure 8b, we removed the column means, thus showing the respective outlets deviation from the mean outlet. While outlets like ntv and faz showed a negative deviation from the mean across all parties, meaning they mention all parties with negative sentiment less frequently than the mean, outlets like taz, CTV and ARD mention parties with negative sentiment more often.

To further examine the situation with the seemingly significant deviations of the media outlets ARD with respect to our sentiment analysis, we investigated the sentiment analysis for each of the main public broadcasting formats in our dataset, although our dataset did not provide enough data for this niche case to allow for conclusive evidence. Figure 9a shows the sentiment analysis results for all three sentiments with respect to the formats in our dataset. One-way ANOVA, however, found no significant differences between parties in all three cases ($p > 0.05$). Tables 14, 15 and 16 show the underlying data basis of this sentiment analysis. While, on average, the AfD was found to be mentioned in the most negative light with 69.4 % of mentions being assigned a negative sentiment, which is a reasonable result, the AfD being mentioned with positive sentiment in 14.6 % of the mentions by the ARD magazine Monitor is rather implausible and most certainly due to at least the lack of data, seeing that this value is based on seven mentions in total.

5 Discussion

The most robust and major finding is a considerable deviation in coverage of the AfD for the 2017-2021 legislature by the two largest German public broadcasters ARD and ZDF compared to the baseline obtained by assessing the corresponding Bundestag seats of the party and comparing it to the reporting. This baseline was also used by Congleton et al. for the analysis of the reporting of Dutch newspapers [23]. An accusation repeatedly made in the past against the public broadcasters in Germany is that their program overrepresents the

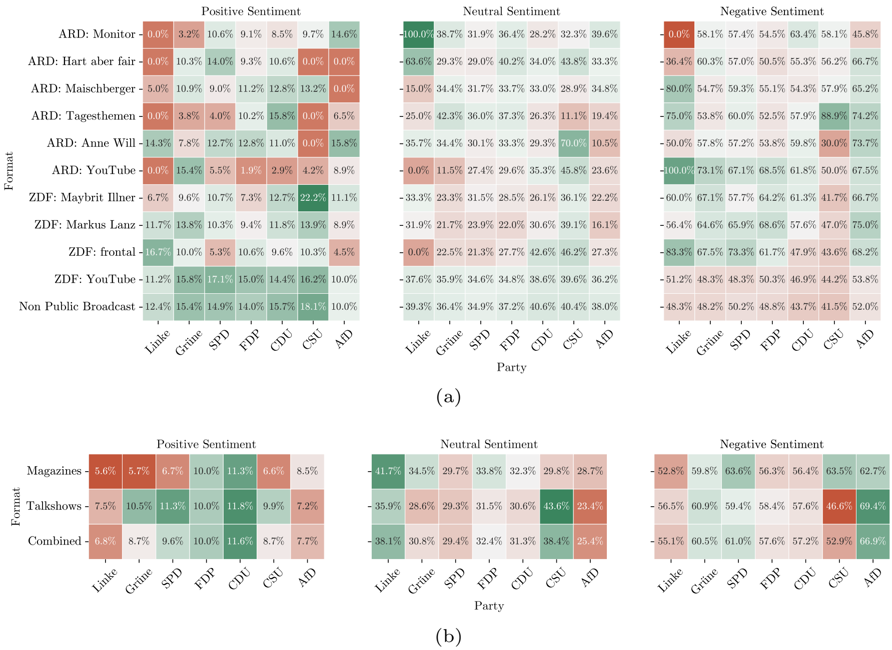


Fig. 9: (a) Proportion of party mentions classified by german-news-sentiment-bert as positive, neutral or negative sentiment broken down by format of ARD and ZDF Mediathek, including political talk shows as well as major political magazines. YouTube data of ARD and ZDF as well as non-public broadcast media were included for comparison. (b) ARD and ZDF Mediathek data consolidated into subsets of magazines and talk shows.

AfD to the benefit of the AfD with respect to their election turnouts [56] (see also [57] for an in-depth discussion). While some of these accusations are about inviting politicians of the AfD to, e.g., talk shows, and this point was not investigated by us (but should show in mentions anyway), our results, see Fig. 4a, are consistent with these accusations for the 2017-2021 legislature. The AfD was mentioned up to about five times (even almost ten times compared to Die Linke for the ARD) as much as other parties if one considers the share of seats in the German Bundestag as a baseline. While the reporting appears to be in favor of the other parties regarding politician mentions, see Fig. 4b, ARD, and to a lesser extent ZDF, are similar to the extreme right media outlet CompactTV, seeing that CompactTV is the only outlet which overrepresents AfD politicians. Overrepresentation here means, that the percentage of mentions surpasses the percentage of seats in the German Bundestag for the respective legislature.

Due to the sensitivity of the research topic, it should be made very clear, that "overrepresentation" with respect to the measure of choice (here: mentions per Bundestag seat) does not imply a definite or actual bias. There are many reasons such an overrepresentation can occur without an actual bias. For example, perhaps the AfD was considered to support a more newsworthy, controversial opinion regarding migration to Germany, which the ARD, for which the largest overpresentation was observed, considered to be more newsworthy due to the AfD being a major oppositional party.

Such a reasoning however does not explain the considerable differences between media outlets.

This line of reasoning is discussed in greater detail in Subsection 5.1.

ARD actually achieves rather balanced mentions of politicians of all parties, unlike other media. Note, that in the 2017-2021 legislature, SPD and CDU were the ruling parties, and thus a few SPD and CDU politicians were part of the government and thus held "special" positions. Regarding private media outlets, Spiegel, NTV, and FAZ, and to a lesser extent WELT and Bild, are most balanced regarding party mentions. Findings are rather different regarding politician mentions, where almost all outlets underrepresent the AfD. Extremes are Nachdenkseite and Bild. Nachdenkseite extremely overrepresents politicians of Die Linke, which is in line with their presumed left wing ideology. CompactTV also overrepresents politicians of Die Linke and to a less degree of Die Grünen. An obvious explanation are attacks against left wing politicians. These observed deviations (regarding the AfD) are not present for the legislature that began in 2021, see Fig. 11b in the appendix. There, CompactTV still heavily overrepresents the AfD as expected, however, ARD and ZDF now mostly underrepresent the AfD. Furthermore, Die Grünen, which some people claim is favored by the public broadcasters, is actually fairly mentioned, even slightly underrepresented by ZDF. A possible explanation for this shift is criticism raised against ARD, ZDF or the public broadcast in general, the time span of the legislature, which in our dataset, which ends in the early second half of the year 2022, and for that reason the change in government influenced journalism a lot. Finally, further reasonable explanations are administrative responsibilities of politicians, which increases arguably their importance, or a shift in the topics of importance to society or politics in general. Depending on party stances, parties can become more or less relevant over time.

The party Die Linke appears to be heavily underrepresented by almost all media outlets, even those considered to be most left wing. However, this is generally plausible, seeing that Die Linke in the last few years appears to be dealing with several internal conflicts, perhaps influencing their political performance. Aside of the AfD, and to a lesser degree Die Linke, there does not seem to be a major under- or overrepresentation of any party and most media. Going by Fig. 5a, the German media landscape, as a whole, actually manages a fairly balanced representation of all parties. Only Die Linke, and, surprisingly, to a smaller degree, Die Grüne are slightly underrepresented on average across the German media landscape.

Regarding statement bias, on average, mentions of the AfD were judged most negatively by German-News-Sentiment-Bert as can be seen in Table 6 with 54.4 % of mentions being classified as negative and the smallest share of all parties in positive mentions. However, when considering only broadcasts by ARD and ZDF, i.e., Fig. 9, the sentiments are generally more negative for all parties and no party can be singled out, supported by non-significant ANOVA results.

More specifically, the magazine Monitor allegedly reporting positively compared to other magazines about the AfD, with 14,6 % positive mentions, second only to the political talk show Anne Will, and only 45,8 % negative sentiments is an unreasonable result as Monitor is known for a clear stance against the AfD. Two explanations are obvious: for one, the amount of data in this case is very small, see Table 14 and onward. As such, random misclassifications cannot sufficiently even out. Furthermore, the BERT model used cannot be trusted for positive sentiments. From Fig. 6, Fig. 7 and Fig. 8 it is somewhat apparent, that there seems to be a slight tendency towards less positive and more negative mentions of the AfD. However, this was in general not significant and future research has to investigate this point further. As such, at best we find indications for a potential differential treatment of the AfD, but no systematic bias can be deduced from our findings.

Our sentiment analysis, which is in principle similar to approaches by Dallmann et al. [26] and Grefenstette et al. [18], considers the surrounding of a keyword for assessing the sentiment. However, this is problematic, as it is not generally obvious, whether the sentiment is actually targeting the party/keyword itself, or speaks positively about the surrounding. This issue belongs to the so-called target-dependent sentiment analysis [58]. Future work has to consider target-dependent sentiment analysis.

Of great importance for the understanding of this work is that any finding is to be understood relatively, i.e., with respect to other media outlets. No result presented in this work allows to conclude objective, absolute bias in the sense that reporting more positively or negatively, more often or less often about a party is objectively wrong or right. For example, we found that media outlets in our dataset mentioned the AfD party less often with positive sentiment and more often with negative sentiment than other parties. However, that does not mean that there has to be a systemic bias in the media landscape against the AfD, since an objective ground truth would have to be established first to come to that conclusion. Similar considerations apply to the mention analysis as will be shortly discussed.

Legitimate journalistic considerations and audience preference can affect reporting patterns of media outlets without representing a bias with respect to certain parties.

While splitting the subtitle data into chunks of one minute seemed to be a natural approach to the initial imbalance of the dataset, it has the downside of at least occasionally splitting subtitles into separate parts that actually should be considered as a whole. However, splitting the subtitles in a meaningful way automatically is a challenging task and a possible future research topic. Recent advances in natural language processing regarding large language models could allow to considerably improve on this issue.

The issue of splitting the subtitles is closely related to the context of mentions. In our work we analyzed word sequences with a length of 21 words, where the center is a politician or party mention very similar to [26], who used the previous and following four words. However, mentions usually depend considerably on the context, be it a political topic, a corresponding entire paragraph etc. Therefore, for a more indepth analysis, both sentiment and mention analysis would benefit from contexualizing the party and politician mentions. That way, it would be possible to assess whether outlets favored certain party stances depending on the topic. Utilizing large language models like GPT-4 could enable a more nuanced consideration of contextual factors, both at the paragraph and topic level [59].

The presented mention analysis could be trivially adapted to other languages, as only a list of politicians and parties is required. The only limit are automatic speech recognition systems, which nowadays achieve very high quality for many major languages, but typically exhibit a considerably increased word-error-rate for smaller languages like Danish, Urdu or Tamil. The higher the word-error-rate, the less reliable mention analysis will become in general.

For sentiment analysis, a pretrained BERT-like (or another larger language model) is required. To the best knowledge of the authors, pretrained language models for the specific task of news media sentiment analysis are very scarce. As such, adaption of the presented method to other languages is currently difficult. Perhaps transfer-learning techniques could alleviate the need for data for the training of such models.

5.1 Actual Bias versus Reasonable Deviations

Our mention analysis is consistent with the claim that the public broadcasters ARD and ZDF overrepresented the AfD, or, paraphrasing commentators, they gave the AfD too much publicity [60]. We chose to compare the share of mentions to the parties' share of seats in the Bundestag and normalized the result to the AFD's proportion for ease of comparison.

An immediate counterargument to considering solely voting behavior to assess party or politician mentions is the relevancy of party or politician's stances. Perhaps, some party has an opinion only regarding one area of politics and does not care about others. Obviously, their relevancy is drastically reduced in discussions about certain political topics. Similar counterarguments can be made for many different areas. However, in the opinion of the authors, none of the counterarguments invalidates the use of the voting behavior, but rather introduces possible confounders. As such, rather than ignoring voting behavior - this would be an extreme position - one should start with the voting behavior, but control for confounders which have to be agreed on. To actually come to a conclusion regarding "real, objective" media bias, one would have to set up a list of agreed, relevant confounders, control for all confounders and then reassess the party or politician mentions. Then, if one agrees with the confounder list, one has to agree with the outcome of the mention analysis. Note further that, aside of mentions, also, e.g. the number of invitations of party politicians to magazines/talk shows and similar things could be used to assess coverage bias or to assist in doing so. The number of invited politicians would certainly correspond to more or less mentions of the respective party, and as such would also be linked to the total number of party or politician mentions. Thus, mentions likely can be considered a proxy for some latent attitude of the broadcasters, albeit, for definite conclusions, many confounders have to be considered, as explained above. This appears to be a very challenging task, as it is not obvious how one would measure, e.g., relevancy automatically.

Because in Germany, similarly as in e.g. England, public broadcast is obliged by law to represent at least to some extent the general population in their broadcasts, mention analysis, perhaps enriched by politician invites and similar metrics, should be compared to the share of seats in parliament in some way, perhaps also considering the state parliaments.

Similar considerations hold for using sentiment analysis as performed in this work to assess statement bias. The "reasonable observer" has to be equipped with some framework to judge right from wrong, good from bad. These require strong assumptions about morality.

5.2 Ethical Considerations and Possible Applications

The general area of automatic multimedia analysis could impact society and the media landscape in the following ways:

Generative models are already in use for certain types of reports published on media websites. For example, text-to-speech models are used for automation of voice-overs. Possibly, artificial intelligence (AI) will play a greater role in the production of journalism. A key aspect in deploying AI in sensitive fields is their trustworthy- and unbiasedness [61]. Automatic bias analysis can help to detect such biases in AI systems, helping media outlets to improve their quality.

The same holds true today, where media outlets mostly rely on human-created journalism, automatic multimedia bias analysis can be useful as additional quality control. It could serve as an automatic feedback loop, allowing journalists to improve their work.

An automatic multimedia journalism "system" of sufficient quality, that was publically accessible, could also benefit the general public by giving transparent information on, for example, reporting style or difference between individual media outlets depending on topic, party and politician. That way, voters could consume news reports/articles in a more informed manner.

This connects to the ethical component of a sufficiently powerful automatic bias analyzing "system": One could break down bias to the individual level and evaluate individual journalists or editors. Such a break down is a very sensitive measure and clear rules regarding such analyzes should be discussed.

Additionally, if such a system was constantly monitoring all relevant media productions, those participating in the production of media - journalists, politicians and so on - perhaps would feel an intense feeling of scrutiny, like a camera watching them at all times. This could potentially be considered unethical or undesired, as it could affect their freedom of expression. While such a state is still in the distant future, potentials, limits and desirable properties of automatic multimedia bias systems should be discussed in the coming years.

5.3 Limitations

This work did not cover the entirety of the media production (video, audio, web/print) and thus any result has to remain a preliminary observation until further research is performed. But, because we considered most major political formats of ARD and ZDF, the results nonetheless could be the starting point for further, possibly manual, research. Additionally, as mentioned earlier, we did not control for confounders that possibly influence politician and party mentions or the sentiment.

While controlling for such confounders in an automatic fashion is highly desirable, it can also be achieved by combining automatic analysis with a human-performed qualitative content analysis of a smaller subset. Screening subsets of the dataset by hand is highly beneficial to increase the validity of automatic approaches until near perfect analysis tools are developed. Due to this, large scale public multimedia datasets are highly desirable and limited the development of automatic multimedia bias analysis.

The automatically generated subtitles derived from YouTube, while being of good quality on average, still contained some errors which certainly impacted, e.g., the sentiment analysis. However, for the subtitles obtained from the Mediathek, as they are handcrafted, no such impact was observed as their quality was close to perfect.

Transformer-based models like Whisper could improve transcription quality, but suffer from hallucinations, which proved to be detrimental to our approach in additional research performed by us due to, e.g., repeating party names without correspondence to any audio material. Additional engineering effort is required to leverage state-of-the-art speech recognition architectures in a reliable, trustworthy manner.

Additionally, our method of extracting party and politicians from the dataset of subtitles is not perfect. While most extracted mentions were correct, some, especially concerning Die Linke, were false positives and led to errors in further examinations. Besides false positive mentions, there likely are false negatives, which given the current state of the art, are impossible to extract. We hardcoded search terms for our analysis, and since the list of ways a political party could be referred to is possibly large, our method is still limited. In this regard, advances in Named Entity Recognition [62] could be of great value. Furthermore, while apparently agreeing on average in tendency with human judgement, german-news-sentiment-bert is far from being an optimal classifier and clear criteria explaining its judgement are missing. However, it is hoped that errors in the sentiment assessment were random in nature and filtered through averaging across a sufficiently large sample. Another issue of the sentiment analysis, shared with [26], is the possible misattribution of the sentiment.

Even if the considered 21 words of the sentiment analysis (per party mention) are correctly classified with regard to their sentiment, this sentiment might not be aimed at the party being mentioned. While this likely does not represent a systematic error, and thus should average out, future work, together with improved automatic transcripts and sentiment assessment algorithms, has to improve this part of the analysis.

In our work we did not assess differences in mentions on a per politician basis, i.e., we did not assess, whether certain politicians of the same party were more frequently mentioned than others. While certain functions like head of party or head of government will certainly and for good reasons lead to more mentions, evaluating differences between politicians of the same party could reveal whether, e.g., certain lines of reasoning or styles of presentation chosen by certain politicians are favored by media outlets, as they might be more suitable for, say, a talk show format.

6 Conclusion and Future Work

This work investigated statement and coverage bias in German video journalism through automatic analysis of video subtitles. A custom dataset was created covering major German media outlets, including the two major German public broadcasting channels ARD and ZDF. Coverage and statement bias were investigated, where coverage was measured through party and politician mentions and the sentiment was automatically assessed using german-news-sentiment-bert. Manual annotations support the validity of the performed mention analysis but revealed a subpar performance of the german-news-sentiment-bert model, which was found to be unreliable for positive sentiments.

Our data is consistent with the claim that ARD, and to a lesser degree, ZDF overrepresented the AfD party during the 2017-2021 legislature, while hurting Die Linke, compared to which the AfD is overrepresented up to a factor of almost ten. Die Linke seems to be underrepresented by both, party and politician mentions. In the 2021 onward legislature, no such overrepresentation of the AfD was observed.

Possible explanations of these overrepresentations are, for instance, topic-dependent party relevancy. Adding to this, different topics are covered more or less frequently by different media outlets potentially resulting in more or less coverage for certain parties. Future work has to investigate such explanations.

However, in contrast, when considering the entire media landscape that was investigated by us, coverage appears to be very even in the 2017-2021 legislature with respect to the respective shares of seats in the German parliament, with the exception of Die Linke and, to a lesser extend, Die Grüne.

6.1 Future Work

Videos contain way more information than just the audio, from which the subtitles used in this work were derived. Assessing bias through, e.g., suggestive images is a very difficult yet interesting task and currently appears to be out of scope of the state of the art if somewhat reliable results are desired, albeit similar attempts have been published [19]. Furthermore, suggestive audio like, e.g., sad music, could be used to introduce bias but was not investigated in this work. Generally, approaches that encompass the entire political journalism production of a media outlet, i.e. covering video, audio, print, and web journalism, has to be investigated in the future. With recent advancements in speech recognition [63], multimedia journalism comes into the reach of automatic bias analysis algorithms.

An interesting aspect not considered in this work is the time-dependency of media coverage. Political or societal events can have considerable impacts on media coverage on certain topics. For example, media reports in Germany varied greatly throughout the Corona pandemic [15]. Relating media reports to a time-varying political situation is therefore of great importance for a reasonable analysis, especially for longer time frames.

Furthermore, improvements in the reliability of sentiment analysis models seem very desirable for the discipline. In the future, sentiment analysis should also be expanded to encompass more nuanced aspects like sarcasm or other rhetoric devices, resulting in a multi-dimensional sentiment-like analysis. This would allow a fine-grained automatic analysis of media outlet slant, but also style of reporting (e.g. fact-driven vs. emotionally).

Declarations

Author Contributions

RH: Main author, main idea, statistical analysis; HS: Programming, writing; HA: Literature review, proof-reading; DB: Statistical analysis advice, proof-reading; JO: Supervision, proof-reading

Ethics and Consent to Participate Declarations

Not applicable.

Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Funding

This research received no external funding.

Code and Data Availability

We plan on making our code and data available upon acceptance of our manuscript.

Appendix

Table 7: Topics in the dataset as classified by BERTopic, amount of classifications (corresponds to minutes of video subtitle data). Topic labels were assigned manually by the authors based on the entire bag of words making up a topic.

Name	Topic Label	Amount [min]
0_ukraine_russland_russischen_putin english: 0_ukraine_russia_russian_putin	Ukraine War and Refugees	114996
1_deutschland_deutschen_europa_deutsche english: 1_germany_german_europe_germans	Germany-EU Relation	53416
2_tiere_fleisch_wald_tier english: 2_animals_meat_forest_animal	Climate Change and Diet	27137
3_polizei_polizisten_täter_beamten english: 3_police_police_officers_perpetrators_civil servant	Justice, Criminality and Discrimination	25141
...	...	
84_gehörlosen_gehörlose_gebärdensprache_hörenden english: 84_deaf_people_deaf_people_sign_language_hearing_people	Deaf	559
85_kokain_drogen_heroin_dealer english: 85_cocaine_drugs_heroin_dealer	Drug Trade	533
86_bundeswehr_soldaten_soldatinnen_deutsche english: 86_bundeswehr_soldiers_female_soldiers_germans	Bundeswehr/Defence	522
87_cannabis_legalisierung_marihuana_drogen english: 87_cannabis_legalization_marihuana_drugs	Legalization of Cannabis	521
88_abtreibung_schwangerschaftsabbruch_abbrüche english: 88_abortion_termination_of_pregnancy_termination	Abortion	514

Table 8: Search Terms used for and number of hits generated by mention extraction process.

Party	Search Term	Hits
Die Linke	die linke	3072
	linkspartei	1817
Bündnis 90/Die Grünen	die grünen	10225
SPD	spd	28808
FDP	fdp	19041
	freien demokraten	335
CDU	cdu	28157
CSU	csu	10641
AfD	afg	7003
	afd	3920
	alternative für deutschland	335

Table 9: Raw data for party mentions.

party medium	linke	grüne	spd	fdp	cdu	csu	afd
NachDenkSeiten	461	207	1551	350	819	313	572
taz	326	507	688	421	712	113	410
DER SPIEGEL	241	592	1936	1021	1498	764	890
ARD	59	269	574	427	453	138	293
ZDF	1114	1880	5016	3938	5030	1576	2510
Bayerischer Rundfunk	61	123	265	97	81	357	99
ntv Nachrichten	366	729	2041	1175	2097	990	831
faz	210	381	1377	799	1326	353	701
WELT	1230	3404	9049	6931	9184	3621	1702
BILD	550	1702	5589	3524	6089	2140	1388
COMPACTTV	271	431	722	693	868	276	1662

Table 10: Raw data for politician mentions.

party medium	linke	grüne	spd	fdp	cdu	csu	afd
NachDenkSeiten	168	140	361	28	220	25	13
taz	55	139	123	40	93	13	7
DER SPIEGEL	58	347	526	201	508	70	56
ARD	33	183	437	95	296	22	21
ZDF	222	1432	3247	695	3395	184	90
Bayerischer Rundfunk	8	41	36	8	30	32	15
ntv Nachrichten	53	413	1112	219	1300	86	55
faz	40	307	569	97	441	85	22
WELT	343	4813	9618	1723	5786	509	89
BILD	121	1885	5337	1409	5874	302	49
COMPACTTV	86	88	123	27	92	2	62

Table 11: Raw data for party mentions with positive sentiment.

party medium	linke	grüne	spd	fdp	cdu	csu	afd
NachDenkSeiten	35	20	119	24	55	24	38
taz	38	73	103	61	83	14	33
DER SPIEGEL	36	117	316	180	238	118	91
ARD	3	24	58	42	44	9	26
ZDF	125	287	807	559	706	252	244
Bayerischer Rundfunk	6	22	44	17	5	45	10
ntv Nachrichten	67	125	391	214	401	246	104
faz	16	54	167	92	136	41	54
WELT	152	526	1446	908	1636	687	168
BILD	85	262	813	540	928	415	130
COMPACTTV	25	44	61	68	70	25	197

Table 12: Raw data for party mentions with neutral sentiment.

party medium	linke	grüne	spd	fdp	cdu	csu	afd
NachDenkSeiten	196	83	692	153	392	146	226
taz	120	165	214	125	251	46	139
DER SPIEGEL	86	224	659	353	535	291	339
ARD	17	87	176	150	142	54	77
ZDF	411	633	1673	1313	1889	624	876
Bayerischer Rundfunk	27	44	79	34	32	139	31
ntv Nachrichten	166	268	706	443	830	411	335
faz	98	162	583	342	656	179	277
WELT	464	1269	3137	2638	3819	1487	669
BILD	199	578	1778	1250	2411	807	477
COMPACTTV	104	147	246	241	279	100	643

Table 13: Raw data for party mentions with negative sentiment.

party medium	linke	grüne	spd	fdp	cdu	csu	afd
NachDenkSeiten	230	104	740	173	372	143	308
taz	168	269	371	235	378	53	238
DER SPIEGEL	119	251	961	488	725	355	460
ARD	39	158	340	235	267	75	190
ZDF	578	960	2536	2066	2435	700	1390
Bayerischer Rundfunk	28	57	142	46	44	173	58
ntv Nachrichten	133	336	944	518	866	333	392
faz	96	165	627	365	534	133	370
WELT	614	1609	4466	3385	3729	1447	865
BILD	266	862	2998	1734	2750	918	781
COMPACTTV	142	240	415	384	519	151	822

Table 14: Raw data for party mentions with positive sentiment by public broadcast.

party format	linke	grüne	spd	fdp	cdu	csu	afd
ARD: Monitor	0	1	5	1	6	3	7
ARD: Hart aber fair	0	6	13	9	5	0	0
ARD: Maischberger	1	7	13	10	12	5	0
ARD: Tagesthemen	0	1	2	6	9	0	2
ARD: Anne Will	2	5	21	15	9	0	3
ARD: YouTube	0	4	4	1	3	1	14
ZDF: Maybrit Illner	1	7	16	9	18	8	1
ZDF: Markus Lanz	11	26	47	33	60	21	10
ZDF: frontal	1	4	4	5	9	4	5
ZDF: YouTube	112	250	740	512	619	219	228
Non Public Broadcast	460	1243	3460	2104	3552	1615	825

Table 15: Raw data for party mentions with neutral sentiment by public broadcast.

party format	linke	grüne	spd	fdp	cdu	csu	afd
ARD: Monitor	1	12	15	4	20	10	19
ARD: Hart aber fair	7	17	27	39	16	7	5
ARD: Maischberger	3	22	46	30	31	11	8
ARD: Tagesthemen	1	11	18	22	15	1	6
ARD: Anne Will	5	22	50	39	24	14	2
ARD: YouTube	0	3	20	16	36	11	37
ZDF: Maybrit Illner	5	17	47	35	37	13	2
ZDF: Markus Lanz	30	41	109	77	156	59	18
ZDF: frontal	0	9	16	13	40	18	30
ZDF: YouTube	376	566	1501	1188	1656	534	826
Non Public Broadcast	1460	2940	8094	5579	9205	3606	3136

Table 16: Raw data for party mentions with negative sentiment by public broadcast.

party format	linke	grüne	spd	fdp	cdu	csu	afd
ARD: Monitor	0	18	27	6	45	18	22
ARD: Hart aber fair	4	35	53	49	26	9	10
ARD: Maischberger	16	35	86	49	51	22	15
ARD: Tagesthemen	3	14	30	31	33	8	23
ARD: Anne Will	7	37	95	63	49	6	14
ARD: YouTube	9	19	49	37	63	12	106
ZDF: Maybrit Illner	9	49	86	79	87	15	6
ZDF: Markus Lanz	53	122	301	240	293	71	84
ZDF: frontal	5	27	55	29	45	17	75
ZDF: YouTube	511	762	2094	1718	2010	597	1225
Non Public Broadcast	1796	3893	11664	7328	9917	3706	4294

Fig. 10: Results of our unique mention analysis. Mentions of parties or politicians were only counted once per video. (a) Percentage of all unique mentions of a given party in the respective medium. CDU and SPD are mentioned most frequently. (b) Same data but with columnwise zero mean. (c) Percentage of all unique mentions of politicians of a given party in the respective medium. CDU and SPD are mentions most frequently, with politicians of Die Linke, FDP, CSU and AfD rarely getting mentioned. (d) Same data but with columnwise zero mean.

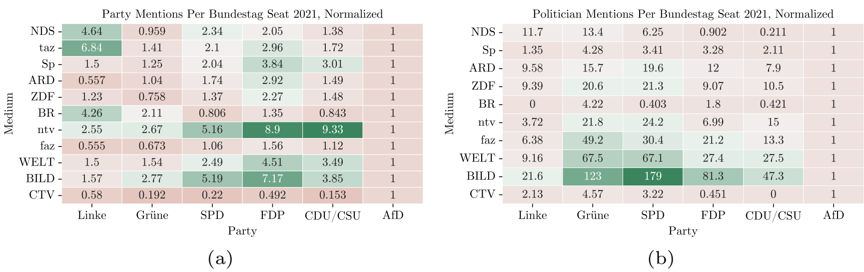


Fig. 11: (a) Party and (b) politician mentions per Bundestag seat by party and medium in the 2021 legislature. Each row is normalized to the value of AfD. Since data for this specific case was too scarce, we decided not to include it in our analysis.

References

[1] Müller, H.: Die Vierte Gewalt – Medien und Journalismus Unter die Lupe Genommen. Eckartschriften. Österreichische Landsmannschaft, Vienna, Austria (2008). ISBN: 3902350261

[2] D'Alessio, D., Allen, M.: Media bias in presidential elections: A meta-analysis. Journal of Communication **50**(4), 133–156 (2000). https://doi.org/10.1111/j.1460-2466.2000.tb02866.x

[3] Maurer, M., Jost, P., Haßler, J., Kruschinski, S.: Auf den spuren der lügenpresse. Publizistik **64**(1), 15–35 (2018). https://doi.org/10.1007/s11616-018-00466-y

[4] Felix Hamborg, Karsten Donnay, Bela Gipp: Automated identification of media bias in news articles: an interdisciplinary literature review. International Journal on Digital Libraries **20**(4), 391–415 (2018). https://doi.org/10.1007/s00799-018-0261-y

[5] Medienanstalten, D.: Medienvielfaltsmonitor 2021. https://www.die-medienanstalten.de/themen/forschung/medienvielfaltsmonitor. Last Accessed: 24.03.2025 (2021)

[6] Deutschland, B.: Medienstaatsvertrag. https://www.die-medienanstalten.de/fileadmin/user_upload/Rechtsgrundlagen/Gesetze_Staatsvertraege/Medienstaatsvertrag_MStV.pdf. 3. Abschnitt, §26, Abs. 2. Last Accessed: 24.03.2025 (2020)

[7] Lünenborg, M., Berghofer, S.: Politikjournalistinnen und -journalisten. Fachjournalist (2010). Last Accessed: 24.03.2025

[8] Eisenegger, M., Oehmer, F., Udris, L., Vogler, D.: In: Öffentlichkeit und Gesellschaft, F. (ed.) Die Qualität der Medienberichterstattung zur Corona-Pandemie, pp. 29–50. Schwabe, ??? (2020). https://doi.org/10.5167/uzh-192622

[9] Maurer, M., Haßler, J., Kruschinski, S., Jost, P.: Looking over the channel: The balance of media coverage about the "refugee crisis" in germany and the uk. Communications **47**(2), 219–240 (2022). https://doi.org/10.1515/commun-2020-0016

[10] Newman, N., Fletcher, R., Robertson, C.T., Nielsen, R.K.: Reuters Institute Digital News Report 2022. Reuters Institute for the Study of Journalism, Oxford, England (2022)

[11] Williams, A.: Unbiased study of television news bias. Journal of Communication **25**(4), 190–199 (1975). https://doi.org/10.1111/j.1460-2466.1975.tb00656.x

[12] Patterson, T.E.: News Coverage of Donald Trump's First 100 Days. https://research.hks.harvard.edu/publications/getFile.aspx?Id=1590. Last Accessed: 24.03.2025 (2017)

[13] Gearhart, S., Moe, A., Zhang, B.: Hostile media bias on social media: Testing the effect of user comments on perceptions of news bias and credibility. Human Behavior and Emerging Technologies **2**(2), 140–148 (2020) https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbe2.185. https://doi.org/10.1002/hbe2.185

[14] McCarthy, J.D., Mcphail, C., Smith, J.: Images of protest: Dimensions of selection bias in media coverage of washington demonstrations, 1982 and 1991. American Sociological Review **61**, 478–499 (1996). https://doi.org/10.2307/2096360

[15] Maurer, M., Reinemann, C., Kruschinski, S.: Einseitig, unkritisch, regierungsnah? Eine empirische Studie zur Qualität der journalistischen Berichterstattung über die Corona-Pandemie. Rudolf Augstein Stiftung (2021)

[16] Shultziner, D., Stukalin, Y.: Politicizing what's news: How partisan media bias occurs in news production. Mass Communication and Society **24**(3), 372–393 (2020). https://doi.org/10.1080/15205436.2020.1812083

[17] Shultziner, D., Stukalin, Y.: Appendix b instructions for coders in research on partisan media coverage of the social justice protest movement in israel (2020). https://doi.org/10.13140/RG.2.2.22767.66721

[18] Grefenstette, G., Qu, Y., Shanahan, J.G., Evans, D.A.: Coupling Niche Browsers and Affect Analysis for an Opinion Mining Application. In: RIAO (2004)

[19] Peng, Y.: Same candidates, different faces: Uncovering media bias in visual portrayals of presidential candidates with computer vision. Journal of Communication **68**(5), 920–941 (2018). https://doi.org/10.1093/joc/jqy041

[20] Kim, E., Lelkes, Y., McCrain, J.: Measuring dynamic media bias. Proceedings of the National Academy of Sciences **119**(32), 2202197119 (2022) https://www.pnas.org/doi/pdf/10.1073/pnas.2202197119. https://doi.org/10.1073/pnas.2202197119

[21] Standford: Standford Cable TV News Analyzer. https://tvnews.stanford.edu/. Last Accessed: 24.03.2025 (2021)

[22] Hong, J., Crichton, W., Zhang, H., Fu, D.Y., Ritchie, J., Barenholtz, J., Hannel, B., Yao, X., Murray, M., Moriba, G., Agrawala, M., Fatahalian,

K.: Analysis of faces in a decade of us cable tv news. In: Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Association for Computing Machinery, ??? (2021)

[23] Congleton, C., van der Putten, P., Verberne, S.: Tracing political positioning of dutch newspapers. In: Spezzano, F., Amaral, A., Ceolin, D., Fazio, L., Serra, E. (eds.) Disinformation in Open Online Media, pp. 27–43. Springer, Cham (2022)

[24] Garz, M., Sörensen, J., Stone, D.F.: Partisan selective engagement: Evidence from facebook. Journal of Economic Behavior & Organization **177**, 91–108 (2020). https://doi.org/10.1016/j.jebo.2020.06.016

[25] Dewenter, R., Dulleck, U., Thomas, T.: Does the 4th estate deliver? the political coverage index and its application to media capture. Constitutional Political Economy **31**(3), 292–328 (2020). https://doi.org/10.1007/s10602-019-09291-5

[26] Dallmann, A., Lemmerich, F., Zoller, D., Hotho, A.: Media Bias in German Online Newspapers. In: Proceedings of the 26th ACM Conference on Social Media 15. ACM Press, New York, NY, USA (2015). https://doi.org/10.1145%2F2700171.2791057

[27] Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., Kappas, A.: Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology **61**(12), 2544–2558 (2010) https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.21416. https://doi.org/10.1002/asi.21416

[28] Lüdke, S., Grau, J., Drawitsch, M.: german-news-sentiment-bert. https://huggingface.co/mdraw/german-news-sentiment-bert. Last Accessed: 24.03.2025 (2021)

[29] Lees, C.: The 'alternative for germany': The rise of right-wing populism at the heart of europe. Politics **38**(3), 295–310 (2018). https://doi.org/10.1177/0263395718777718

[30] Swiss Policy Research: Medien-Navigator. https://swprs.org/medien-navigator/. Last Accessed: 24.03.2025 (2021)

[31] Zandt, D.V.: Media Bias/Fact Check. https://mediabiasfactcheck.com/. Last Accessed: 24.03.2025 (2015)

[32] Busch-Janser, S.: Die deutsche Medienlandschaft in einer Karte . https://www.polisphere.eu/blog/die-deutsche-medienlandschaft-in-einer-karte/. Last Accessed (webarchive): 24.03.2025 (2017)

[33] YouTube LLC: YouTube. https://www.youtube.com. Last Accessed: 24.03.2025

[34] ARD: ARD Mediathek. https://www.ardmediathek.de. Last Accessed: 24.03.2025

[35] ZDF: ZDF Mediathek. https://www.zdf.de. Last Accessed: 24.03.2025

[36] ARD: ANNE WILL. https://daserste.ndr.de/annewill/. Last Accessed (webarchive): 24.03.2025

[37] ARD: hart aber fair. https://www1.wdr.de/daserste/hartaberfair/index.html. Last Accessed: 24.03.2025

[38] ARD: Maischberger. https://www.daserste.de/information/talk/maischberger/maischberger-index100.html. Last Accessed: 24.03.2025

[39] ZDF: Markus Lanz. https://www.zdf.de/gesellschaft/markus-lanz. Last Accessed: 24.03.2025

[40] ZDF: maybrit illner. https://www.zdf.de/politik/maybrit-illner. Last Accessed: 24.03.2025

[41] ARD: Tagesthemen. https://www.ardmediathek.de/sendung/tagesthemen/Y3JpZDovL2Rhc2Vyc3RlLmRlL3RhZ2VzdGhlbWVu. Last Accessed: 24.03.2025

[42] ZDF: frontal. https://www.zdf.de/politik/frontal/. Last Accessed: 24.03.2025

[43] ARD: Monitor. https://www1.wdr.de/daserste/monitor/index.html. Last Accessed: 24.03.2025

[44] Finkhäuser, A., Wiegandt, N.: MediathekView. https://mediathekview.de. Last Accessed: 24.03.2025 (2008)

[45] Saini, H.K.: YouTube-Search-Python. https://pypi.org/project/youtube-search-python/. Last Accessed: 24.03.2025 (2020)

[46] Depoix, J.: YouTube-Transcript-API. https://pypi.org/project/youtube-transcript-api/. Last Accessed: 24.03.2025 (2020)

[47] Bundestag, D.: Deutscher Bundestag - Abgeordnete. https://www.bundestag.de/abgeordnete. Last Accessed: 24.01.2023

[48] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv: 1810.04805 (2018)

[49] Tymann, K., Lutz, M., Palsbröker, P., Gips, C.: Gervader - a german adaptation of the vader sentiment analysis tool for social media texts. In: Lernen, Wissen, Daten, Analysen (2019). https://api.semanticscholar.org/CorpusID:202760140

[50] Na, J.-C., Kyaing, W.: Sentiment analysis of user-generated content on drug review websites. Journal of Information Science Theory and Practice **3**, 6–23 (2015). https://doi.org/10.1633/JISTaP.2015.3.1.1

[51] Landis, J., Koch, G.: The measurement of observer agreement for categorical data. Biometrics **33**, 159–74 (1977). https://doi.org/10.2307/2529310

[52] Lüdke, S., Grau, J., Drawitsch, M.: News Sentiment Development on the Example of 'Migration' . https://github.com/text-analytics-20/news-sentiment-development/blob/main/news-sentiment-report-public.pdf. Last Accessed: 24.03.2025 (2022)

[53] Tagesschau: Sitzplätze im Bundestag. https://www.tagesschau.de/inland/btw17/bundestag-sitzverteilung-101.html. Last Accessed (webarchive): 24.03.2025 (2017)

[54] Bundesregierung: Bundeskanzlerin Angela Merkel 2005-2021. https://www.bundesregierung.de/breg-de/suche/bundeskanzlerin-angela-merkel-2005-2021-2045938. Last Accessed: 24.03.2025 (2022)

[55] Bundestag: Deutscher Bundestag - So setzt sich die neue Bundesregierung zusammen. https://www.bundestag.de/dokumente/textarchiv/2021/kw49-de-kanzlerwahl-bundesregierung-870146. Last Accessed: 24.03.2025 (2021)

[56] Deutschlandfunk: AfD und Medien: 'Für alle Parteien müssen die gleichen Standards gelten' . https://www.deutschlandfunk.de/afd-und-medien-fuer-alle-parteien-muessen-die-gleichen-100.html. Last Accessed (webarchive): 31.08.2023 (2017)

[57] Heyen, S.: Die AfD in Den Medien : Eine Analyse des Medialen Framings der Partei "Alternative Für Deutschland" Am Beispiel Politischer Talkshows. Carl Grossmann, ??? (2020). https://doi.org/10.24921/2020.94115952

[58] Hamborg, F., Donnay, K., Gipp, B.: Towards target-dependent sentiment classification in news articles. In: Diversity, Divergence, Dialogue, pp. 156–166. Springer, ??? (2021). https://doi.org/10.1007/978-3-030-71305-8_12

[59] Pham, C.M., Hoyle, A., Sun, S., Resnik, P., Iyyer, M.: TopicGPT: A

prompt-based topic modeling framework. In: Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), pp. 2956–2984. Association for Computational Linguistics, ??? (2024). https://doi.org/10.18653/v1/2024.naacl-long.164

[60] Huber, J.: AfD - Medien - Gesellschaft: 'Wer nur entlarven will, verstärkt die Einigelung' . https://www.tagesspiegel.de/gesellschaft/medien/wer-nur-entlarven-will-verstarkt-die-einigelung-6306778.html. Last Accessed (webarchive): 20.03.2025 (2017)

[61] Zhou, M., Abhishek, V., Derdenger, T., Kim, J., Srinivasan, K.: Bias in Generative AI (2024). https://arxiv.org/abs/2403.02726

[62] Li, J., Sun, A., Han, J., Li, C.: A survey on deep learning for named entity recognition. IEEE Transactions on Knowledge and Data Engineering **34**(1), 50–70 (2022). https://doi.org/10.1109/TKDE.2020.2981314

[63] Radford, A., Kim, J.W., Xu, T., Brockman, G., McLeavey, C., Sutskever, I.: Robust Speech Recognition via Large-Scale Weak Supervision. arXiv (2022). https://doi.org/10.48550/ARXIV.2212.04356. https://arxiv.org/abs/2212.04356