Connecting the Dots in News Analysis: A Cross-Disciplinary Survey of Media Bias and Framing

Anonymous ACL submission

Abstract

The manifestation and effect of bias in news reporting have been central topics in the social sciences for decades, and have received increasing attention in the NLP community recently. While NLP can help to scale up analyses or contribute automatic procedures to investigate the impact of biased news in society, we argue that methodologies that are currently dominant fall short of capturing the complex questions and effects addressed in theoretical media studies. This is problematic because it diminishes the validity and safety of resulting tools and applications. Here, we review and critically compare task formulations, methods and evaluation schemes in the social sciences and NLP. We discuss open questions and suggest possible directions to close identified gaps between theory and predictive models, and their evaluation. These include model transparency, considering document-external information, and cross-document reasoning.

1 Introduction

The depiction of complex issues in the media strongly impacts public opinion, politics, and policies (Ghanem, 1997; Giles and Shaw, 2009). Because a handful of global corporations own an increasing proportion of news outlets, the reach and impact of biased reporting are amplified (Hamborg, 2020). Although perfect neutrality is neither realistic nor desirable, media bias turns into an issue when it becomes systematic. If the public is unaware of the presence of bias, this can lead to dangerous consequences, including intolerance and ideological segregation (Baly et al., 2020).

Figure 1 illustrates the concepts of framing and media bias adopted in this paper, using the passing of the Respect for Marriage Act as an example. *Framing* refers to the emphasis of selected facts with the goal of eliciting a desired interpretation or reaction in the reader (Entman, 2007). The

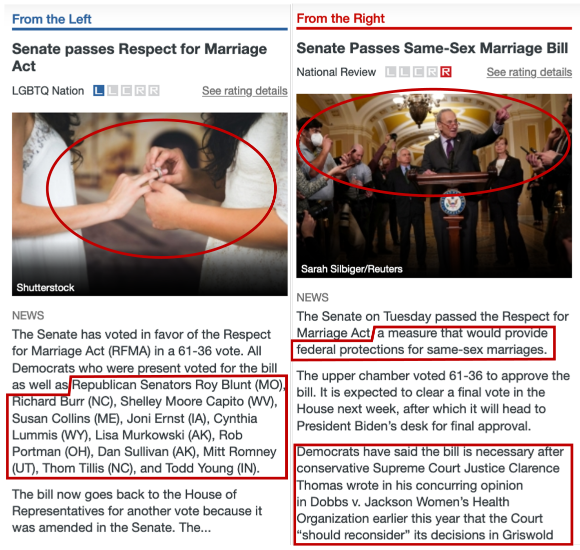

Figure 1: Two articles about the same event written from different political ideologies.

left-leaning article in Figure 1 leads with an uplifting picture of a wedding and emphasizes bill support, evoking a positive framing by emphasizing new opportunities for same-sex couples; while the right-leaning article focuses on concerns and debates in both image and text, framing the issue in a more negative light. *Political bias* refers to partisan slanted news stories, or the "tendency to deviate from an accurate, neutral, balanced, and impartial representation of 'reality' of events and social world" (McQuail and Deuze, 2020), which can be a result of a selected framing. In Figure 1, each document was flagged as far-left and far-right ideological leaning, respectively, on the basis of their publishing media outlets.[1] Political bias is typically deliberate (Williams, 1975) while framing may be inadvertent as a result of focusing on selective information due to external pressures such as space limitations.

For decades, framing and media bias have been

[1]Examples and categorization taken from Allsides.com.

under active research in different subfields of the social sciences. Angles of study include the manifestation of frames in the mass media and their effects on public opinion (communication sciences); the impact of frames in groups' and individuals' sensemaking of the world (social psychology; sociology) or on their observable behaviour (economics and political science), to name a few. This paper primarily focusses on the first notion of bias and framing: its systematic analysis in the mass media, through manual coding, or with NLP technology.[2]

The increasing pace of news reporting and move to digital news sources suggests both a need and opportunity to scale the process of media bias detection (Parasie, 2022). Besides, there is evidence that exposing media bias promotes healthy public debate, helps journalists to increase thoroughness and objectivity, and promotes critical and conscious news consumption (Dallmann et al., 2015). While we abstract from concrete applications of NLP technologies for the main part of the paper, we discuss the role of NLP in the context of news analysis in Section 5.

1.1 Contribution and Approach

In this paper, we survey work on framing and media bias prediction in NLP and relate it to typical research questions and hypotheses in the social sciences. We tease out disconnects across disciplines, and make concrete suggestions on how social science approaches can improve NLP methodology, and how NLP methods can more effectively aid social science scholars in their analyses and underpin technology to raise awareness of media bias.

Hamborg et al. (2019) present an overview of traditional and computational approaches to media bias, including detailed definitions of bias types and their emergence in the context of news production. We complement the survey by providing a more in-depth review of research methodologies in NLP, more recent computational approaches, and a unified focus on the phenomenon of framing and its manifestation as media bias. A recent survey by Ali and Hassan (2022) reviews computational approaches to modelling framing providing a detailed systematic analysis of NLP and machine learning methods. We refer the reader their work for an exhaustive list of common approaches.

In contrast, we critically survey the methodological decisions along the higher-level NLP pipeline: data (Section 4.1), problem formulation (Section 4.2), and evaluation (Section 4.3), link all levels back to social science methodology, and pinpoint gaps between the two disciplines. We motivate our focus with a case study in Section 3.

We obtained a comprehensive body of literature which bridges NLP and social science work, as follows. First, we departed from two influential cross-disciplinary papers: (1) a review of media bias and framing across disciplines, but without a technical focus (Hamborg et al., 2019); and (2) one of the first and most influential NLP framing data sets, with a strong theoretical grounding (Card et al., 2015). We then identified other relevant work by following both papers' citation graphs, both inward and outward. [3] It is important to note that this survey is primarily (U.S.) English-centred, because currently-available datasets and work predominantly focus on U.S. news sources. Diversifying research to other countries, cultures, and languages is an important step for future work.

2 Background: Framing and Media Bias

Framing and *politically biased news reporting* are two strategies to systematically promote specific perspectives on contested issues. They are overlapping concepts which have been addressed jointly or with similar methods in NLP. As such, we include both strategies in this survey.

Framing has been conceptualized variously in different social science disciplines. Prevalent notions of framing include *equivalence framing:* presenting the same logical information in different forms (Cacciatore et al., 2016) and *emphasis framing:* highlighting particular aspects of an issue to promote a particular interpretation (Entman, 2007). Additionally, framing has been conceptualised as a process (de Vreese, 2005; Entman, 2007; Chong and Druckman, 2007), a communication tool (Scheufele, 1999), or a political strategy (Roy and Goldwasser, 2020). Frames have been conceptualised within different dichotomies. de Vreese (2005) distinguishes *issue-specific* and *issue-general* frames which apply to only a single or across several issues, respectively. Scheufele (1999) differentiates between *media frames*, as embedded in the political discourse, and *audience*

[2] We use the terms social sciences and communication sciences interchangeably.

[3] Note that we intentionally depart from the standard approach of selecting the top N results from Google Scholar or the ACL Anthology for few simple queries, as this would would not capture the diversity of works both in terminology and publication venues.

frames, as the reader's interpretation of an issue. Finally, Gross (2008) defines *episodic framing* as portraying an issue with an individual example compared to *thematic framing*, which takes broader context into account. Here, we cover both issue-specific and issue-generic frames and attach to Entman (2007)'s notion of emphasis framing.

While framing is a priori detached from partisan views, *political bias* refers to an explicit association of an article or media outlet with a specific political leaning. Both concepts result in biased news reporting, and correspondingly NLP researchers have attempted to address them jointly, either by investigating political framing (Roy and Goldwasser, 2020) or by identifying correlations between framing and partisan slanted articles (Ziems and Yang, 2021). NLP studies have attempted automatic media bias identification under several names, including: hyper-partisan news detection (Kiesel et al., 2019), media bias detection (Spinde et al., 2021b; Lei et al., 2022), identification of biased terms (Spinde et al., 2021a), and political ideology detection (Iyyer et al., 2014; Kulkarni et al., 2018). Their common goal is to detect and classify the bias of a data sample towards a particular political ideology. Many of these approaches naturally relate to investigate *how the story is told* (i.e., framing).

3 Three Disconnects

To illustrate the disconnects between the social sciences and NLP, we use a representative study of media bias from the communication sciences (Hernández, 2018) which investigates the framing of domestic violence in the SCMP.[4] The author formulates two research questions:

1. Framing functions: Are femicides recognized as a problem of domestic violence? What are the causes of femicides? And what are the solutions proposed?
2. Frame narratives: What are the main narratives? And what are the sources used to support them?

The first research question considers the *local* aspects within each news article. Specifically, it studies the causes and solutions presented, grounded in Entman (1993)'s conceptualisation of framing in terms of a problem, its cause, and its solution. The second research question relates these local aspects to a *global* (cross-document) view by contrasting narratives that present domestic violence as isolated incidents with those that treat it as a societal problem. It further connects the articles to *extrinsic* variables, including the sources used and cultural contexts of the story (e.g. whether the article refers the role of women in the Chinese family or understands domestic violence through the lens of the Confucian philosophy). Furthermore, the study considers articles over an extended period, capturing the *temporal development* of framing and bias. In contrast, current NLP approaches to frame prediction have predominantly adopted a single-label prediction approach per unit of analysis (Baumer et al., 2015; Naderi and Hirst, 2017; Liu et al., 2019), rather than treating frames as structures which could decompose into aspects like cause vs. solution (but see Akyürek et al. (2020); Mendelsohn et al. (2021); Frermann et al. (2023) for recent exceptions). Current approaches furthermore treat units of analysis (sentences, articles) as independent without considering links across documents, across time, or to document-external context. The multi-level and dynamic understanding of bias and framing is fundamental in the social science studies. In sum, we identify three fundamental properties of bias and framing that underpin social science research on bias and framing, and we also visually represent them in Figure 2:

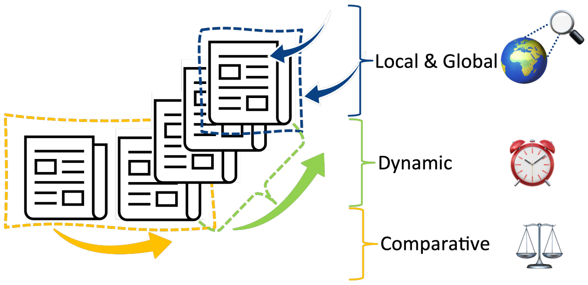

Figure 2: Illustration of the three disconnects: framing is both local and global (blue), dynamic (green) and best identified through comparative analysis (yellow).

Framing/bias is local and global It is local, because because a single document can contain several frames, and it is global because to understand the general framing of an article it is necessary to aggregate local frames and link them to document-external information such as cited (or omitted) sources, or the outlets' political leaning.

Framing/bias is dynamic Frames change over time, across outlets, or across countries or communities. Understanding the *development* of framing

[4] South China Morning Post

can shed light on the impacts of a sustained exposure to biased reporting on readers' opinions, and enables the study of trends.

Framing/bias as a comparative task Media bias and framing often become most apparent when directly contrasting articles from different perspectives, places or times (cf., Figure 1). Incorporating this contrastive notion into the task formulation – rather than predicting labels for instances in isolation – may improve induced frames in terms of quality, reliability and interpretability.

The remainder of this article links these fundamental disconnects to the more practical research design decisions that arise across both disciplines: data, methods and evaluation.

4 A Critical Review of Current Practices in NLP and Social Science

We critically compare approaches across NLP and the social sciences, pointing out discrepancies together with practical suggestions for future work.

4.1 Datasets

Social science studies are characterized by carefully collated data sets which are, however, typically small in size (<<100 articles) and manual labels are rarely released to the public. Hence we focus on limitations and opportunities of NLP framing and bias benchmarks in this section. In Table 1, we list relevant datasets, along with the type of labels they provide, the size of the collection, the associated tasks, and unit of analysis.

Media bias detection At the *sentence level*, Lim et al. (2020) used crowdsourcing to annotate sentences on 46 English-language news articles about 4 different events with four levels of bias (not-biased, slightly biased, biased, or very biased). Spinde et al. (2021b) released BABE ("Bias Annotations By Experts"), a collection of sentences labelled by experts according to binary categories: biased and non-biased, at the sentence and word levels. Fan et al. (2019) provided the BASIL ("Bias Annotation Spans on the Informational Level") dataset containing sentence (span) and word-level annotations of political leaning and sentiment (stance) towards entities in the article.

At the *document level*, the Bitterlemons corpus (Lin et al., 2006), comprises weekly issues about the Palestine–Israel conflict. Each issue contains articles from Palestinian and Israeli perspectives written by the portal's editors and guest authors. Despite being intended for document classification, this dataset can be employed to explore framing and political bias, given the documents' nature of strong bias towards one side of the conflict. Additionally, the web portal AllSides[5] categorises news outlets into three political ideologies: right, centre, and left (they also offer a finer-grained five-point scale annotation: left, lean left, centre, lean right, right) with the aim to provide all political perspectives on a given story (cf., Figure 1) including expert manual assigned categories at the article level. Several research groups have contributed datasets scraped from AllSides (Chen et al., 2018; Baly et al., 2020; Liu et al., 2022b; Lee et al., 2022).

Framing At the *headline level*, Liu et al. (2019) released the Gun Violence Frame Corpus (GVFC). It includes headlines about gun violence in news articles from 2016 and 2018 in the U.S., labelled with frames like politics, economics, and mental health. Tourni et al. (2021) released a multi-modal version of the GVFC collection, including the main image associated with each article, and annotations about relevance and framing at the image level.

At the *document level*, the Media Frames Corpus (MFC, Card et al., 2015) is the currently most extensive frame-labeled data set available. It includes articles from 13 U.S. newspapers on three policy issues: immigration, same-sex marriage, and smoking. This dataset is intended to enable the analysis of policy issue framing, providing annotations at document and span levels with frames like morality, economic, and cultural. Ziems and Yang (2021) contribute a police violence news articles collection (PVFC) that can be categorised in both domains, media bias and framing. They provide annotations for political leaning: conservative, liberal or none and also entity-centric frames, including the victim's age, race, and gender.

Opportunities for Future Work. In Section 3, we propose three main aspects to investigate framing and media bias. (1) *Conducting studies at a local and global level*. McLeod et al. (2022) suggests that framing can occur at different textual units in a document. Building on this idea, we propose a shift from single label classification on NLP datasets like AllSides, and Bitterlemons. Instead, they could be used to identify predictive sentences or spans for particular frames of political

[5] https://www.allsides.com/about

Dataset	Categories	Size	Unit of Analysis	Task
Bitterlemons (Lin et al., 2006)	Perspective (Israel, Palestine)	594	Documents	Classification
Flipper (Chen et al., 2018)	Left, Centre, Right	6,447	Documents	Classification
BASIL (Fan et al., 2019)	Liberal, Conservative, Centre; Pos, Neu, Neg	1.2k / 448 300	Spans/Words Documents	Classification
AllSides (Baly et al., 2020)	Left, Centre, Right	34k	Documents	Classification
BiasedSents (Lim et al., 2020)	not-, slightly-, very-, biased	966	Sentences	Classification
BABE (Spinde et al., 2021b)	Biased, Non-biased	3.7k	Sentences	Classification
BIGNEWSALIGN (Liu et al., 2022b)	Left, Centre, Right	1M	Documents	Classification
NeuS (Lee et al., 2022)	Left, Centre, Right	10.6k	Documents	Cross-Doc Summarisation
MFC (Card et al., 2015)	15 Frames	61.5k/ 11.9k	Sentences/ Documents	Classification
GVFC (Liu et al., 2019)	9 Frames	2.99k	Headlines	Classification
Multimodal GVFC (Tourni et al., 2021)	9 Frames	1.3k	Headlines + Images	Classification
PVFC (Ziems and Yang, 2021)	Entity frames; Conservative, Liberal, none	82k	Documents	Entity frame prediction
Narrative Frames (Frermann et al., 2023)	3 entity roles; 5 frames	428	Documents	Multi-label frame prediction

Table 1: Prominent benchmarks for political bias (top) and framing (bottom). We report size (number of data points), unit of analysis, supported task(s) and labels. All these data sets are in English and most of them U.S. centred.

biases, and investigate commonalities. This can directly inform social scientists in their analyses as well as tools to expose biases to news consumers. Roy and Goldwasser (2020) used point-wise mutual information (Church and Hanks, 1990) over bigrams and trigrams to identify spans but found poor generalization of the approach. Khanehzar et al. (2021) modelled latent frames at the event level, whith not explicit validation. The MFC contains sentence-level annotations for exploring local framing, however to the best of our knowledge no study has attempted to aggregate those labels to a global level. Regarding datasets providing sentence-level (BABE) and headline (GVFC) annotation, this can be considered as a local dimension. However, they generalise from the headline to the entire document, which ignores the subtle signals in the local dimension. (2) *The dynamics of framing* on various levels are captured by current data sets: the MFC, BASIL, GVFC and BABE provide article timestamps, supporting diachronic modeling of bias and framing. While some studies exist in this domain (Kwak et al., 2020; Card et al., 2022), the majority of NLP framing considers articles in isolation. Other dynamics, e.g., across countries, communities or media types (e.g., news vs. blogs) are of central interest in communication studies but less achievable with existing data sets. Constructing cross-language and/or cross-cultural data sets with articles aligned on the event level is an important first step. (3) *Framing as a comparative task*. We propose that researchers explore cross-document differences in their presentation of a specific issue. Several of the datasets obtained from AllSides include event-level alignment and hence enable comparison across documents on the left–centre–right spectrum at a finer granularity.

4.2 Methodologies

In NLP, researchers have approached media bias as political ideology detection or framing categorisation using different task formulations. The first and most common strategy is *single-label classification*, i.e. assigning a single label to each data point. At the *word level*, Recasens et al. (2013) learn linguistic features from word removal edit-logs in Wikipedia. Spinde et al. (2021a) compared the Euclidean distance of word embeddings to identify biased words in articles from Huffington Post (left wing) and Breitbart News (right wing). And Liu et al. (2021) experimented with identifying and replacing bias-inducing words with neutral ones using salience scores over word embeddings.

At the *sentence level*, Iyyer et al. (2014) used RNNs to identify political ideology in sentences in congressional debate transcripts and articles from the Ideological Book corpus. Using the BASIL corpus, Hartmann et al. (2019) correlated sentence and document distributions using a Gaussian mixture model (Reynolds, 2009) to identify biased sentences; Chen et al. (2020) classified biased spans by calculating their probability distributions on news articles; and Guo and Zhu (2022) applied contrastive learning and created sentence graphs to categorise biased sentences. Other researchers translated keywords from GVFC into several languages, and fine-tuned mBERT to classify frames in news headlines in languages other than English (Akyürek et al., 2020; Aksenov et al., 2021).

At the *document level*, there has been substantial work building on the MFC corpus. The task has been approached with RNNs (Naderi and Hirst, 2017), attention and discourse information (Ji and Smith, 2017), and pre-trained transformer models (Khanehzar et al., 2019). Baly et al. (2020) combined adversarial adaptation and adapted triple loss with features like Twitter and Wikipedia information about the readers and the outlet to classify the political ideology of news articles. Scholars have performed similar tasks on languages other than English, e.g. by translating English keywords in MFC to Russian to investigate the U.S. framing in Russian media over 13 years (Field et al., 2018).

Some work has formalized framing/bias detection as *multi-label classification*, typically adopting unsupervised methods like clustering (Ajjour et al., 2019) or topic modelling (Tsur et al., 2015; Menini et al., 2017) which allows to 'softly' assign documents to more than one cluster. In a supervised manner, Mendelsohn et al. (2021) employ RoBERTa to classify multiple framing typologies on immigration-related tweets. Similarly, Akyürek et al. (2020) address multi-label framing over headlines using different configurations of BERT. Both works focus on short documents (headlines or articles capped at 280 characters). The very recent work of Fremann et al. (2023) is the first to address document-level multi-label frame classification. Rather than unstructured, 'topic-like' frame detection, some works anchored framing in the depiction of important stakeholders, also referred to as *entity framing* (Ziems and Yang, 2021; Khanehzar et al., 2023).

While we focus on the task of frame and bias *detection*, NLP has also proposed methods for *bias mitigation*. Methods include the flipping of bias of headlines (from left- to right-leaning) Chen et al. (2018) or generating neutral summaries from a collection of articles with different leanings on the same topic (Lee et al., 2022). These applications come with their own sets of methodological and evaluation challenges, as well as ethical risks, and are beyond the scope of this paper. We advertise for the alternative approach of highlighting frames in multiple articles and presenting them side-by-side as illustrated in Figure 1, as a safer and potentially more effective approach in raising awareness of bias and framing.

In the social sciences, approaches tend to be manual, with fewer data samples. One common approach is to *reason across many documents from a high-level perspective*. For example, Chyi and McCombs (2004) design and evaluate a two-dimensional framework (spatial and temporal) to investigate framing changes over time in 170 news articles in American English about a U.S. school shooting event. They manually annotated articles with the signals indicating both of the frame typologies, quantified those annotations and draw conclusions about the temporal and spatial framing behaviour in the inspected articles. Muschert and Carr (2006) assessed the previously-proposed framework based on 290 news documents, and confirmed that the present temporal dimension frame still holds when using data from more than one school shooting. Hernández (2018) analysed the framing of 124 news stories from the South China Morning Post (SCMP) about femicides by manually coding the articles and quantifying those observations. The author explored whether those cases were portrayed as isolated cases or part of a systematic social problem, by manually analysing signals like narratives, sources, and the role of the entities.

In addition, communication science studies often *correlate features of news reports with extra-textual information to formulate or validate their hypotheses*. For example, McCarthy et al. (2008) assess media bias in reporting on demonstrations. They examine media coverage of protests during Belarus's transition from communism, considering factors like protest size, sponsors' status, arrests, and their correlation with media coverage. Similarly, Gentzkow and Shapiro (2010) investigate media bias by calculating think tank citation frequencies in media outlets and correlating them with U.S. Congress members mentioning the same groups.

Opportunities for Future Work. There is a stark disconnect between largely *local* approaches to frame modelling in NLP and the focus on *dynamic* and *global* questions explored in framing/bias studies in the social sciences. These arguably more complex questions emerging from the social sciences can guide the development of NLP methodologies. Capturing subtle signals, including the metaphoric or technical (legal) language use, the correlation with external features, e.g. report's sources, and the broader cultural context in which an article emerged can enrich news framing and bias analysis. On a linguistic level, framing models could be enriched with notions of metaphoric (Chakrabarty et al., 2022; Liu et al., 2022a) or subjective (Barrón-Cedeño et al., 2023) language. On the cross-document and dynamic level, we propose to address bias and frame classification as a comparative task rather than classifying documents in isolation. This can help *inducing* frames from data by analyzing axes of largest variation; and can naturally support tools and applications to raise readers' bias awareness by exposing them to contrasting perspectives on the same issue. Contextualizing framing models with extra-textual, cultural context is arguably the most challenging gap to fill. While it is tempting to suggest the use of large language models to draw some of these connections, we strongly argue for using them at most as an aid for human domain experts, and to scrutinize any automatic predictions due to the known intrinsic biases in these models.

4.3 Evaluation

We consider two levels of validation: validating data annotations, and validating model predictions.

Validating annotations Validating the quality of labelled data applies to both the social sciences and NLP. In a typical social science study, the distribution of manual labels is the main factor for accepting or rejecting hypotheses or drawing larger conclusions. As such, measures for data quality such as inter-coder reliability (ICR) are routinely reported and a core requisite of the study. This validation ensures that the codebook was correctly conceptualised, and coding often includes discussions and several iterations on trial data or pilot studies (Hernández, 2018), leading to relatively high ICR scores from carefully trained annotators, often with domain knowledge.

Validating (model) predictions Social science studies are largely analytical examining labelled data, qualitatively based on manual analysis, and quantitatively based on statistical tests. In contrast, NLP framing studies primarily rely on empirical methods, evaluating through numerical comparisons with ground truth labels. We propose a shift towards deeper insights, assessing a model's ability to capture framing and political bias on a higher, more abstract level, while also fostering fresh insights into the data. Current approaches fall short of drawing inferences from explicit information, such as assessing story objectivity and factuality. These nuanced, graded strategies require more comprehensive metrics than binary accuracy.

Opportunities for Future Work. We suggest the consequent adoption of three levels of evaluation: (1) model performance, (2) error analysis, and (3) measuring model certainty. While the three levels are by no means new, NLP work continues to focus on (1), with (2) and (3) given less thought and rigor. NLP research on media bias would benefit from established standards that guide the error analysis well as measures of model reliability and (un)certainty. Such standards might include reporting of 'most challenging' classes and/or instances; categorization of errors; as well as exploring reasons for such short comings (Vilar et al., 2006; Kummerfeld and Klein, 2013). Finally, with the increasing impact of NLP technology on the broader public, users of resulting models (be it news consumers or social science researchers), must have access model confidence scores to assess the reliability of model predictions, as per point (3).

5 Discussion

Harmonizing depth and scale The differences in data sets and evaluation between the disciplines naturally follow from their respective goals. Framing studies in the social sciences aim to uncover the principles underlying framing and its effects through careful, manual analysis of limited amounts of data, typically grounded in theoretical constructs. The primary goal of NLP in the space of media analysis is automation and scalability. Complex annotation of large training data sets as required for supervised approaches is infeasible. Besides, the required structured annotation paradigms would result in sparse observations of label co-occurrence which in turn would require even larger labelled corpora – and exploding annotation costs. Harmonizing the goal of scalability with depth and theoretical rigour is a difficult problem (that is not specific to the domain of framing and media bias). One approach towards addressing this problem is the use of semi- or unsupervised approaches, which limit the annotations to evaluation sets of more manageable size. Incorporating small amounts of labelled data with powerful pre-trained models is an obvious methodological approach, however, ensuring the validity of predictions and interplay of biases encoded in these models with the target task at hand is an open and important research problem – particularly in a sensitive domain like media bias analysis.

Feasible yet valid annotation How can we obtain ecologically valid annotations in an efficient way and sufficient quantity? We suggest to follow a common strategy in the social sciences: break articles into self-contained segments, on the event or argument level (Muschert and Carr, 2006). While recent work on argumentation in online debates has followed a similar approach of segmenting contributions into arguments and annotating frames on the argument level (Ajjour et al., 2019), it has not been applied in the news media context. Localized rather than article-level annotations have three advantages: (1) a cognitively easier task for annotators; (2) interpretability through the possibility to provide local, extractive evidence for frame predictions; and (3) a richer document-model of framing that goes beyond the single most likely frame.

Cross-disciplinary expertise for document-external grounding Section 3 pointed to a need for multi-level bias analysis, incorporating local, cross-document and broader cultural contexts. Most NLP work models individual articles without integrating external information or other articles in the collection. A few exceptions exist, including Baly et al. (2020) who incorporate readership demographics from Twitter and publisher information from Wikipedia; and Kulkarni et al. (2018) who incorporate article link structure into their models. Both works still model data points in isolation, and fall short of incorporating the more subtle cultural, political or societal contexts that inevitably interact with news framing. We argue for a strong role of cross-disciplinarity and human oversight when incorporating those factors, involving domain experts at every step from formulating research questions to model design, transparency, robustness, and evaluation. Cross-disciplinary projects would guide NLP researchers to develop novel methods that are valid and useful for studying the fundamentals of framing and media bias, and equip social scientists with enlarged data sets of high quality and relevance to enrich their research.

Open data NLP has a strong culture of sharing code and annotated data sets to encourage collaboration and reproducibility. This is less common in the humanities. Sharing this data more explicitly through cross-disciplinary dialogue could provide critical assessment and feedback from domain experts. It could drive research into combining large (and potentially noisier) data with small-scale (but high-quality) data sets from the social sciences, to address increasingly complex questions on the emergence and effects of media biases and framing.

The role of NLP in media bias analysis Despite a surge in data sets and models for automatic analysis of frames and media bias, the *ultimate goal* of these works receives surprisingly little attention. With the broader adoption of NLP methods diverse applications emerge — from supporting social scientists in scaling their research to larger data samples, to tools that highlight (or even edit) biased news to general public news consumers to expose slanted reporting. An explicit notion of goals and applications (and corresponding statement in research papers) will inform model evaluation, risks and ethical concerns to be discussed in the paper. A mandatory adoption of model cards (Mitchell et al., 2019) is one step in this direction. Irrespective of the final application of NLP research, we argue that NLP can contribute safe and valuable tools and methods only if it recognizes the complexity of bias an framing both in its data sets and annotations as well as in its evaluation procedures.

6 Conclusion

We surveyed recent work in NLP on framing and media bias, and identified disconnects and synergies in datasets, methodologies, and validation techniques to research practices in the social sciences. Despite the opportunities for NLP to support and scale social science scholarship on media bias, a current oversimplification in conceptualisation, modelling, and evaluation limits the validity and reliability of contributions. We have teased out three disconnects and proposed directions for future work, including: (1) analysing news articles from a local and global perspective, incorporating external non-textual features; (2) taking into account the dynamics of framing and bias across documents, cultures or over time; and (3) tackling the issue of media bias as a comparative task, defining frames on the basis of systematic differences between articles whose origins differ on pre-defined characteristics. This would allow for a more complex characterisation of bias than the currently dominant approach of single-label classification.

Limitations

This survey focuses on media bias and 'frame building' – i.e., the manifestation of biases and frames in news articles. This constrains the scope of our analysis to mainstream print news outlets; and leaves aside the dimension of 'frame setting' – i.e., the effects of those frames on the news consumers. Additionally, we are aware that regardless of the approach taken for sampling the body of previous work included in this paper, given the vast literature in the social sciences there will be remaining bias in our selection. With the aim of mitigating this bias, we point the reader to complement our work with previous surveys in this field i.e. Hamborg et al. (2019) and Ali and Hassan (2022).

Ethics Statement

Identifying framing and political bias in news articles is a sensitive application area, and inevitably influenced by social and structural biases in the academic investigators and the pool of annotators. Datasets and technologies intending to tackle these phenomena comprise the social bias of annotators and researchers developing them in an environment lacking diversity. Besides there is a potential for dual use of models and benchmarks to promote polarisation and misinformation through framing, rather than reduce it. We see this paper as an opportunity to identify new directions to diversify NLP methodologies and data sets, grounded in best-practices from the media sciences which have been developed for decades. We anticipate that these steps will, together with a better documentation of models and intended use cases, will help to address the above concerns.

References

Yamen Ajjour, Milad Alshomary, Henning Wachsmuth, and Benno Stein. 2019. Modeling frames in argumentation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2922–2932, Hong Kong, China. Association for Computational Linguistics.

Dmitrii Aksenov, Peter Bourgonje, Karolina Zaczynska, Malte Ostendorff, Julian Moreno-Schneider, and Georg Rehm. 2021. Fine-grained classification of political bias in German news: A data set and initial experiments. In *Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)*, pages 121–131, Online. Association for Computational Linguistics.

Afra Feyza Akyürek, Lei Guo, Randa Elanwar, Prakash Ishwar, Margrit Betke, and Derry Tanti Wijaya. 2020. Multi-label and multilingual news framing analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8614–8624, Online. Association for Computational Linguistics.

Mohammad Ali and Naeemul Hassan. 2022. A survey of computational framing analysis approaches. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9335–9348, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Ramy Baly, Giovanni Da San Martino, James Glass, and Preslav Nakov. 2020. We can detect your bias: Predicting the political ideology of news articles. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4982–4991, Online. Association for Computational Linguistics.

Alberto Barrón-Cedeño, Firoj Alam, Tommaso Caselli, Giovanni Da San Martino, Tamer Elsayed, Andrea Galassi, Fatima Haouari, Federico Ruggeri, Julia Maria Struß, Rabindra Nath Nandi, et al. 2023. The clef-2023 checkthat! lab: Checkworthiness, subjectivity, political bias, factuality, and authority. In *European Conference on Information Retrieval*, pages 506–517. Springer.

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 the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1472–1482, Denver, Colorado. Association for Computational Linguistics.

Michael A Cacciatore, Dietram A Scheufele, and Shanto Iyengar. 2016. The end of framing as we know it... and the future of media effects. *Mass communication and society*, 19(1):7–23.

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 Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 438–444, Beijing, China. Association for Computational Linguistics.

Dallas Card, Serina Chang, Chris Becker, Julia Mendelsohn, Rob Voigt, Leah Boustan, Ran Abramitzky, and Dan Jurafsky. 2022. Computational analysis of 140 years of US political speeches reveals more positive but increasingly polarized framing of immigration.

Proceedings of the National Academy of Sciences, 119(31):e2120510119.

Tuhin Chakrabarty, Yejin Choi, and Vered Shwartz. 2022. It's not rocket science: Interpreting figurative language in narratives. *Transactions of the Association for Computational Linguistics*, 10:589–606.

Wei-Fan Chen, Khalid Al Khatib, Benno Stein, and Henning Wachsmuth. 2020. Detecting media bias in news articles using Gaussian bias distributions. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4290–4300, Online. Association for Computational Linguistics.

Wei-Fan Chen, Henning Wachsmuth, Khalid Al-Khatib, and Benno Stein. 2018. Learning to flip the bias of news headlines. In *Proceedings of the 11th International Conference on Natural Language Generation*, pages 79–88, Tilburg University, The Netherlands. Association for Computational Linguistics.

Dennis Chong and James N. Druckman. 2007. Framing theory. *Annual Review of Political Science*, 10(1):103–126.

Kenneth Ward Church and Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1):22–29.

Hsiang Iris Chyi and Maxwell McCombs. 2004. Media salience and the process of framing: Coverage of the Columbine school shootings. *Journalism & Mass Communication Quarterly*, 81(1):22–35.

Alexander Dallmann, Florian Lemmerich, Daniel Zoller, and Andreas Hotho. 2015. Media bias in German online newspapers. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, HT '15, page 133–137, New York, NY, USA. Association for Computing Machinery.

Claes H. de Vreese. 2005. News framing: Theory and typology. *Information Design Journal*, 13(1):51–62.

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

Robert M. Entman. 2007. Framing bias: Media in the distribution of power. *Journal of Communication*, 57(1):163–173.

Lisa Fan, Marshall White, Eva Sharma, Ruisi Su, Prafulla Kumar Choubey, Ruihong Huang, and Lu Wang. 2019. In plain sight: Media bias through the lens of factual reporting. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6343–6349, Hong Kong, China. Association for Computational Linguistics.

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 the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3570–3580, Brussels, Belgium. Association for Computational Linguistics.

Lea Frermann, Jiatong Li, Shima Khanehzar, and Gosia Mikolajczak. 2023. Conflicts, villains, resolutions: Towards models of narrative media framing. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8712–8732, Toronto, Canada. Association for Computational Linguistics.

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

Salma Ghanem. 1997. *Filling in the Tapestry: The Second Level of Agenda Setting*, 1 edition, pages 3–14. Routledge, New York.

David Giles and Rachel L Shaw. 2009. The psychology of news influence and the development of media framing analysis. *Social and personality psychology compass*, 3(4):375–393.

Kimberly Gross. 2008. Framing persuasive appeals: Episodic and thematic framing, emotional response, and policy opinion. *Political Psychology*, 29(2):169–192.

Shijia Guo and Kenny Q. Zhu. 2022. Modeling multi-level context for informational bias detection by contrastive learning and sentential graph network. *CoRR*, abs/2201.10376.

Felix Hamborg. 2020. Media bias, the social sciences, and NLP: Automating frame analyses to identify bias by word choice and labeling. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 79–87, Online. Association for Computational Linguistics.

Felix Hamborg, Karsten Donnay, and Bela Gipp. 2019. Automated identification of media bias in news articles: An interdisciplinary literature review. *International Journal on Digital Libraries (IJDL)*, 20(4):391–415.

Mareike Hartmann, Tallulah Jansen, Isabelle Augenstein, and Anders Søgaard. 2019. Issue framing in online discussion fora. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1401–1407, Minneapolis, Minnesota. Association for Computational Linguistics.

Miriam Hernández. 2018. "Killed out of love": A frame analysis of domestic violence coverage in Hong Kong. *Violence Against Women*, 24(12):1454–1473.

Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. 2014. Political ideology detection using recursive neural networks. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1113–1122, Baltimore, Maryland. Association for Computational Linguistics.

Yangfeng Ji and Noah A. Smith. 2017. Neural discourse structure for text categorization. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 996–1005, Vancouver, Canada. Association for Computational Linguistics.

Shima Khanehzar, Trevor Cohn, Gosia Mikolajczak, and Lea Frermann. 2023. Probing power by prompting: Harnessing pre-trained language models for power connotation framing. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 873–885, Dubrovnik, Croatia. Association for Computational Linguistics.

Shima Khanehzar, Trevor Cohn, Gosia Mikolajczak, Andrew Turpin, and Lea Frermann. 2021. Framing unpacked: A semi-supervised interpretable multi-view model of media frames. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2154–2166, Online. Association for Computational Linguistics.

Shima Khanehzar, Andrew Turpin, and Gosia Mikolajczak. 2019. Modeling political framing across policy issues and contexts. In *Proceedings of the The 17th Annual Workshop of the Australasian Language Technology Association*, pages 61–66, Sydney, Australia. Australasian Language Technology Association.

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. Association for 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 the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3518–3527, Brussels, Belgium. Association for Computational Linguistics.

Jonathan K. Kummerfeld and Dan Klein. 2013. Error-driven analysis of challenges in coreference resolution. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 265–277, Seattle, Washington, USA. Association for Computational Linguistics.

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. In *12th ACM Conference on Web Science*, page 305–314, Southampton United Kingdom. ACM.

Nayeon Lee, Yejin Bang, Tiezheng Yu, Andrea Madotto, and Pascale Fung. 2022. NeuS: Neutral multi-news summarization for mitigating framing bias. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3131–3148, Seattle, United States. Association for Computational Linguistics.

Yuanyuan Lei, Ruihong Huang, Lu Wang, and Nick Beauchamp. 2022. Sentence-level media bias analysis informed by discourse structures. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 10040–10050, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Sora Lim, Adam Jatowt, Michael Färber, and Masatoshi Yoshikawa. 2020. Annotating and analyzing biased sentences in news articles using crowdsourcing. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 1478–1484, Marseille, France. European Language Resources Association.

Wei-Hao Lin, Theresa Wilson, Janyce Wiebe, and Alexander Hauptmann. 2006. Which side are you on? identifying perspectives at the document and sentence levels. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)*, pages 109–116, New York City. Association for Computational Linguistics.

Emmy Liu, Chenxuan Cui, Kenneth Zheng, and Graham Neubig. 2022a. Testing the ability of language models to interpret figurative language. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4437–4452, Seattle, United States. Association for Computational Linguistics.

Ruibo Liu, Lili Wang, Chenyan Jia, and Soroush Vosoughi. 2021. Political depolarization of news articles using attribute-aware word embeddings. In *Proceedings of the Fifteenth International AAAI Conference on Web and Social Media, ICWSM 2021, held virtually, June 7-10, 2021*, pages 385–396. AAAI Press.

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 the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 504–514, Hong Kong, China. Association for Computational Linguistics.

Yujian Liu, Xinliang Frederick Zhang, David Wegsman, Nicholas Beauchamp, and Lu Wang. 2022b. POLITICS: Pretraining with same-story article comparison for ideology prediction and stance detection. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 1354–1374, Seattle, United States. Association for Computational Linguistics.

John McCarthy, Larissa Titarenko, Clark McPhail, Patrick Rafail, and Boguslaw Augustyn. 2008. Assessing stability in the patterns of selection bias in newspaper coverage of protest during the transition from communism in Belarus. *Mobilization: An International Quarterly*, 13(2):127–146.

Douglas M McLeod, Hyesun Choung, Su Min-Hsin, Kim Sang-Jung, Ran Tao, Jiawei Liu, and ByungGu Lee. 2022. Navigating a diverse paradigm: A conceptual framework for experimental framing effects research. *Review of communication research*, 10.

Denis McQuail and Mark Deuze. 2020. *Mcquail's media and mass communication theory*, 7th ed edition. SAGE Publications, Thousand Oaks.

Julia Mendelsohn, Ceren Budak, and David Jurgens. 2021. Modeling framing in immigration discourse on social media. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2219–2263, Online. Association for Computational Linguistics.

Stefano Menini, Federico Nanni, Simone Paolo Ponzetto, and Sara Tonelli. 2017. Topic-based agreement and disagreement in US electoral manifestos. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2938–2944, Copenhagen, Denmark. Association for Computational Linguistics.

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 220–229.

Glenn W. Muschert and Dawn Carr. 2006. Media salience and frame changing across events: Coverage of nine school shootings, 1997–2001. *Journalism & Mass Communication Quarterly*, 83(4):747–766.

Nona Naderi and Graeme Hirst. 2017. Classifying frames at the sentence level in news articles. In *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*, pages 536–542, Varna, Bulgaria. INCOMA Ltd.

Sylvain Parasie. 2022. *Computing the news : data journalism and the search for objectivity*. Columbia University Press New York, New York.

Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2013. Linguistic models for analyzing and detecting biased language. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1650–1659, Sofia, Bulgaria. Association for Computational Linguistics.

Douglas A. Reynolds. 2009. Gaussian mixture models. In Stan Z. Li and Anil K. Jain, editors, *Encyclopedia of Biometrics*, pages 659–663. Springer US.

Shamik Roy and Dan Goldwasser. 2020. Weakly supervised learning of nuanced frames for analyzing polarization in news media. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7698–7716, Online. Association for Computational Linguistics.

Dietram A. Scheufele. 1999. Framing as a theory of media effects. *Journal of Communication*, 49(1):103–122.

Timo Spinde, Felix Hamborg, Lada Rudnitckaia, and Bela Gipp. 2021a. Identification of biased terms in news articles by comparison of outlet-specific word embeddings. In *Proceedings of the iConference 2021*.

Timo Spinde, Manuel Plank, Jan-David Krieger, Terry Ruas, Bela Gipp, and Akiko Aizawa. 2021b. Neural media bias detection using distant supervision with BABE - bias annotations by experts. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 1166–1177, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Isidora Tourni, Lei Guo, Taufiq Husada Daryanto, Fabian Zhafransyah, Edward Edberg Halim, Mona Jalal, Boqi Chen, Sha Lai, Hengchang Hu, Margrit Betke, Prakash Ishwar, and Derry Tanti Wijaya. 2021. Detecting frames in news headlines and lead images in U.S. gun violence coverage. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 4037–4050, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Oren Tsur, Dan Calacci, and David Lazer. 2015. A frame of mind: Using statistical models for detection of framing and agenda setting campaigns. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1629–1638, Beijing, China. Association for Computational Linguistics.

David Vilar, Jia Xu, Luis Fernando D'Haro, and Hermann Ney. 2006. Error analysis of statistical machine translation output. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy. European Language Resources Association (ELRA).

Alden Williams. 1975. Unbiased study of television news bias. *Journal of Communication*, 25(4):190–199.

Caleb Ziems and Diyi Yang. 2021. To protect and to serve? analyzing entity-centric framing of police violence. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 957–976, Punta Cana, Dominican Republic. Association for Computational Linguistics.