A Survey on Predicting the Factuality and the Bias of News Media

Preslav Nakov[1] **Jisun An**[2] **Haewoon Kwak**[2] **Muhammad Arslan Manzoor**[1]
Zain Muhammad Mujahid[1] **Husrev Taha Sencar**[3]
[1] MBZUAI [2] Indiana University Bloomington
[3] Qatar Computing Research Institute, HBKU
{preslav.nakov}@mbzuai.ac.ae

Abstract

The present level of proliferation of fake, biased, and propagandistic content online has made it impossible to fact-check every single suspicious claim or article, either manually or automatically. An increasing number of scholars are focusing on a coarser granularity, aiming to profile entire news outlets, which allows fast identification of potential "fake news" by checking the reliability of their source. Source factuality is also an important element of systems for automatic fact-checking and "fake news" detection, as they need to assess the reliability of the evidence they retrieve online. Political bias detection, which in the Western political landscape is about predicting left-center-right bias, is an equally important topic, which has experienced a similar shift toward profiling entire news outlets. Moreover, there is a clear connection between the two, as highly biased media are less likely to be factual; yet, the two problems have been addressed separately. In this survey, we review the state of the art on media profiling for factuality and bias, arguing for the need to model them jointly. We also shed light on some of the major challenges for modeling bias and factuality jointly. We further discuss interesting recent advances in using different information sources and modalities, which go beyond the text of the articles the target news outlet has published. Finally, we discuss current challenges and outline future research directions.

1 Introduction

The rise of the web has made it possible for anybody to create a website and to become a *news medium*. This was a hugely positive development as it elevated freedom of expression to a whole new level, allowing anybody to have their voice heard. With the subsequent rise of social media, anybody could potentially reach out to a vast audience, something that until recently was only possible for major news outlets. One of the consequences was a *trust crisis*: with traditional news media stripped of their gatekeeping role, society was left unprotected against potential manipulation.

In an attempt to solve the trust problem, several initiatives, such as PolitiFact[1], Snopes[2], FactCheck[3], and Full Fact[4], have been launched to fact-check suspicious claims manually. However, given the scale of the proliferation of false information online, it was unfeasible to fact-check every single suspicious claim, even when this was done automatically (Pérez-Rosas et al., 2017), not only for computational reasons but also due to timing. In order to fact-check a claim manually or automatically, it is required to verify the stance of mainstream media concerning that claim and/or the reaction of users on social media. Accumulating this evidence takes time, and delay means more potential sharing of the malicious content (Zhou, 2021; Liu et al., 2022).

Therefore, a much more promising alternative is to profile the medium that initially published the news article with the suspicious claim. Since media that have published fake or biased content in the past are more likely to do so in the future, profiling media in advance makes it possible to detect likely "fake news" the moment it is published by simply checking the reliability of its source (Baly et al., 2020b; Mehta et al., 2022; Panayotov et al., 2022). Factuality labels at the media level can also be used for distant supervision, labeling all their articles with the medium's label. This approach is frequently used for "fake news" detection where manually annotating large datasets can be challenging (Nørregaard et al., 2019; Spinde et al., 2022).

Estimating news source reliability is important for claim fact-checking (Nguyen et al., 2018), and it also gives an important prior when solving article-

[1] https://www.gdeltproject.org/
[2] https://www.snopes.com/
[3] https://www.factcheck.org/
[4] https://fullfact.org/

Findings of the Association for Computational Linguistics: ACL 2024, pages 15947–15962
August 11-16, 2024 ©2024 Association for Computational Linguistics

level tasks such as “fake news” and click-bait detection (Hardalov et al., 2016; Karadzhov et al., 2017a; De Sarkar et al., 2018; Pérez-Rosas et al., 2018; Brill, 2001; Finberg et al., 2002; Pan et al., 2018; Nguyen et al., 2020). Moreover, the CLEF-2023 CheckThat! lab had a recent shared task on predicting the factuality and the bias of news media (Barrón-Cedeño et al., 2023; Nakov et al., 2023; Martino et al., 2023; Alam et al., 2023).

Recently, Vallejo et al. (2023) conducted a comprehensive examination of research methodologies, exploring the influence of media framing on public opinions, political dynamics, and policies. They stress the importance of understanding media bias and framing to enhance media literacy, critical thinking, and informed decision-making. Several surveys on fake news (Shu et al., 2017; da Silva et al., 2019; Zhou, 2021; Mridha et al., 2021; Mishra et al., 2022), mis/dis-information (Islam et al., 2020; Alam et al., 2022; Hardalov et al., 2022a), fact-checking (Thorne and Vlachos, 2018; Kotonya and Toni, 2020; Nakov et al., 2021; Guo et al., 2022b), truth discovery (Li et al., 2016; Xu et al., 2021; Ahmed et al., 2022), and propaganda detection (Martino et al., 2020; Chaudhari and Pawar, 2021) have been conducted. However, they have focused on claims or articles, while here we survey research on profiling entire news outlets. We delve into the capacity of large language models (LLMs) to estimate the credibility and political leaning of sources through carefully crafted questions. We also provide a holistic perspective on the use of LLMs in the context of media profiling.

2 Factuality

Veracity of information has been studied at different levels: (*i*) claim-level (e.g., *fact-checking*), (*ii*) article-level (e.g., *“fake news” detection*), (*iii*) user-level (e.g., *hunting for trolls*), and (*iv*) medium-level (e.g., *source reliability estimation*). Our primary interest here is in the latter. At the claim-level, significant effort has been paid to fact-checking and rumor detection using information from social media, i.e., how users reply to the claim (Canini et al., 2011; Castillo et al., 2011; Ma et al., 2015, 2016; Zubiaga et al., 2015; Ma et al., 2017; Dungs et al., 2018; Kochkina et al., 2018; Lim et al., 2020; Hardalov et al., 2022b; Nguyen et al., 2020), yet, the need for more nuanced methodologies that concentrate on a multitude of characteristics is paramount (Thorne and Vlachos, 2018; Guo et al., 2022b). A set of web pages and snippets from search engines have also been used as a source of information (Mukherjee and Weikum, 2015; Popat et al., 2016, 2017; Karadzhov et al., 2017b; Mihaylova et al., 2018; Baly et al., 2018b). In either case, the most important information for the claim-level tasks are *stance* (does a tweet or a news article agree or disagree with the claim?) and *source reliability* (do we trust the user who posted the tweet or the medium that published the news article?). The problem of source reliability remains largely under-explored. In the case of social media and community fora, it concerns modeling the user, e.g., there has been research on finding opinion manipulation *trolls* (Mihaylov and Nakov, 2016), *sockpuppets* (Maity et al., 2017), *Internet water army* (Chen et al., 2013), and *seminar users* (Darwish et al.). In the case of the Web, it is about source trustworthiness (the URL domain, the medium).

In early work, the source reliability of news media has often been estimated automatically based on the general stance of the target medium with respect to known true/false claims without access to gold labels about the overall medium-level factuality of reporting (Dong et al., 2015; Mukherjee and Weikum, 2015; Popat et al., 2016, 2017; Popat et al., 2018).

More recent work has addressed the task as one on its own right. Baly et al. (2018a) used gold labels from Media Bias/Fact Check (MBFC)[5], and rich information sources available before disinformation campaign begins: articles published by the medium, what is said about it on Wikipedia, metadata from its Twitter profile, URL structure, and traffic information to characterize the media. In follow-up work, Baly et al. (2019) uncovered the ordinal relationship between media bias and factuality using a multi-task ordinal regression setup, as detailed in Section 4. Then, Baly et al. (2020b) considered the social context, extended the information sources to include Facebook followers and speech signals from the news medium’s channel on YouTube (if any). Hounsel et al. (2020) focused on infrastructure features such as domain registrations, TLS/SSL certificates, and web hosting configurations. They posit that these features could potentially reveal significant disparities between disinformation and authentic news websites, prior to content dissemination.

[5]http://mediabiasfactcheck.com

Mehta et al. (2022) leveraged the concept of *communities* to connect users, the content they prefer, and with the source which provide that content. They approached the task as a problem of reasoning over relationships between sources, their published articles, and the engagement patterns of users on social media within a graph framework. Panayotov et al. (2022) proposed a graph-based framework to profile news media outlets, with nodes representing the outlets and the connecting edges indicating audience overlap. Taking inspiration from homophily considerations, they constructed a network reflecting the hypothesis that similar users consume similar types of media.

Large language models (LLMs) can also be used to estimate source reliability as they can capture knowledge (Qin et al., 2023). Yang and Menczer (2023) showed that with well-crafted instructions, ChatGPT showcases its capability to provide credibility ratings for an extensive spectrum of news outlets. Their results showed that these ratings correlate with those from human experts and that LLMs could be an affordable reference for credibility ratings in media profiling applications. Mehta and Goldwasser (2023) introduced an innovative interactive framework for news media profiling, combining the capabilities of graph-based models, pre-trained LLMs, and human expertise to delve into the social context of news on social media platforms. Their findings highlighted the framework's ability to rapidly identify fake news media, even in the challenging terrain of emerging news events, with as few as five human interactions.

3 Bias

Compared to factuality, which can be objectively determined by whether a piece of information is true or not, media bias has more complex dimensions. For the last few decades, many scholars have conceptualized media bias in different ways. For instance, a bias can be defined as "imbalance or inequality of coverage rather than as a departure from truth" (Stevenson et al., 1973). A departure from truth, however, can be measured only when an accurate record of the event is available (e.g., trial transcript and reporting).

A different definition, namely "any systematic slant favoring one candidate or ideology over another" (Waldman and Devitt, 1998), is proposed to capture various dimensions rather than coverage imbalance, such as favorability conveyed in visual representations (i.e., news photos). E.g., smiling, speaking at the podium, cheering crowd, and eye-level shots are preferred over frowning, sitting, being alone, and shots from above, respectively. Guo et al. (2022a) utilized pre-trained BERT (Devlin et al., 2019) models, fine-tuned on news articles from various media outlets, to capture linguistic biases through masked language modeling. To validate their model's ability to detect media bias, they compare its results with established news bias datasets from sources like Pew Research and `allsides.com`

D'Alessio and Allen (2000) reviewed 59 studies about partisan media bias in presidential elections. They proposed to categorize media bias into the following three types: (*i*) *gatekeeping bias*, where editors and journalists "determine" which content reaches the audience within various forms of media (Smith et al., 2001), (*ii*) *coverage bias*, where the amount of news coverage (e.g., the length of newspapers articles, or the time given on television) each party receives is systematically biased to one party at the expense of the other one (Hassell et al., 2020), and (*iii*) *statement bias*, where news media interject their attitudes or opinions in the news reporting. Groeling (2013) proposed a more relaxed concept of media bias, which is "a portrayal of reality that is significantly and systematically (not randomly) distorted," to take a variety of media bias dimensions into account. In particular, he focused on two main forms of media bias—*selection bias* (i.e., what to cover) and *presentation bias* (i.e., how to cover it)—driven by the choices of newsmakers.

Selection bias has been studied in various ways, including qualitative interviews or surveys of journalists and editors about the decision-making process they use to select the stories in their newsroom (Tandoc Jr, 2014). Here, news selection is not necessarily confined to the political context. News reporting about any news items can be considered as a unit of analysis.

Data-driven research on selection bias commonly follows three steps: (*i*) collect news articles (for newspapers or online news) or transcripts (for TV news) for a target period, (*ii*) conduct content analysis to find the news coverage of politicians, parties, or events. Optionally, study the tone of the news articles (e.g., negative news are more frequently reported) (Soroka, 2012), and (*iii*) identify systematic biases by comparing news coverage. An exhaustive database of news stories is thus essential for selection bias research. While commercial databases, such as Lexis Nexis, have been widely used (Soroka, 2012; Padgett et al., 2019; Gilens and Hertzman, 2000; Boykoff and Boykoff, 2004), publicly available datasets, such as GDELT, start to get attention (Boudemagh and Moise, 2017; Kwak and An, 2014) and are getting validated by comparing multiple sources (Kwak and An, 2016; Weaver and Bimber, 2008). The availability of such datasets also enables researchers to compare news coverage across countries (Guo and Vargo, 2017; Kwak et al., 2018; Litterer et al., 2023).

Presentation bias has been characterized from diverse perspectives, including framing (Entman, 2007), visuals (Barrett and Barrington, 2005), sources (Baum and Groeling, 2008), tone (Soroka, 2012), and more. Particularly, framing bias has been actively studied in many disciplines.

Framing Bias refers to a bias that highlights a certain aspect of an event or an issue more than the others (Entman, 1993). Emphasizing a particular aspect can deliver a distorted view toward the issue even without the use of biased expressions.

Framing biases have been typically studied at issue level (Kim and Johnson, 2022). Researchers collect news articles about an issue or an event, conduct manual content analysis, and build a frame detection model (Baumer et al., 2015). Open-source tools to help the analysis have been proposed (Bhatia et al., 2021; Morstatter et al., 2018). While this approach can characterizes diverse frames, it is not trivial to compare framing across issues.

The Media Frames Corpus (MFC) was proposed to address this limitation (Card et al., 2015). It contains articles annotated with 15 generic frames (including *others*) across three policy issues. Several studies have demonstrated reasonable prediction performance of the general media frames with different datasets (Field et al., 2018; Kwak et al., 2020). These 15 general frames were also used for analyzing political discourse on social media (Johnson et al., 2017). These frames are often customized to a specific issue by adding issue-specific frames (Liu et al., 2019), even though doing so somewhat contradicts the original motivation of general media frames, namely to be able to compare frames across various issues.

News slant was proposed to characterize how framing in news reports favors one side over the other (Entman, 2007). The media-level slant thus could differ across issues (Ganguly et al., 2020).

A variety of methods have been proposed to quantify the extent of news slant in traditional news media by (*i*) linking media outlets to politicians with known political positions, (*ii*) directly analyzing news content, and (*iii*) using shared audience among media outlets. Groseclose and Milyo (2005) assigned an ADA (Americans for Democratic Action) score for each media outlet by investigating co-citations of think-tanks by members of Congress and media outlets. Gentzkow and Shapiro (2010) proposed an ideological slant index of news media in a seminal study. The news slant is measured by the extent of phrases in news coverage that are more frequently used by one political party (i.e., Democratic or Republican) congress members than by another one in the 2005 Congress Record. Their frequency-based approach successfully finds politically charged phrases such as *death tax* or *war on terror* by Republicans and associated media and *estate tax* or *war in Iraq* by Democrats and associated media, and they further computed media slant index for 433 newspapers. The choice of words by political party members and news media is considered framing because they purposely highlight some aspects of the issue over others.

An et al. (2011, 2012) proposed a method to compute media slant scores by measuring distances between media sources by their mutual followers on Twitter. Stefanov et al. (2020) identified the political leanings of media outlets and influential people on Twitter based on their stance on controversial topics. They built clusters of users around core vocal ones based on their behavior on Twitter, such as retweeting, using a procedure proposed in (Darwish et al., 2019).

Left-center-right bias (or left-right bias) was studied based on media-level annotation from specialized online platforms, such as News Guard, AllSides, and Media Bias/Fact Check, where journalists use carefully designed guidelines to make judgments. Researchers have then trained systems to predict this bias using a variety of information sources such as analyzing the corresponding YouTube channels (Dinkov et al., 2019), and using information from the articles the target news outlet has published, what is there about them in social media and in Wikipedia (Baly et al., 2020b).

There has also been work on predicting the left-center-right bias of articles, which is somewhat relevant here as it can be an element of media-level analysis. Such systems are typically trained using distant supervision, projecting the label from a medium to each article from that medium, which is an easy way to obtain large datasets, needed to train contemporary deep learning models. For example, Kulkarni et al. (2018) used site-level annotations from the AllSides website for political bias detection. The same approach was used to study hyperpartisanship, i.e., extremely one-sided reporting (Potthast et al., 2018), as part SemEval-2019 task 4 on Hyper-partisan News Detection (Kiesel et al., 2019). Recent studies have underscored the potential problems of inaccuracy and noise linked to distant supervision. This has led to the development of a left-center-right bias dataset, which features careful manual annotations at the article level (Baly et al., 2020a).

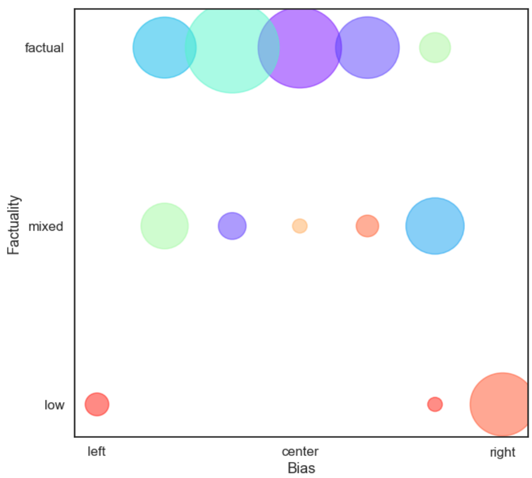

Figure 1: Correlation between bias and factuality for the news outlets in the Media Bias/Fact Check website.

4 Joint Modeling

Figure 1 shows the correlation between the factuality and the bias of the news outlets analyzed by MBFC. We can see that extreme-left and extreme-right websites are not factual, and that center media are most factual overall[6]. Moreover, some of the datasets used for the two tasks have media-level annotations for both factuality and bias. Thus, it makes sense to model factuality and bias jointly. Yet, joint modeling of the two tasks remains severely underexplored. In fact, there has been a single attempt at doing so to date: (Baly et al., 2019) proposed a multi-task learning formulation based on the copula ordinal regression framework (Walecki et al., 2016), which jointly predicts factuality and bias on ordinal scales. They further used several auxiliary tasks, modeling centrality, hyper-partisanship, as well as left-vs.-right bias on a coarse-grained scale. They further took into account the ordinal nature of the labels for both factuality *(high, mix, low)* and bias *(ext. left, left, center left, center, center right, right, ext. right)* tasks, noting that classifying an *extreme right* medium as *extreme-left* is a huge error, while classifying it as a *center* is a smaller one, and predicting *right* is an even smaller error. The chart[7] where Reuters[8] and AirForceTimes[9] are center-bias and exhibit high factuality.

5 Basis of Prediction

5.1 Textual Content

Linguistic Features focus on language use, and they have been shown to be useful for detecting fake articles, as well as for predicting the political bias and the factuality of reporting of news media (Horne et al., 2018; Baly et al., 2018a). For example, Horne and Adali (2017) showed that "fake news" pack a lot of information in the title (as many people do not read beyond the title, e.g., in social media), and use shorter, simpler, and repetitive content in the body (as writing fake information takes a lot of effort). Such features can be calculated based on the Linguistic Inquiry and Word Count (LIWC) lexicon and used to distinguish articles from trusted sources vs. hoaxes vs. satire vs. propaganda (Pennebaker et al., 2001). They can be also modeled using linguistic markers (Mihaylova et al., 2018) such as *factives* (Hooper, 1975), *assertives* (Hooper, 1975), *implicatives* from (Karttunen, 1971), *hedges* (Hyland, 2005), *Wiki-bias* terms (Recasens et al., 2013), *subjectivity* cues (Riloff and Wiebe, 2003), and *sentiment* cues (Liu et al., 2005). There are 141 such features in the NELA toolkit (Horne et al., 2018).
Embedding representations: An alternative way to represent an article is to use embedding representations, typically based on large pre-trained language models, such as BERT (Devlin et al., 2019). This can be done without fine-tuning, e.g., by encoding an article (possibly truncated, e.g., BERT-base can take up to 512 tokens as an input) and then

[6]https://adfontesmedia.com/interactive-media-bias-chart/

[7]https://adfontesmedia.com/interactive-media-bias-chart/

[8]https://www.reuters.com/

[9]https://www.airforcetimes.com/

averaging the word representations extracted from the second-to-last layer. Alternatively, one can use pre-trained sentence encoders such as Sentence BERT (Reimers and Gurevych, 2019).

5.2 Multimedia Content

Nowadays, almost all news websites heavily rely on multimedia content. This dependence, however, also makes multimedia a very effective means for dispensing intended and even manipulated messages. The increasing availability of automated and AI-powered multimedia editing and synthesis tools, combined with massive computational power, makes such capabilities accessible to everyone.

Given that the multimedia editors of a news site typically follow a defined workflow when creating, acquiring, editing, and curating content for their pages, this pattern adds a crucial dimension to profiling the factuality and the bias of a news source. In fact, questions around the origin and the veracity of photographic images and videos have long been the subject of multimedia forensics research (Sencar and Memon, 2013; Sencar et al., 2022). There has been research on verifying metadata integrity (Kee et al., 2011; Iuliani et al., 2019; Yang et al., 2020), digital integrity (Korus, 2017; Cozzolino and Verdoliva, 2018), physical integrity (O'Brien et al., 2012; Iuliani et al., 2017; Matern et al., 2020; Riess et al., 2017; Peng et al., 2017) identification of processing traces (Hadwiger et al., 2019), and discrimination of synthesized (i.e., GAN generated) media (Agarwal et al., 2020; Li et al., 2018; Verdoliva, 2020). However, these capabilities have only been sparsely explored in the context of predicting factuality and bias.

Existing work mainly considered characteristics of images appearing in trustworthy vs. unreliable sources. It was proposed to use visual characteristics (Jin et al., 2017), deep-learning representations (Qi et al., 2019; Khattar et al., 2019; Singhal et al., 2019), image provenance information from reverse image search (Zlatkova et al., 2019), and self-consistency with respect to metadata (Huh et al., 2018). Overall, multimedia characteristics have a strong potential that is yet to be fully used for news media profiling.

5.3 Audience Homophily

The well-known homophily principle, "birds of a feather flock together," crucially asserts that similar individuals interact with each other at a higher rate. Therefore, audience representation could be another approach to describe a news media outlet whereby an overall descriptive characteristic of followers of the outlet is obtained. Then, by evaluating the similarity of audience-centric representations with previously categorized news media, its factuality and bias can be inferred.

Ribeiro et al. (2018) used Facebook's targeted advertising tool to infer the ideological leaning of online media based on the political leaning of the users who interacted with these media. An et al. (2012) relied on follow relationships on Twitter to ascertain the ideological leaning of news media and users. Wong et al. (2013) studied retweet behavior to infer the ideological leanings of online media sources and of popular Twitter accounts. Barberá (2015) proposed a model based on the follower relationships to media sources and Twitter personalities to estimate their ideological leaning.

Stefanov et al. (2020) predicted the political leaning of media with respect to a topic by observing the users of which side of the debate on a polarizing topic were sharing content from which media in support of their position. They constructed a user-media graph and then used label propagation and graph neural networks to derive representations for media, which they used for classification. They further aggregated the leanings across several polarizing topics to come up with a left-center-right polarization prediction.

In (Baly et al., 2020b), audience characterization was conducted across three social media platforms. Twitter profile descriptions, YouTube audience reactions, and Facebook advertising platform's demographic information were used to create representative models of each medium. The aggregated data was then categorized into five political bias labels: very conservative, conservative, moderate, liberal, and very liberal. Mehta et al. (2022) and Yang et al. (2023) addressed this challenge using a graph framework that maps relationships between news sources, their articles, and user engagement. In a similar vein, motivated by homophily considerations, Panayotov et al. (2022) modeled audience overlap to induce a graph and produce embeddings that model the similarity between news outlets.

5.4 Infrastructure Characteristics

Beyond textual, visual, and audience features, news sites also exhibit distinct characteristics that relate to the underlying infrastructure and technological components deployed to serve their content online.

In this regard, the prediction problem is analogous to a well-studied one in the cybersecurity domain, where the goal has been to identify infrastructure characteristics of malicious domains (Anderson et al., 2007; Invernizzi et al., 2014) that are used for malware distribution (Wang et al., 2013; Invernizzi et al., 2014), phishing (James et al., 2013; Mohammad et al., 2012, 2014; Purwanto et al., 2020), online scams (Alrwais et al., 2017; Konte et al., 2009; Hao et al., 2016), and spamming (Anderson et al., 2007; Hao et al., 2009). Since establishing the infrastructure of a news medium involves several decisions with respect to technological aspects, it is plausible to expect that news media with varying IT practices and different levels of access to IT resources will differ in their characteristics.

There has been very little work on network, web design, and data elements of a news website to characterize new sites for factuality and bias. At the network level, Hounsel et al. (2020) aimed to distinguish disinformation websites vs. authentic websites vs. sites not related to news or politics, and found that features related to a website's domain name, registration, and DNS configuration work best. Concerning the web design aspect, Castelo et al. (2019) introduced a web page classifier based on several features that govern the structure and the style of a page in addition to three categories of linguistic features. Their binary classification results (real vs. fake news) on several datasets showed that the web-markup features consistently perform well and are complementary to linguistic ones. Hanley et al. (2023) used partial Granger-causality to reveal positive correlations between the frequency of hyperlinks from misinformation websites and the popularity of conspiracy theory websites. Similar work suggested that news outlets known for spreading misinformation may play a significant role in popularizing conspiracy theories (Hanley et al., 2021; Sehgal et al., 2021).

Finally, at the data level, Fairbanks et al. (2018) examined the source of web pages to identify shared data objects, such as mutually linked sites, scripts, and images, across websites.

6 Major Challenges

Ordinal scales: While ideological bias (news slant) is typically modeled as left-center-right, a spectrum can exist within each bias based on bias intensity. A hyperpartisan bias prediction task has been tested to differentiate far-right from right and far-left from left, but it does not model the political bias using an ordinal scale. Difficulties in labeling the bias (i.e., creating ground-truth datasets) by experts or through crowdsourcing presents a major hurdle for modeling ideological bias as an ordinal variable.

Multimodality: In news reporting, photos typically attract high attention, and readers can sometimes understand a news story from the photos alone, even without reading the text. Indeed, news text and photos are strongly coupled and deliver relevant information about news stories to readers. Thus, modeling news text and photos together should benefit our understanding of their factuality (Alam et al., 2022) and potential harmfulness (Sharma et al., 2022).

Evaluation granularity: The label of a news medium is inferred from a sample of observations. This can introduce measurement bias if a news medium does not exhibit the same reporting behavior across all its publications. This is especially true for media with a particular stance on specific issues (Ganguly et al., 2020). Thus, reliable estimation of factuality and bias labels requires analyzing a relatively large amount of content covering a range of issues.

Variability in factuality & bias ratings: These ratings are inherently not static and may change over time when a news medium takes corrective action to address issues raised by fact-checkers. Thus, the ground truth needed for building a learning approach varies, triggering the need for re-evaluating the performance of the proposed approaches. Thus, there is a need to take into account the sensitivity of a learning approach to such small but nevertheless inevitable variations.

Dataset size: The datasets for media-level factuality and bias are relatively small, typically of a few hundred examples.They are derived from few sites, such as Media Bias/Fact Check and AllSides, where domain experts perform manual analysis.

Annotation vs. modeling: One problem is that human annotators judge the factuality of reporting and the bias of media based on criteria that are not easy to automate or based on information that may not be accessible to automatic systems. For example, if a news outlet is judged to be of mixed factuality based on it having failed just 2-3 fact-checks, for an automatic system to arrive at the same conclusion using the same idea, it would have to select for analysis the exact same articles where the false claims were made.

Data availability: Primarily due to copyright issues, there are only a few publicly available datasets of the full text of news for research purposes. Instead, indexed data (e.g., GDELT dataset) by mentioned actors, events, locations, sources, or tones are available and have been analyzed in many studies. A set of news headlines collected from news websites or aggregated websites (e.g., AllSides) are also shared more actively for research purposes. Considering the importance of social media channels in news dissemination, researchers collect and analyze social media posts of official accounts of news media. As social media posts are relatively more informal than news articles to fit for social media audience (Park et al., 2021), more studies are required for understanding their biases and factuality correctly.

Hallucinations in LLMs: While LLMs have demonstrated remarkable capabilities in various applications, including estimating source reliability as discussed in Section 2, it is crucial to acknowledge a major challenge associated with these models. LLMs, such as ChatGPT, can occasionally generate incorrect, false, or misleading content. This raises concerns about the reliability of credibility ratings and information provided in sensitive domains, such as news media profiling. Combining the strengths of LLMs with human expertise and other models can offer a more robust approach to mitigating the risks associated with hallucination in sensitive domains.

7 Future Forecasting

Support for non-English corpora and different political systems: Most studies we review are in English, and we anticipate more research on bias and factuality in other languages. Recently, various approaches have been proposed to accelerate NLP research for resource-scarce languages, such as multilingual word embeddings or large language models. We believe that those efforts help conduct bias and factuality research for non-English corpora. One non-technical issue here is that not all the countries have US-like left-center-right political biases. For example, in some countries with a multiparty system, understanding political biases is the initial step in media bias research.

Incorporation of video content: TV news accounts for significant portions of the news industry. Also, the presence of news media has become strong in video-driven social media platforms (e.g., TikTok) over time. To get high user engagement, news media outlets upload short video clips curated for social media use, particularly on existing social media. Previous studies on video news bias typically analyzed transcripts, not the video itself. Commercial databases, such as Lexis Nexis, or open-source libraries to create subtitles are used to analyze news transcripts. We expect that more studies on analyzing video contents in an end-to-end manner will be presented to fully understand the bias and factuality of video news.

Bringing practical implications: Since the factuality and the bias of news media largely influence the public, it is crucial to implement working systems so that readers can benefit from a rich stream of research. Several stand-alone websites, such as Media Bias/Fact Check, AllSides, and Tanbih (Zhang et al., 2019), aim to make media bias and factuality transparent to end-users, thus promoting media literacy. We expect new tools and services to support more media and languages.

Foundation Models: In the future, researchers can explore innovative ways to understand media better using LLMs. By asking these models specific questions, e.g., whether a media outlet has a history of sharing false information or if it leans toward certain views. Also, combining this with looking at pictures in news articles using vision-language models could give a more complete understanding of media content.

8 Lessons Learned

Factuality and bias have some commonalities as they exert negative influences on the public by delivering skewed information.

Political Landscape Positioning: News media often take a biased position in the political landscape to appeal to partisan audiences, particularly as the news industry becomes more competitive.

Reporting Perspective: Despite the bias, many journalists and editors are concerned about the issue and try to report diverse perspectives.

Multi-dimensionality: Media bias is multi-dimensional and can be conveyed through different means such as text, photos, and videos, with ideological bias being a significant conceptualization.

Factuality Challenges: Accurate prediction of ideological bias of a news medium is easier than assessing factuality, as factuality relies on verification from other sources and observations.

Sophisticated Analysis: More sophisticated analysis of text style and multimedia characteristics could improve accuracy, but other elements of a news medium need to be complemented.

Audience Homophily: Recent studies emphasize audience homophily and infrastructure characteristics in bridging performance gaps, valuable for early discovery and categorization of news media due to their content-agnostic nature.

9 Conclusion

We surveyed the emerging field of news media profiling for factuality and bias, which can enable early fake news detection. While this is a relatively understudied direction, there has been a lot of recent interest in the problem, including a shared task at the CLEF-2023 CheckThat! lab. We discussed that the factuality and the bias of news outlets are a matter of degree, and we advocated for their joint modeling. Besides textual representations, we examined the contribution of contextual frameworks. Finally, we discussed the major challenges and made a forecast for the future.

10 Limitations

Characterizing an entire media outlet presents a significant challenge. While this study reviews a variety of features used in the literature to profile outlets, dealing with the dynamic nature of news outlets in terms of their bias and factuality and handling the vast volume of digital content they produce remains a daunting task. Moreover, the research largely relies on Western definitions of political bias (left-center-right), which may not accurately capture the nuanced ideological biases present in news outlets from different cultural or political contexts.

11 Ethics Statement

Environmental Impact The energy efficiency of model training and inference operations is a critical aspect of mitigating their environmental impact. Instead of utilizing extensive computational resources for training complex models, which significantly contribute to carbon emissions, we advocate for optimizing model architectures to achieve better performance with less computational power. The methodologies reviewed in this study typically do not train large-scale models from scratch but fine-tune them on comparatively smaller datasets. This fine-tuning process is less resource-intensive. Additionally, using Central Processing Units (CPUs) for inference after the model has been fine-tuned is a viable approach that contributes less to global warming compared to more power-demanding alternatives. We thus endorse these environmentally-conscious practices in the continued development and application of AI technologies.

Misuse potential While the research, datasets, and models related to entire news outlets analysis hold immense potential, they also carry the risk of misuse. They can serve as valuable resources for fact-checkers, journalists, social media platforms, and policymakers, aiding in accurate reporting, fact verification, and policy formulation. However, the same tools can be exploited for spreading disinformation and manipulating public opinion if they fall into the wrong hands. Hence, it's crucial for researchers and practitioners in the field to exercise caution and ethical restraint. We encourage stringent safeguards and responsible use of such tools to prevent their exploitation for malicious purposes.

References

Shruti Agarwal, Hany Farid, Ohad Fried, and Maneesh Agrawala. 2020. Detecting deep-fake videos from phoneme-viseme mismatches. In *Proceedings of CVPR*, pages 2814–2822. IEEE.

Sajjad Ahmed, Knut Hinkelmann, and Flavio Corradini. 2022. Combining machine learning with knowledge engineering to detect fake news in social networks-a survey. *arXiv preprint arXiv:2201.08032*.

Firoj Alam, Alberto Barrón-Cedeño, Gullal S. Cheema, Gautam Kishore Shahi, Sherzod Hakimov, Maram Hasanain, Chengkai Li, Rubén Míguez, Hamdy Mubarak, Wajdi Zaghouani, and Preslav Nakov. 2023. Overview of the clef-2023 checkthat! lab task 1 on check-worthiness in multimodal and multigenre content. In *Conference and Labs of the Evaluation Forum*.

Firoj Alam, Stefano Cresci, Tanmoy Chakraborty, Fabrizio Silvestri, Dimiter Dimitrov, Giovanni Da San Martino, Shaden Shaar, Hamed Firooz, and Preslav Nakov. 2022. A survey on multimodal disinformation detection. In *Proceedings of COLING*, pages 6625–6643, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Sumayah A. Alrwais, Xiaojing Liao, Xianghang Mi, Peng Wang, Xiaofeng Wang, Feng Qian, Raheem A.

Beyah, and Damon McCoy. 2017. Under the shadow of sunshine: Understanding and detecting bulletproof hosting on legitimate service provider networks. *2017 IEEE Symposium on Security and Privacy (SP)*, pages 805–823.

Jisun An, M. Cha, Krishna P. Gummadi, and Jon A. Crowcroft. 2011. Media landscape in twitter: A world of new conventions and political diversity. *Proceedings of the International AAAI Conference on Web and Social Media*.

Jisun An, Meeyoung Cha, Krishna Gummadi, Jon Crowcroft, and Daniele Quercia. 2012. Visualizing media bias through Twitter. In *Proceedings of AAAI ICWSM*, volume 6.

David S. Anderson, Chris Fleizach, Stefan Savage, and Geoffrey M. Voelker. 2007. Spamscatter: Characterizing internet scam hosting infrastructure. In *USENIX Security Symposium*.

R. Baly, Georgi Karadzhov, Abdelrhman Saleh, James R. Glass, and Preslav Nakov. 2019. Multi-task ordinal regression for jointly predicting the trustworthiness and the leading political ideology of news media. In *North American Chapter of the Association for Computational Linguistics*.

Ramy Baly, Giovanni Da San Martino, James Glass, and Preslav Nakov. 2020a. We can detect your bias: Predicting the political ideology of news articles. In *Proceedings of EMNLP*, pages 4982–4991, Online. Association for Computational Linguistics.

Ramy Baly, Georgi Karadzhov, Dimitar Alexandrov, James Glass, and Preslav Nakov. 2018a. Predicting factuality of reporting and bias of news media sources. In *Proceedings of EMNLP*, pages 3528–3539, Brussels, Belgium. ACL.

Ramy Baly, Georgi Karadzhov, Jisun An, Haewoon Kwak, Yoan Dinkov, Ahmed Ali, James Glass, and Preslav Nakov. 2020b. What was written vs. who read it: News media profiling using text analysis and social media context. In *Proceedings of the 58th Annual Meeting of ACL*, pages 3364–3374, Online. ACL.

Ramy Baly, Mitra Mohtarami, James Glass, Lluís Màrquez, Alessandro Moschitti, and Preslav Nakov. 2018b. Integrating stance detection and fact checking in a unified corpus. In *Proceedings of NAACL: Human Language Technologies, Volume 2 (Short Papers)*, pages 21–27, New Orleans, Louisiana. ACL.

Pablo Barberá. 2015. Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political Analysis*, 23:76 – 91.

Andrew W. Barrett and Lowell W. Barrington. 2005. Bias in newspaper photograph selection. *Political Research Quarterly (formerly WPQ)*, 58:609 – 618.

Alberto Barrón-Cedeño, Firoj Alam, Andrea Galassi, Giovanni Da San Martino, Preslav Nakov, T. Elsayed, Dilshod Azizov, Tommaso Caselli, Gullal S. Cheema, Fatima Haouari, Maram Hasanain, Mucahid Kutlu, Chengkai Li, Federico Ruggeri, Julia Maria Struß, and Wajdi Zaghouani. 2023. Overview of the clef-2023 checkthat! lab on checkworthiness, subjectivity, political bias, factuality, and authority of news articles and their source. In *Conference and Labs of the Evaluation Forum*.

Matthew A Baum and Tim Groeling. 2008. New media and the polarization of American political discourse. *Political Communication*, 25(4):345–365.

Eric Baumer, Elisha Elovic, Ying Qin, Francesca Polletta, and Geri Gay. 2015. Testing and comparing computational approaches for identifying the language of framing in political news. In *Proceedings of NAACL: Human Language Technologies*, pages 1472–1482, Denver, Colorado. ACL.

Vibhu Bhatia, Vidya Prasad Akavoor, Sejin Paik, Lei Guo, Mona Jalal, Alyssa Smith, David Assefa Tofu, Edward Edberg Halim, Yimeng Sun, Margrit Betke, Prakash Ishwar, and Derry Tanti Wijaya. 2021. OpenFraming: Open-sourced tool for computational framing analysis of multilingual data. In *Proceedings of EMNLP*, pages 242–250, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Emina Boudemagh and Izabela Moise. 2017. News media coverage of refugees in 2016: A GDELT case study. *Proceedings of the International AAAI Conference on Web and Social Media*.

Maxwell Boykoff and Jules Boykoff. 2004. Balance as bias: global warming and the us prestige press. *Global Environmental Change-human and Policy Dimensions*, 14:125–136.

Ann M. Brill. 2001. Online journalists embrace new marketing function. *Newspaper Research Journal*, 22:28 – 40.

Canini et al. 2011. Finding credible information sources in social networks based on content and social structure. In *Proceedings of IEEE SocialCom/PASSAT*.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the ACL and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 438–444, Beijing, China. ACL.

Sonia Castelo, Thais G. Almeida, Anas Elghafari, Aécio Santos, Kien Pham, Eduardo Freire Nakamura, and Juliana Freire. 2019. A topic-agnostic approach for identifying fake news pages. *Companion Proceedings of The 2019 World Wide Web Conference*.

Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *Proceedings of WWW Conference*, pages 675–684. ACM.

Deptii Devendra Chaudhari and Ambika Vishal Pawar. 2021. Propaganda analysis in social media: a bibliometric review. *Information Discovery and Delivery*, 49(1):57–70.

Cheng Chen, Kui Wu, Srinivasan Venkatesh, and Xudong Zhang. 2013. Battling the internet water army: detection of hidden paid posters. In *Proceedings of ASONAM*, pages 116–120. ACM.

Davide Cozzolino and Luisa Verdoliva. 2018. Camera-based image forgery localization using convolutional neural networks. *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 1372–1376.

Fernando Cardoso Durier da Silva, Rafael Vieira, and Ana Cristina Bicharra Garcia. 2019. Can machines learn to detect fake news? a survey focused on social media. In *Hawaii International Conference on System Sciences*.

Dave D'Alessio and Mike Allen. 2000. Media bias in presidential elections: a meta-analysis. *Journal of Communication*, 50:133–156.

Kareem Darwish, Dimitar Alexandrov, Preslav Nakov, and Yelena Mejova. Seminar users in the Arabic Twitter sphere. In *Proceedings of SocInfo*.

Kareem Darwish, Peter Stefanov, Michaël Aupetit, and Preslav Nakov. 2019. Unsupervised user stance detection on twitter. *ArXiv*, abs/1904.02000.

Sohan De Sarkar, Fan Yang, and Arjun Mukherjee. 2018. Attending sentences to detect satirical fake news. In *Proceedings of COLING*, pages 3371–3380, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the NAACL: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. ACL.

Yoan Dinkov, Ahmed Ali, Ivan Koychev, and Preslav Nakov. 2019. Predicting the leading political ideology of Youtube channels using acoustic, textual, and metadata information. In *Proceedings of Interspeech ISCA*, pages 501–505. ISCA.

Xin Luna Dong, Evgeniy Gabrilovich, Kevin Murphy, Van Dang, Wilko Horn, Camillo Lugaresi, Shaohua Sun, and Wei Zhang. 2015. Knowledge-based trust: Estimating the trustworthiness of web sources. *VLDB Endow.*, 8(9):938–949.

Sebastian Dungs, Ahmet Aker, Norbert Fuhr, and Kalina Bontcheva. 2018. Can rumour stance alone predict veracity? In *Proceedings of COLING*, pages 3360–3370, Santa Fe, New Mexico, USA. ACL.

Robert M. Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4):51–58.

Robert M. Entman. 2007. Framing bias: Media in the distribution of power. *Journal of Communication*, 57:163–173.

James David Fairbanks, Nathan Knauf, and Erica Briscoe Georgia. 2018. Credibility assessment in the news : Do we need to read ?

Anjalie Field, Doron Kliger, Shuly Wintner, Jennifer Pan, Dan Jurafsky, and Yulia Tsvetkov. 2018. Framing and agenda-setting in Russian news: a computational analysis of intricate political strategies. In *Proceedings of EMNLP*, pages 3570–3580, Brussels, Belgium. Association for Computational Linguistics.

Finberg et al. 2002. Digital journalism credibility study. *Online News Association. Retrieved November*, 3:2003.

Soumen Ganguly, Juhi Kulshrestha, Jisun An, and Haewoon Kwak. 2020. Empirical evaluation of three common assumptions in building political media bias datasets. In *International Conference on Web and Social Media*.

Matthew Gentzkow and Jesse M. Shapiro. 2010. What drives media slant? evidence from U.S. daily newspapers. *Econometrica*, 78(1):35–71.

Martin Gilens and C. Hertzman. 2000. Corporate ownership and news bias: Newspaper coverage of the 1996 telecommunications act. *The Journal of Politics*, 62:369 – 386.

Tim J. Groeling. 2013. Media bias by the numbers: Challenges and opportunities in the empirical study of partisan news. *Annual Review of Political Science*, 16:129–151.

Tim Groseclose and Jeffrey Milyo. 2005. A measure of media bias. *Quarterly Journal of Economics*, 120:1191–1237.

Lei Guo and Chris J Vargo. 2017. Global intermedia agenda setting: A big data analysis of international news flow. *Journal of Communication*, 67(4):499–520.

Xiaobo Guo, Weicheng Ma, and Soroush Vosoughi. 2022a. Measuring media bias via masked language modeling. In *Proceedings of AAAI ICWSM*, volume 16, pages 1404–1408.

Zhijiang Guo, Michael Schlichtkrull, and Andreas Vlachos. 2022b. A survey on automated fact-checking. *Transactions of the Association for Computational Linguistics*, 10:178–206.

Benjamin Hadwiger, Daniele Baracchi, Alessandro Piva, and Christian Riess. 2019. Towards learned color representations for image splicing detection. In *Proceedings of IEEE ICASSP*, pages 8281–8285. IEEE.

Hans W. A. Hanley, Deepak Kumar, and Zakir Durumeric. 2021. No calm in the storm: Investigating qanon website relationships. *ArXiv*, abs/2106.15715.

Hans W. A. Hanley, Deepak Kumar, and Zakir Durumeric. 2023. A golden age: Conspiracy theories' relationship with misinformation outlets, news media, and the wider internet. *Proceedings of the ACM on Human-Computer Interaction*, 7:1 – 33.

Shuang Hao, Alex Kantchelian, Brad Miller, Vern Paxson, and Nick Feamster. 2016. Predator: Proactive recognition and elimination of domain abuse at time-of-registration. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*.

Shuang Hao, Nadeem Ahmed Syed, Nick Feamster, Alexander G. Gray, and Sven Krasser. 2009. Detecting spammers with snare: Spatio-temporal network-level automatic reputation engine. In *USENIX Security Symposium*.

Momchil Hardalov, Arnav Arora, Preslav Nakov, and Isabelle Augenstein. 2022a. A survey on stance detection for mis- and disinformation identification. In *Findings of the ACL: NAACL*, pages 1259–1277, Seattle, United States. ACL.

Momchil Hardalov, Anton Chernyavskiy, Ivan Koychev, Dmitry I. Ilvovsky, and Preslav Nakov. 2022b. Crowdchecked: Detecting previously fact-checked claims in social media. In *AACL*.

Momchil Hardalov, Ivan Koychev, and Preslav Nakov. 2016. In search of credible news. In *Proceedings of AIMSA*, pages 172–180.

Hans J. G. Hassell, John B. Holbein, and Matthew R. Miles. 2020. There is no liberal media bias in which news stories political journalists choose to cover. *Science Advances*, 6.

Joan Bybee Hooper. 1975. On assertive predicates.

Benjamin D. Horne and Sibel Adali. 2017. This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. *ArXiv*, abs/1703.09398.

Benjamin D. Horne, William Dron, Sara Khedr, and Sibel Adali. 2018. Sampling the news producers: A large news and feature data set for the study of the complex media landscape. In *International Conference on Web and Social Media*.

Austin Hounsel, Jordan Holland, Ben Kaiser, Kevin Borgolte, Nick Feamster, and Jonathan R. Mayer. 2020. Identifying disinformation websites using infrastructure features. *arXiv: Computers and Society*.

Minyoung Huh, Andrew Liu, Andrew Owens, and Alexei A. Efros. 2018. Fighting fake news: Image splice detection via learned self-consistency. *ArXiv*, abs/1805.04096.

Ken Hyland. 2005. Metadiscourse: Exploring interaction in writing.

Luca Invernizzi, Stanislav Miskovic, Rubén Torres, Christopher Krügel, Sabyasachi Saha, Giovanni Vigna, Sung-Ju Lee, and Marco Mellia. 2014. Nazca: Detecting malware distribution in large-scale networks. In *Network and Distributed System Security Symposium*.

Md. Rafiqul Islam, S. Liu, Xianzhi Wang, and Guandong Xu. 2020. Deep learning for misinformation detection on online social networks: a survey and new perspectives. *Social Network Analysis and Mining*, 10.

Massimo Iuliani, Marco Fanfani, Carlo Colombo, and Alessandro Piva. 2017. Reliability assessment of principal point estimates for forensic applications. *J. Vis. Commun. Image Represent.*, 42:65–77.

Massimo Iuliani, Dasara Shullani, Marco Fontani, Saverio Meucci, and Alessandro Piva. 2019. A video forensic framework for the unsupervised analysis of mp4-like file container. *IEEE Transactions on Information Forensics and Security*, 14:635–645.

Joby James, Sandhya L, and Ciza Thomas. 2013. Detection of phishing urls using machine learning techniques. *2013 International Conference on Control Communication and Computing (ICCC)*, pages 304–309.

Zhiwei Jin, Juan Cao, Yongdong Zhang, Jianshe Zhou, and Qi Tian. 2017. Novel visual and statistical image features for microblogs news verification. *IEEE Transactions on Multimedia*, 19:598–608.

Kristen Johnson, Di Jin, and Dan Goldwasser. 2017. Leveraging behavioral and social information for weakly supervised collective classification of political discourse on Twitter. In *Proceedings of the 55th Annual Meeting of the ACL (Volume 1: Long Papers)*, pages 741–752, Vancouver, Canada. ACL.

Georgi Karadzhov, Pepa Gencheva, Preslav Nakov, and Ivan Koychev. 2017a. We built a fake news / click bait filter: What happened next will blow your mind! In *Proceedings of the RANLP*, pages 334–343, Varna, Bulgaria. INCOMA Ltd.

Georgi Karadzhov, Preslav Nakov, Lluís Màrquez, Alberto Barrón-Cedeño, and Ivan Koychev. 2017b. Fully automated fact checking using external sources. In *Proceedings of RANLP*, pages 344–353, Varna, Bulgaria.

Lauri Karttunen. 1971. Implicative verbs. *Language*, 47(2):340–358.

Eric Kee, Micah K. Johnson, and Hany Farid. 2011. Digital image authentication from JPEG headers. *IEEE TIFS*, 6(3):1066–1075.

Dhruv Khattar, Jaipal Singh Goud, Manish Gupta, and Vasudeva Varma. 2019. MVAE: multimodal variational autoencoder for fake news detection. In *Proceedings of WWW Conference*, pages 2915–2921. ACM.

Johannes Kiesel, Maria Mestre, Rishabh Shukla, Emmanuel Vincent, Payam Adineh, David Corney, Benno Stein, and Martin Potthast. 2019. SemEval-2019 task 4: Hyperpartisan news detection. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 829–839, Minneapolis, Minnesota, USA. ACL.

Michelle YoungJin Kim and Kristen Marie Johnson. 2022. CLoSE: Contrastive learning of subframe embeddings for political bias classification of news media. In *Proceedings of the COLING*, pages 2780–2793, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2018. All-in-one: Multi-task learning for rumour verification. In *Proceedings of COLING*, pages 3402–3413, Santa Fe, New Mexico, USA. ACL.

Maria Konte, Nick Feamster, and Jaeyeon Jung. 2009. Dynamics of online scam hosting infrastructure. In *Passive and Active Network Measurement Conference*.

Pawel Korus. 2017. Digital image integrity - a survey of protection and verification techniques. *Digit. Signal Process.*, 71:1–26.

Neema Kotonya and Francesca Toni. 2020. Explainable automated fact-checking: A survey. In *Proceedings of COLING*, pages 5430–5443, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Vivek Kulkarni, Junting Ye, Steve Skiena, and William Yang Wang. 2018. Multi-view models for political ideology detection of news articles. In *Proceedings of EMNLP*, pages 3518–3527, Brussels, Belgium. ACL.

Haewoon Kwak and Jisun An. 2014. A first look at global news coverage of disasters by using the gdelt dataset. In *Social Informatics*.

Haewoon Kwak and Jisun An. 2016. Two tales of the world: Comparison of widely used world news datasets GDELT and EventRegistry. In *Proceedings of AAAI ICWSM*.

Haewoon Kwak, Jisun An, and Yong-Yeol Ahn. 2020. A systematic media frame analysis of 1.5 million new york times articles from 2000 to 2017. *Proceedings of the 12th ACM Conference on Web Science*.

Haewoon Kwak, Jisun An, Joni Salminen, Soon-Gyo Jung, and Bernard J Jansen. 2018. What we read, what we search: Media attention and public attention among 193 countries. In *Proceedings of the 2018 World Wide Web Conference*, pages 893–902.

Yaliang Li, Jing Gao, Chuishi Meng, Qi Li, Lu Su, Bo Zhao, Wei Fan, and Jiawei Han. 2016. A survey on truth discovery. *SIGKDD Explor. Newsl.*, 17(2):1–16.

Yuezun Li, Ming-Ching Chang, and Siwei Lyu. 2018. In ictu oculi: Exposing ai created fake videos by detecting eye blinking. *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–7.

Sora Lim, Adam Jatowt, Michael Färber, and Masatoshi Yoshikawa. 2020. Annotating and analyzing biased sentences in news articles using crowdsourcing. In *Proceedings of LREC*, pages 1478–1484.

Benjamin Litterer, David Jurgens, and Dallas Card. 2023. When it rains, it pours: Modeling media storms and the news ecosystem. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 6346–6361.

Bing Liu, Minqing Hu, and Junsheng Cheng. 2005. Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of ACM WWW*, pages 342–351. ACM.

Siyi Liu, Lei Guo, Kate Mays, Margrit Betke, and Derry Tanti Wijaya. 2019. Detecting frames in news headlines and its application to analyzing news framing trends surrounding U.S. gun violence. In *Proceedings of CoNLL*, pages 504–514, Hong Kong, China. ACL.

Yujian Liu, Xinliang Frederick Zhang, David Wegsman, Nick Beauchamp, and Lu Wang. 2022. Politics: Pretraining with same-story article comparison for ideology prediction and stance detection. *ArXiv*, abs/2205.00619.

Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J. Jansen, Kam-Fai Wong, and Meeyoung Cha. 2016. Detecting rumors from microblogs with recurrent neural networks. In *Proceedings of IJCAI*, pages 3818–3824. IJCAI/AAAI Press.

Jing Ma, Wei Gao, Zhongyu Wei, Yueming Lu, and Kam-Fai Wong. 2015. Detect rumors using time series of social context information on microblogging websites. *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*.

Jing Ma, Wei Gao, and Kam-Fai Wong. 2017. Detect rumors in microblog posts using propagation structure via kernel learning. In *Proceedings of the 55th Annual Meeting of ACL (Volume 1: Long Papers)*, pages 708–717, Vancouver, Canada. ACL.

Suman Kalyan Maity, Aishik Chakraborty, Pawan Goyal, and Animesh Mukherjee. 2017. Detection of sockpuppets in social media. In *Proceedings of CSCW*, pages 243–246.

Giovanni Da San Martino, Firoj Alam, Maram Hasanain, Rabindra Nath Nandi, Dilshod Azizov, and Preslav Nakov. 2023. Overview of the clef-2023 checkthat! lab task 3 on political bias of news articles and news media. In *Conference and Labs of the Evaluation Forum*.

Giovanni Da San Martino, Stefano Cresci, Alberto Barrón-Cedeño, Seunghak Yu, Roberto Di Pietro, and Preslav Nakov. 2020. A survey on computational propaganda detection. In *Proceedings of IJCAI*, pages 4826–4832. ijcai.org.

Falko Matern, Christian Riess, and Marc Stamminger. 2020. Gradient-based illumination description for image forgery detection. *IEEE Transactions on Information Forensics and Security*, 15:1303–1317.

Nikhil Mehta and Dan Goldwasser. 2023. An interactive framework for profiling news media sources. *arXiv preprint arXiv:2309.07384*.

Nikhil Mehta, Maria Leonor Pacheco, and Dan Goldwasser. 2022. Tackling fake news detection by continually improving social context representations using graph neural networks. In *Annual Meeting of the Association for Computational Linguistics*.

Todor Mihaylov and Preslav Nakov. 2016. Hunting for troll comments in news community forums. In *Proceedings of the 54th Annual Meeting of the ACL (Volume 2: Short Papers)*, pages 399–405, Berlin, Germany. ACL.

Tsvetomila Mihaylova, Preslav Nakov, Lluís Màrquez, Alberto Barrón-Cedeño, Mitra Mohtarami, Georgi Karadzhov, and James R. Glass. 2018. Fact checking in community forums. In *Proceedings of AAAI*, pages 5309–5316, New Orleans, Louisiana, USA.

Shubha Mishra, Piyush Kumar Shukla, and Ratish Agarwal. 2022. Analyzing machine learning enabled fake news detection techniques for diversified datasets.

R. M. Mohammad, F. Thabtah, and L. McCluskey. 2012. An assessment of features related to phishing websites using an automated technique. In *Proceedings of IEEE ICITST*, pages 492–497.

Rami Mustafa A. Mohammad, Fadi A. Thabtah, and Lee Mccluskey. 2014. Predicting phishing websites based on self-structuring neural network. *Neural Computing and Applications*, 25:443–458.

Morstatter et al. 2018. Identifying framing bias in online news. *ACM Transactions on Social Computing*, 1(2):1–18.

Muhammad Firoz Mridha, Ashfia Jannat Keya, Md. Abdul Hamid, Muhammad Mostafa Monowar, and Md Saifur Rahman. 2021. A comprehensive review on fake news detection with deep learning. *IEEE Access*, 9:156151–156170.

Subhabrata Mukherjee and Gerhard Weikum. 2015. Leveraging joint interactions for credibility analysis in news communities. *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*.

Preslav Nakov, Firoj Alam, Giovanni Da San Martino, Maram Hasanain, Dilshod Azizov, Rabindra Nath Nandi, and Panayot P. Panayotov. 2023. Overview of the clef-2023 checkthat! lab task 4 on factuality of reporting of news media. In *Conference and Labs of the Evaluation Forum*.

Preslav Nakov, David P. A. Corney, Maram Hasanain, Firoj Alam, Tamer Elsayed, Alberto Barr'on-Cedeno, Paolo Papotti, Shaden Shaar, and Giovanni Da San Martino. 2021. Automated fact-checking for assisting human fact-checkers. *ArXiv*, abs/2103.07769.

An T. Nguyen, Aditya Kharosekar, Matthew Lease, and Byron C. Wallace. 2018. An interpretable joint graphical model for fact-checking from crowds. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 1511–1518. AAAI Press.

Van-Hoang Nguyen, Kazunari Sugiyama, Preslav Nakov, and Min-Yen Kan. 2020. FANG: leveraging social context for fake news detection using graph representation. In *Proceedings of ACM CIKM*, pages 1165–1174. ACM.

Jeppe Nørregaard, Benjamin D. Horne, and Sibel Adalı. 2019. NELA-GT-2018: A large multi-labelled news dataset for the study of misinformation in news articles. In *Proceedings of AAAI ICWSM*, pages 630–638.

O'Brien et al. 2012. Exposing photo manipulation with inconsistent reflections. *ACM Trans. Graph.*, 31(1):4:1–4:11.

Jeremy Padgett, Johanna L. Dunaway, and Joshua P. Darr. 2019. As seen on tv? how gatekeeping makes the u.s. house seem more extreme. *Journal of Communication*, 69:696–719.

Jeff Z. Pan, Siyana Pavlova, Chenxi Li, Ningxi Li, Yangmei Li, and Jinshuo Liu. 2018. Content based fake news detection using knowledge graphs. In *International Workshop on the Semantic Web*.

Panayot Panayotov, Utsav Shukla, Husrev Taha Sencar, Mohamed Nabeel, and Preslav Nakov. 2022. Greener: Graph neural networks for news media profiling. *ArXiv*, abs/2211.05533.

Kunwoo Park, Haewoon Kwak, Jisun An, and Sanjay Chawla. 2021. Understanding effects of editing tweets for news sharing by media accounts through a causal inference framework. In *Proceedings of AAAI ICWSM*.

Bo Peng, Wei Wang, Jing Dong, and Tieniu Tan. 2017. Optimized 3D lighting environment estimation for image forgery detection. *IEEE TIFS*, 12(2):479–494.

Pennebaker et al. 2001. Linguistic inquiry and word count: LIWC 2001. *LIWC Operators Manual 2001*.

Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2017. Automatic detection of fake news. *arXiv preprint arXiv:1708.07104*.

Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2018. Automatic detection of fake news. In *Proceedings of COLING*, pages 3391–3401, Santa Fe, New Mexico, USA.

Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. 2016. Credibility assessment of textual claims on the web. In *Proceedings of ACM CIKM*.

Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. 2017. Where the truth lies: Explaining the credibility of emerging claims on the Web and social media. In *Proceedings of WWW Companion*, pages 1003–1012.

Popat et al. 2018. CredEye: A credibility lens for analyzing and explaining misinformation. In *Proceedings of WWW Conference*.

Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2018. A stylometric inquiry into hyperpartisan and fake news. In *Proceedings of the 56th Annual Meeting of the ACL (Volume 1: Long Papers)*, pages 231–240, Melbourne, Australia. ACL.

Purwanto et al. 2020. PhishZip: A new compression-based algorithm for detecting phishing websites. In *Proceedings of IEEE CNS*.

Peng Qi, Juan Cao, Tianyun Yang, Junbo Guo, and Jintao Li. 2019. Exploiting multi-domain visual information for fake news detection. In *Proceedings of IEEE ICDM*, pages 518–527.

Chengwei Qin, Aston Zhang, Zhuosheng Zhang, Jiaao Chen, Michihiro Yasunaga, and Diyi Yang. 2023. Is ChatGPT a general-purpose natural language processing task solver? *arXiv preprint arXiv:2302.06476*.

Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2013. Linguistic models for analyzing and detecting biased language. In *Annual Meeting of the Association for Computational Linguistics*.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of EMNLP-IJCNLP*, pages 3982–3992, Hong Kong, China. ACL.

Ribeiro et al. 2018. Media bias monitor: Quantifying biases of social media news outlets at large-scale. In *Proceedings of AAAI ICWSM*, pages 290–299.

Christian Riess, Mathias Unberath, Farzad Naderi, Sven Pfaller, Marc Stamminger, and Elli Angelopoulou. 2017. Handling multiple materials for exposure of digital forgeries using 2-d lighting environments. *MM. Tools Appl.*, 76(4):4747–4764.

Ellen Riloff and Janyce Wiebe. 2003. Learning extraction patterns for subjective expressions. In *Proceedings of EMNLP*.

Vibhor Sehgal, Ankit Peshin, Sadia Afroz, and Hany Farid. 2021. Mutual hyperlinking among misinformation peddlers. *ArXiv*, abs/2104.11694.

Husrev Taha Sencar and Nasir Memon, editors. 2013. *Digital Image Forensics*. Springer.

Sencar et al., editor. 2022. *Multimedia Forensics*. Springer.

Shivam Sharma, Firoj Alam, Md. Shad Akhtar, Dimitar Dimitrov, Giovanni Da San Martino, Hamed Firooz, Alon Halevy, Fabrizio Silvestri, Preslav Nakov, and Tanmoy Chakraborty. 2022. Detecting and understanding harmful memes: A survey. In *Proceedings of IJCAI*, IJCAI-ECAI '22, Vienna, Austria.

Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *SIGKDD Explor. Newsl.*, 19(1).

Shivangi Singhal, Rajiv Ratn Shah, Tanmoy Chakraborty, Ponnurangam Kumaraguru, and Shin'ichi Satoh. 2019. Spotfake: A multi-modal framework for fake news detection. In *Proceedings of IEEE BigMM*, pages 39–47.

Jackie Smith, John D. McCarthy, Clark Mcphail, and Boguslaw Augustyn. 2001. From protest to agenda building: Description bias in media coverage of protest events in washington, d.c. *Social Forces*, 79:1397 – 1423.

Stuart N Soroka. 2012. The gatekeeping function: Distributions of information in media and the real world. *The Journal of Politics*, 74(2):514–528.

Timo Spinde, Manuel Plank, Jan-David Krieger, Terry Ruas, Bela Gipp, and Akiko Aizawa. 2022. Neural media bias detection using distant supervision with BABE–bias annotations by experts. *arXiv preprint arXiv:2209.14557*.

Peter Stefanov, Kareem Darwish, Atanas Atanasov, and Preslav Nakov. 2020. Predicting the topical stance and political leaning of media using tweets. In *Proceedings of the 58th Annual Meeting of ACL*, pages 527–537, Online. ACL.

Stevenson et al. 1973. Untwisting the news twisters: A replication of Efron's study. *Journalism Quarterly*, 50(2):211–219.

Edson C Tandoc Jr. 2014. Journalism is twerking? how web analytics is changing the process of gatekeeping. *New media & society*, 16(4).

James Thorne and Andreas Vlachos. 2018. Automated fact checking: Task formulations, methods and future directions. In *Proceedings of COLING*, pages 3346–3359, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Gisela Vallejo, Timothy Baldwin, and Lea Frermann. 2023. Connecting the dots in news analysis: A cross-disciplinary survey of media bias and framing. *arXiv preprint arXiv:2309.08069*.

Luisa Verdoliva. 2020. Media forensics and deepfakes: An overview. *IEEE Journal of Selected Topics in Signal Processing*, 14:910–932.

Paul Waldman and James Devitt. 1998. Newspaper photographs and the 1996 presidential election: The question of bias. *Journal Mass Commun Q*, 75(2):302–311.

Robert Walecki, Ognjen Rudovic, Vladimir Pavlovic, and Maja Pantic. 2016. Copula ordinal regression for joint estimation of facial action unit intensity. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4902–4910.

G. Wang, J. W. Stokes, C. Herley, and D. Felstead. 2013. Detecting malicious landing pages in malware distribution networks. In *Proceedings of IEEE/IFIP DSN*, pages 1–11.

David Weaver and Bruce Bimber. 2008. Finding news stories: A comparison of searches using lexisnexis and google news. *Journalism & Mass Communication Quarterly*, 85:515 – 530.

Felix Ming Fai Wong, Chee Wei Tan, Soumya Sen, and Mung Chiang. 2013. Quantifying political leaning from tweets and retweets. *Proceedings of the International AAAI Conference on Web and Social Media*.

Fan Xu, Victor S Sheng, and Mingwen Wang. 2021. A unified perspective for disinformation detection and truth discovery in social sensing: A survey. *ACM Computing Surveys (CSUR)*, 55(1):1–33.

Kai-Cheng Yang and Filippo Menczer. 2023. Large language models can rate news outlet credibility. *arXiv preprint arXiv:2304.00228*.

Pengpeng Yang, Daniele Baracchi, Massimo Iuliani, Dasara Shullani, Rongrong Ni, Yao Zhao, and Alessandro Piva. 2020. Efficient video integrity analysis through container characterization. *IEEE Journal of Selected Topics in Signal Processing*, 14:947–954.

Sin-Han Yang, Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2023. Entity-aware dual co-attention network for fake news detection. In *Findings*.

Yifan Zhang, Giovanni Da San Martino, Alberto Barrón-Cedeño, Salvatore Romeo, Jisun An, Haewoon Kwak, Todor Staykovski, Israa Jaradat, Georgi Karadzhov, Ramy Baly, Kareem Darwish, James Glass, and Preslav Nakov. 2019. Tanbih: Get to know what you are reading. In *Proceedings of EMNLP-IJCNLP*, pages 223–228, Hong Kong, China. ACL.

Xinyi Zhou. 2021. A survey of fake news: Fundamental theories, detection methods, and opportunities.

Dimitrina Zlatkova, Preslav Nakov, and Ivan Koychev. 2019. Fact-checking meets fauxtography: Verifying claims about images. In *Proceedings of EMNLP-IJCNLP*, pages 2099–2108, Hong Kong, China. Association for Computational Linguistics.

Arkaitz Zubiaga, Geraldine Wong Sak Hoi, Maria Liakata, Rob Procter, and Peter Tolmie. 2015. Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PLoS ONE*, 11.