Mapping the Media Landscape: Predicting Factual Reporting and Political Bias Through Web Interactions

Dairazalia Sánchez-Cortés[1]([✉]) [iD], Sergio Burdisso[1] [iD], Esaú Villatoro-Tello[1] [iD],
and Petr Motlicek[1,2] [iD]

[1] Idiap Research Institute, Martigny, Switzerland
{dairazalia.sanchez-cortes,sergio.burdisso,esau.villatoro,
petr.motlicek}@idiap.ch
[2] Brno University of Technology, Brno, Czech Republic

Abstract. Bias assessment of news sources is paramount for professionals, organizations, and researchers who rely on truthful evidence for information gathering and reporting. While certain bias indicators are discernible from content analysis, descriptors like political bias and fake news pose greater challenges. In this paper, we propose an extension to a recently presented news media reliability estimation method that focuses on modeling outlets and their longitudinal web interactions. Concretely, we assess the classification performance of four reinforcement learning strategies on a large news media hyperlink graph. Our experiments, targeting two challenging bias descriptors, factual reporting and political bias, showed a significant performance improvement at the source media level. Additionally, we validate our methods on the CLEF 2023 *Check-That!* Lab challenge, outperforming the reported results in both, F1-score and the official MAE metric. Furthermore, we contribute by releasing the largest annotated dataset of news source media, categorized with factual reporting and political bias labels. Our findings suggest that profiling news media sources based on their hyperlink interactions over time is feasible, offering a bird's-eye view of evolving media landscapes.

Keywords: news media profiling · media bias descriptors · factual reporting · political bias

1 Introduction

Given its open and distributed nature, the World Wide Web (WWW) has become the main information source worldwide, democratizing content creation and making it easy for everybody to share and spread information online. On the bright side, this phenomenon enables a faster dissemination of information compared to what was possible with traditional newspapers, radio, and TV. On the downside, at the moment of removing the "gate-keeper" role from traditional

media, it opens the door for additional problems, e.g., the spread of misinformation, at breaking-news speed, that can potentially mislead the users and even impact their behavior [3,25].

Thus, while the goal of this democratic channel is to provide users with the necessary tools to acquire greater knowledge about a topic, the reality is that in the way this knowledge (i.e., news) is presented and reported is not necessarily always impartial [2,18], and there is a growing concern regarding the biases of different media outlets when reporting specific events [14]. For example, in polarizing topics like politics, many of the news can be biased towards one political perspective or the other, i.e., *political bias*, which may influence citizens' voting decisions and preferences of undecided individuals [13].

To mitigate the impact of misinformation and to favor critical assessment for the newsreaders, independent bias assessment services like MBFC[1] and allsides[2] perform information verification. The review process is performed manually by professionals at the event or article level, clearly this is a challenging schema to maintain on the long term given the fast-speed proliferation of both news media websites and news articles. Automation comes handy to perform certain fact-checking tasks, like gathering information (e.g. articles with similar topics, metadata on the media-publisher, etc.); for the more complex parts of the verification analysis, advances in AI continues pushing the boundaries in order to provide valuable tools (for example, search and retrieval, summarization, transformers, LM and LLMs). While the latest LLMs performance on several tasks is remarkable, they are still prone to carry unauthenticated information [17].

While many existing tools are being adopted to support verification tasks at the article level (with and without human supervision), there are very few advances to fully automate news media profiling at the source level (other than popularity). Previous research has shown evidence that some news bias descriptors can be inferred by just inspecting the outlet website metadata [10,16]. Other approaches have addressed source reliability, factuality of reporting or political bias, by assembling information from multiple external and social media sources, metadata and/or content-based features [3–5,7,9,21,22]. Unfortunately, methodologies relying on social media metadata can not longer be reproduced at scale given the current access restrictions.

A recent research shifting from the social media and text-based approach, is presented in [8]. Burdisso *et al.*, proposed a highly performing and robust graph-based methodology to score news media reliability. Their method considers the longitudinal interactions on the web to learn a reliability value from their source neighbors. Based on the research evidence that neighboring properties can be spread among news media outlets, we extend their work and we propose to address the following research question: to what extent it is possible to profile news media outlets (i.e., different properties) based solely on their interaction with other media sources? To address this question, in this paper we focus on two challenging media bias descriptors: factuality of reporting and political bias.

We choose to extend Burdisso *et al.* methodology given that it is both language and content independent (political, religious, racial, etc.), it can be applied at a larger scale and their 17k English news outlets dataset is publicly available.

Our main contributions are as follows: (a) we show that it is possible to predict/estimate bias descriptors, i.e., political bias and factual reporting; of the source media based on their interactions with other sources (outperforming the baseline); (b) we validate the robustness of our approach on the publicly available dataset from the CLEF CheckThat! challenge, specifically collected to classify political bias with currently active news outlets, and we established a new SOTA result; (c) we release the biggest dataset at the source media level with standard political bias and factual reporting labels[3].

2   Related Work

The bias in news media is a pervasive and ubiquitous problem [14,15,25]. The need for applied research on news media descriptors has increased since 2000 due to the generalized adoption of social media platforms, and the proliferation of tools that facilitate both websites and news-content creation [6,12]. Bias in news media has a wide descriptors spectrum [14,15,27], for example Racial Bias refers to preferences of coverage or not of events related to minorities or group of individuals [24]. Gender Bias refers to the inclination towards one gender over another, resulting in unequal treatment, coverage and perception [1,23]. Political Bias, refers to partial representation of political issues or tendency to favor a particular political ideology.

A significantly large NLP community has reported advances on news media bias at the article level (i.e., based on content), also referred as bias at the event-label or a short-term bias on a selected event [14]. However, in this paper, we contribute towards the news media source profiling (i.e., at the source level). We focus our research work on the following two long-term bias descriptors:

Factual Reporting. Recent task challenges, particularly the *CheckThat!* Lab challenge at CLEF 2023, have addressed the factuality of reporting based on three classes (*High*, *Mixed* and *Low*) at the article level [21]. Submitted models range from traditional supervised models (such as SVMs, Random Forest, gradient Boost) to Deep Learning-based ones [19,21]. Due to the challenging nature to perform factuality assessment, graph-based models emerged to address the problem disclosing better performance when combined with text-based approaches [3,4,11,22]. Fairbanks *et al.*, [11] proposed a structural model based on the metadata from the article's news web links. Their findings revealed that credibility, a descriptor in close relation with factual reporting, is harder to determine from merely the content. Baly *et al.* [3] analyzed the factual reporting focusing on the source media. Their approach used text-based features from articles content and metadata including Wikipedia pages, Twitter, URL-related features (domain, orthography, char n-grams), and Web traffic (Alexa service). Also targeting the factuality at the media level, Panayotov *et al.*, [22] proposed

---

[3] https://github.com/idiap/Factual-Reporting-and-Political-Bias-Web-Interactions.

to model the factuality of reporting using graph neural network and similarity between news media based on their audience overlap. Although the latest models revealed significant improvements at the media level, the methods in [3,22] rely on the Alexa website ranking and web traffic information, which is now discontinued.

More recent approaches are focusing on state-of-the-art LMs and LLMs, from adversarial training, ensemble of models based on RoBERTa or GPTs [20,26]. Li *et al.*, heuristics on adversarial training revealed the importance of semantics in the title and the summary of the news captured at the beginning and end of the article. Their best performing political inference results from a majority voting from four implemented models from which, two are RoBERTa-based. Tran *et al.*, examined the impact of imbalanced training data between *High*, *Mixed* and *Low* factual reporting. The authors introduced a RoBERTa-based back-translation framework that significantly surpassed the baseline performance. Their approach ranked among the top three performers at the *CheckThat!* Lab challenge in 2023. To the best of our knowledge, the state-of-the-art methodology in media profiling, outperforming ensembles of content-based and external data was recently introduced in [8]. Burdisso *et al.*, propose an hyperlink-interactions graph to infer News source reliability degree (a continuous value) based on reinforcement learning techniques. In addition to the standing performance, authors contribute with the largest reported dataset in source media profiling with 17k English-speaking news outlets.

Political Bias. In the recent years, the inference of political bias at the outlet level has been approached by applying SVMs, CatBoost and applied oversampling techniques, mostly enhancing content-features from articles [2,4,9]. Baly *et al.*, [3,4] proposed a framework based on SVMs reporting significant results when complementing content-based data with Wikipedia and social media metadata. Recently, Azizov *et al.*, [2] proposed a majority voting ensemble of CatBoost models and TF-IDF, showing better performance than LM-frameworks at the *CheckThat!* lab challenge at CLEF 2023 [9] given a benchmark dataset with three political classes (*Left*, *Center*, *Right*). In Panayotov *et al.* [22], the political bias was modeled using a graph neural network augmented with audience/social media data. Graph-based approaches showed evidence that metadata capturing information other than the article content improved classification of political stance. Given the still open challenge to accurately infer political bias at the news source level, more recent approaches are exploring the pertinence of using LMs [26,27]. Tran *et al.* [26], analyzed and addressed the three-class (*Left*, *Center*, *Right*) imbalance by translating to Spanish and back to English the classes with less articles. Then, they fine-tuned RoBERTa English-large, and performed a majority voting at the article-level to infer the news source political leaning, showing a significant performance above the baseline. Wessel *et al.*, [27] proposed a framework using transformers to infer 9 bias descriptors. For the case of political bias, the original bias annotation provided at the outlet level is transformed into two classes bias and not-bias. Despite the 2 million political news articles used in this work, they were exclusively gathered from the top 11 most popular US media outlets. Authors concluded that cognitive and political bias at

the content-level are the most challenging bias descriptors to detect, in contrast with for example gender or racial bias.

Although some approaches show significant improvement over majority baselines, the robustness and scalability of the models is not sufficient to consider the factual reporting and political bias problem solved. Contrary to previous research that depends on content, audience feedback, and/or metadata, in this paper we extend a very recent work that models the problem in a scalable fashion relaying on network interactions among the news sources [8]. Following sections describe the proposed methodology and obtained results.

3 Methodology and Strategies

In order to validate our research question and based on the evidence presented by Burdisso *et al.* that longitudinal interactions can spread the news media reliability degree among their neighbors [8], we extend their work to address factual reporting and political bias.

The introduced approach consists of first building a news media graph from the WWW and then applying different reinforcement learning strategies to infer the reliability values. More precisely, constructing a weighted directed graph $G = \langle S, E, w \rangle$ where there is an edge $(s, s') \in E$ if source s contains articles (hyper) linked to s' and where the weight $w(s, s') \in [0, 1]$ is the proportion of total hyperlinks in s linked to s'.[4] In this work, we hypothesize that the political bias and factual reporting of sources s can be estimated from the sources it interacts with, by inheriting their properties.

Following the original work in [8], we model the estimation as a Markov Decision Process (MDP) $\langle \mathbb{S}, A, P, r \rangle$ such that: (1) The set of states $\mathbb{S}$ are all the news outlets websites—*i.e.* $\mathbb{S} = S$; (2) The set of actions A contains only one element, the *"move to a different news media website"* action; (3) The probability P of moving from the origin s to s' will be given by the proportion of hyperlinks in s connecting to s'—*i.e.* we have $P(s, s') = w(s, s')$; and (4) The reward r of moving to another news source (s') is determined only by the origin source(s), and it will be positive or negative depending on the known property—*e.g.* $r(s) = 1$ if we know for this s we have *Right* or *High*, for political bias or factual reporting, respectively; $r(s) = -1$ if s is *Left* or *Low*, for political bias or factual reporting, respectively; $r(s) = 0$ otherwise. Finally, the property (political bias or factual reporting level) value for all news sources s in the graph will be estimated by a function $\rho(s)$ following 4 different strategies:

– **F**-*property*: The property value is proportional to the *expected* perceived reward given by the following Bellman equation where π is the unique policy

---

[4] Note that this simple hyperlink-based representation is also implicitly capturing content-based references to and from other sources.

(i.e. the probability of taking action $a \in A$ in state s) and $\gamma \in [0, 1)$ the discount factor:[5]

$$\rho(s) = \sum_{s' \in \mathbb{S}} P^{\pi}(s, s')[r(s') + \gamma \rho(s')] \tag{1}$$

That is, under this strategy, the value of source s will be inherent from the sources it connects in the **Future**.

– **P**-*property*: The property value is interpreted a proportion of the accumulated perceived reward, i.e., the value is inherited by the sources that lead to it in the **P**ast. The value is thus, giving by the following the reverse Bellman equation:

$$\rho(s) = r(s) + \gamma \sum_{s' \in \mathbb{S}} P^{\pi}(s', s)\rho(s') \tag{2}$$

– **FP**-*property*: This strategy combines the previous two strategies by considering **F**uture and **P**ast information. A source s increases its positive value $\rho(s)$ as more positive sources link to it ($\rho_{\mathbf{P}}^{+}(s)$), while losing value as it links to more negative sources ($\rho_{\mathbf{F}}^{-}(s)$).[6] Thus, $\rho(s)$ is simply defined as:

$$\rho(s) = \rho_{\mathbf{F}}^{-}(s) + \rho_{\mathbf{P}}^{+}(s) \tag{3}$$

– **I**-*property*: **I**nvestment Strategy (invest and collect credits) consisting of two iterative steps, repeated n times: (1) all sources invest their property value to the neighboring sources proportionally to the strength of their links ($w(s, s')$) following Eq. 4, (2) sources collect the credits back proportionally to the investment and update its own property value following Eq. 5.

$$totalcredits(s) = \sum_{s' \in S} w(s', s) \cdot \rho(s') \tag{4}$$

$$\rho(s) = \rho(s) + \sum_{s' \in S} w(s, s') \cdot credits_s(s') \tag{5}$$

where credits are distributed among investors s', in proportion to their contribution to s, i.e., $credits_{s'}(s) = w_{s'}(s) \cdot totalcredits(s)$.

3.1   Datasets

There are several attempts to unify existing datasets to assess Bias in news media. Recently, a unified bias dataset was presented including several Bias descriptors [27], nevertheless, the collection of articles, sentences, comments, etc., are on one hand targeting rather short-term bias (text-based), and on the other hand large part of the data do not have URLs to existing news media

---

[5] The discount factor controls the distance of looking back/forward; $\gamma \approx 0$ focuses mostly on present reward $r(s)$, while $\gamma \approx 1$ considers all history/future to compute $\rho(s)$.

[6] Algorithm 1 and Algorithm 2 in [8] detail how these updates are applied.

Table 1. Label distribution on both datasets.

Dataset	Political Bias			Factual Rep.			Total
	Left	Center	Right	Low	Mixed	High	
MBFC (ours)	2078	763	1079	408	1391	2121	3920
CLEF *CheckThat!*	272	359	392	-	-	-	1023

sources. Recently, [8] released the largest dataset with URLs annotated with reliability labels constructed by collecting and consolidating annotations from different sources. In this work, we follow a similar process as described by the authors in [8] to build our own dataset with political bias and factual reporting annotation, which we refer to as "MBFC".

MBFC. Following the methodology described in [8], we crawled 3920 news media URL domains from the *Media Bias/Fact Check* (MBFC)[4] service including annotated bias descriptors that are further transformed and normalized into political and factual reporting labels as follows: for Political bias, the final normalized categories are *Left, Center, Right*; for the case of Factual Reporting, labels include *High* (which aggregated high and very high), *Mixed* and *Low* (which aggregates low and very low).

CLEF *CheckThat!* Additionally, in order to compare results with previously published approaches, we use the dataset released for the CLEF 2023 *Check-That!* lab which focused on political bias identification. This dataset contains a total of 1023 news media URL domains with political bias labels crawled from *allsides*[5], a website that gathers news articles with balanced representation of the different political perspectives. The data is officially divided into fixed train, dev, and test set splits containing 817 *(Left-216, Center-296 and Right-305)*, 104 *(Left-31, Center-34 and Right-39)*, and 102 *(Left-25, Center-29 and Right-48)* news sources, respectively. More details about the data and the labeling process can be found in [9]. Table 1 summarizes the label distribution and size of both introduced datasets.

4 Experiments and Results

In this work we used the graph G built in [8] consisting of 17K news sources obtained after processing 100M news articles from Common Crawl News. Following [3,8] we report 5-fold cross-validation evaluation results on our MBFC datasets, whereas for CLEF's *CheckThat!* we report results on the official test set. In order to estimate the factual score of reporting from the graph, we first convert the factuality/bias ground truth labels from the training set into rewards as follows: $r(s) = 1$ if the media label is *High/Right*, $r(s) = -1$ if *Low/Left*, and $r(s) = 0$ otherwise. Then, at inference time, sources s are classified with the label *Right/High* if $\rho(s) > 0$ and *Left/Low* otherwise. Even though one limitation of the proposed strategies is that they are essentially binaries, in order to

Table 2. 5-fold cross-validation average results for Political Bias and Factual Reporting classification. The best-performing values are <u>**underlined**</u>, while the 2nd-best results appear in **bold** font.

Task	Strategy	F_1 score			Accuracy
		Macro avg.	High/Righ	Low/Left	
Factual Rep.	*Majority*	38.94 ± 0.04	87.88 ± 0.09	0.00 ± 0.00	83.84 ± 0.16
	Random	36.44 ± 0.88	65.69 ± 1.64	7.18 ± 1.45	49.93 ± 1.69
	F-Factuality	57.60 ± 4.38	95.00 ± 0.97	20.19 ± 7.86	90.60 ± 1.76
	P-Factuality	$\mathbf{85.13} \pm 2.73$	$\mathbf{98.70} \pm 0.35$	$\mathbf{71.55} \pm 5.15$	$\mathbf{97.52} \pm 0.66$
	FP-Factuality	71.35 ± 2.33	96.76 ± 0.65	45.93 ± 4.09	93.89 ± 1.19
	I-Factuality	$\underline{\mathbf{87.99}} \pm 4.60$	$\underline{\mathbf{99.02}} \pm 0.43$	$\underline{\mathbf{76.96}} \pm 8.79$	$\underline{\mathbf{98.12}} \pm 0.81$
Political Bias	*Majority*	38.04 ± 0.07	0.00 ± 0.00	76.08 ± 0.14	65.40 ± 0.18
	Random	45.42 ± 1.84	30.29 ± 2.86	60.55 ± 1.75	49.65 ± 1.73
	F-Political	60.42 ± 3.74	41.56 ± 6.27	79.29 ± 1.42	69.44 ± 2.30
	P-Political	$\mathbf{74.08} \pm 2.31$	$\mathbf{65.80} \pm 3.23$	$\mathbf{82.36} \pm 1.39$	$\mathbf{76.73} \pm 1.95$
	FP-Political	64.90 ± 3.15	52.47 ± 4.82	77.33 ± 1.94	69.34 ± 2.55
	I-Political	$\underline{\mathbf{77.77}} \pm 2.45$	$\underline{\mathbf{70.97}} \pm 3.39$	$\underline{\mathbf{84.56}} \pm 1.54$	$\underline{\mathbf{79.85}} \pm 2.12$

compare results in *CheckThat!* three-label classification task, we use the official dev set to find an ϵ value to classify sources s as follow: *Left/Low* if $\rho(s) < -\epsilon$; *Right/High* if $\rho(s) > \epsilon$; *Center/Mixed* otherwise. More precisely, we selected the hyper-parameters $\epsilon = 3\mathrm{e}{-3}$, $\gamma = 0.15$ (Eq. 1, 2, 3), and $n = 2$ (Eq. 5) after performing a grid search maximizing the Macro avg. F1 score with $\epsilon \in [1\mathrm{e}{-3}, 1\mathrm{e}{-1}]$ (1e−3 increments), $\gamma \in [0.05, 0.95]$ (0.05 increments), $n \in [1, 10]$, respectively.

4.1 Factuality of Reporting

Table 2 shows the results from the 5-fold cross-validation for Factual Reporting. The baseline for comparison includes Random and Majority class classification. The **F**-Factuality strategy performed at 57.60 F1-score overall, for the individual classes *High* Factual reporting performance is 95.00 F1-score and 20.19 for *Low* Factual Reporting. For all cases there is significant improvement with respect to the baselines. For **P**-Factuality F1-score performance is 85.13, and 98.7 and 71.55 for the *High* and *Low* classes. The significantly high performance reveals that indeed the graph with past reward strategy captures close interacting networks on both sides, *High* score and *Low* score of factual reporting. The strategy **FP**-Factuality performs at 71.35 F1-score, although it outperforms **F**-Factuality and the baselines, it remains behind **P**-Factuality. Finally, the **I**-Factuality strategy outperforms all the other strategies up to 87.99 F1-score, 76.96 for class Low and 99.02 for the class High. The results show that for the case of **I**-Factuality (the invest and collect strategy), the gathered information from the hyperlinks and its neighbors can accurately capture the level of factuality, significantly better for the class Low.

Table 3. Results on CLEF's *CheckThat!* dataset on Political Bias of news media. MAE: Mean Absolute Error. The smaller MAE value translates into better predictions.

Team	MAE($\downarrow$)	F_1 score($\uparrow$)	Accuracy($\uparrow$)
Baseline [9]	0.902	-	-
Awakened	0.765	-	-
Accenture [26]	0.549	0.625	0.627
Frank [2]	0.320	0.727	0.725
F-Political	0.333	0.632	0.667
P-Political	**0.238**	**0.760**	**0.762**
FP-Political	0.309	0.670	0.690
I-Political	**0.214**	**0.784**	**0.786**

I-Factuality accurately identifies almost all sources with *Low* Factual Reporting, which is indeed a key contribution of this paper. We assume that high performance of the reward value might be due to capturing unintentionally the lifespan of a news media domain, which has been reported as a high contributor in the identification of disinformative websites [16]. Both strategies **P-**Factuality and **I-**factuality are highly performing on F1-score and Accuracy, similarly to findings on Reliabilty of news media in [8], disclosing an accurate profiling of Bias given only their network interactions overtime.

4.2   Political Bias

Results on MBFC. Table 2 shows the 5-fold cross-validation results for F1-score and Accuracy, we included two baselines Random, and Majority class for comparison. For the political leaning the **F-**Political performs at 60.42 F1-score, and 79.29 F1-score for Right at the class level, showing a modest improvement over the baseline (76.08). For **P-**Political the overall F1-score performance is 74.08, with 65.8 for the class *Left* and 82.36 for the class *Right*. For the combined **FP-**Political the F1-score of 64.90 outperforms the **F-**Political but does not improve the **P-**Political performance, for both the overall and the class level, which indicates that past information contributes more to the predictions. The best performing strategy is **I-**Political performing at 77.77 F1-score and, 70.97 and 84.56 for the classes *Left* and *Right* respectively. At the class level, our results on political bias show significantly better performance on the class *Right*. Figure 1 shows part of the graph for the news media source `www.newrepublic.com`, where the values are estimated with **I-**political. The size of the node is proportional to their political bias, as newrepublic predominantly engages with *Left*-wing sources, its final value leaned significantly towards the *Left* (red).

Results on the CLEF *CheckThat!* Table 3 shows the F1-score performance and the official scoring metric MAE (Mean Absolute Error) for the Labs at CLEF 2023. The political labels were coded as ordinal values (*Left*-0, *Center*-1, *Right*-2), a smaller MAE value translates into better predictions from the
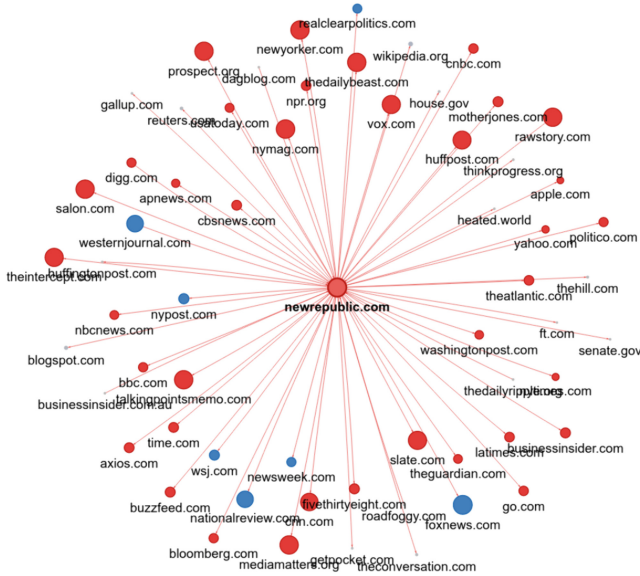
Fig. 1. Example showing how `newrepublic.com` relates with neighboring news sources. *Left* and *Right* wing sources are colored red and blue respectively, in addition, size of the node reflects the degree of the bias (learned by our **I**-political strategy). We can see that since `newrepublic.com` interacts mostly with *Left*-wing sources, its final bias degree ended up being considerable *Left*-wing. (Color figure online)

proposed models. The baseline with MAE of 0.902 uses an SVM classification model based on N-Grams. The top performed participating model [2] achieved a MAE of 0.320, outperforming the baseline and the other participating models. However, our proposed strategies (**P**-Political and **I**-Political) outperform the best-performing participating model in all reported metrics the top (MAE, F1-score and Accuracy). The MAE top performance (smaller MAE) indicates that the miss-predictions are less severe (from *Center* to the extremes or vice-versa), otherwise inferences will result on a higher penalization if predicting completely opposite extremes *Left* ↔ *Right*.

5   Conclusions

This research extends the methodology proposed in [8] by addressing long-term news media profiling, contrasting with approaches focused solely on short-term bias. Our experiments on two challenging bias descriptors-factual reporting and political bias-utilize four reinforcement learning strategies for classification performance evaluation. We provide compelling evidence supporting the longitudinal view of news media and their web interactions as a robust and scalable proxy for profiling, particularly regarding political bias and factual reporting. Concretely, performed experiments show that the proposed approach allows

superior performance in estimating outlet media bias descriptors compared to baseline methods. Furthermore, we present promising results from comparisons with other participating models submitted to the CLEF 2023 *CheckThat!* lab, designed for inferring political bias in currently active news outlets. Our approach surpasses top results in both F1-score and the official MAE performance measure, establishing a new SOTA result for this particular task. Finally, as an additional contribution, we release the largest dataset at the source media level, annotated with standard political bias and factual reporting labels.

As part of future efforts, we aim to investigate the dynamics of political bias changes over time within news media, such as shifts from center to extreme positions. Additionally, we plan to explore the integration of other bias descriptors, such as press freedom, in multi-task bias identification.

Disclosure of Interests. The authors have no competing interests.

References

1. Asr, F.T., et al.: The gender gap tracker: using natural language processing to measure gender bias in media. PLoS ONE **16**(1), e0245533 (2021)
2. Azizov, D., Liang, S., Nakov, P.: Frank at checkthat! 2023: detecting the political bias of news articles and news media. Working Notes of CLEF (2023)
3. Baly, R., Karadzhov, G., Alexandrov, D., Glass, J., Nakov, P.: Predicting factuality of reporting and bias of news media sources. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 3528–3539 (2018)
4. Baly, R., et al.: What was written vs. who read it: news media profiling using text analysis and social media context. In: Proceedings of the Association for Computational Linguistics (2020). https://doi.org/10.18653/v1/2020.acl-main.308
5. Baly, R., Karadzhov, G., Saleh, A., Glass, J., Nakov, P.: Multi-task ordinal regression for jointly predicting the trustworthiness and the leading political ideology of news media. In: Proceedings of the North American Chapter of the Association for Computational Linguistics (2019). https://doi.org/10.18653/v1/N19-1216
6. Billard, T.J., Moran, R.E.: Designing trust: design style, political ideology, and trust in "fake" news websites. Digit. Journal. **11**(3), 519–546 (2023)
7. Bozhanova, K., Dinkov, Y., Koychev, I., Castaldo, M., Venturini, T., Nakov, P.: Predicting the factuality of reporting of news media using observations about user attention in their youtube channels. In: Proceedings of the International Conference on Recent Advances in Natural Language Processing, pp. 182–189 (2021)
8. Burdisso, S., Sanchez-Cortes, D., Villatoro-Tello, E., Motlicek, P.: Reliability estimation of news media sources: birds of a feather flock together. In: Proceedings of the North American Chapter of the Association for Computational Linguistics (2024). https://aclanthology.org/2024.naacl-long.383
9. Da San Martino, G., Alam, F., Hasanain, M., Nandi, R.N., Azizov, D., Nakov, P.: Overview of the clef-2023 checkthat! lab task 3 on political bias of news articles and news media. Working Notes of CLEF (2023)

10. Esteves, D., Reddy, A.J., Chawla, P., Lehmann, J.: Belittling the source: trustworthiness indicators to obfuscate fake news on the web. In: Proceedings of the First Workshop on Fact Extraction and VERification (FEVER), pp. 50–59 (2018)
11. Fairbanks, J., Fitch, N., Knauf, N., Briscoe, E.: Credibility assessment in the news: do we need to read. In: Proceedings of the MIS2 Workshop Held in Conjunction with 11th International Conference on Web Search and Data Mining, pp. 799–800. ACM (2018)
12. Fang, X., Che, S., Mao, M., Zhang, H., Zhao, M., Zhao, X.: Bias of AI-generated content: an examination of news produced by large language models. Sci. Rep. **14**(1), 1–20 (2024)
13. Gezici, G.: Quantifying political bias in news articles (2022)
14. Hamborg, F.: Media bias analysis. In: Hamborg, F. (ed.) Revealing Media Bias in News Articles, pp. 11–53. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-17693-7_2
15. Hanimann, A., Heimann, A., Hellmueller, L., Trilling, D.: Believing in credibility measures: reviewing credibility measures in media research from 1951 to 2018. Int. J. Commun. **17**, 214–235 (2023)
16. Hounsel, A., Holland, J., Kaiser, B., Borgolte, K., Feamster, N., Mayer, J.: Identifying disinformation websites using infrastructure features. In: USENIX Workshop on Free and Open Communications on the Internet (FOCI) (2020)
17. Kamalloo, E., Dziri, N., Clarke, C.L., Rafiei, D.: Evaluating open-domain question answering in the era of large language models. In: Proceedings of the North American Chapter of the Association for Computational Linguistics (2023). https://doi.org/10.18653/v1/2023.acl-long.307
18. Kulshrestha, J., et al.: Search bias quantification: investigating political bias in social media and web search. Inf. Retrieval J. **22**, 188–227 (2019)
19. Leburu-Dingalo, T., Thuma, E., Motlogelwa, N., Mudongo, M., Mosweunyane, G.: UBCS at checkthat! 2023: stylometric features in detecting factuality of reporting of news media. Working Notes of CLEF (2023)
20. Li, C., Xue, R., Lin, C., Fan, W., Han, X.: Cucplus at checkthat! 2023: text combination and regularized adversarial training for news media factuality evaluation. Working Notes of CLEF (2023)
21. Nakov, P., et al.: Overview of the clef-2023 checkthat! lab task 4 on factuality of reporting of news media. Working Notes of CLEF (2023)
22. Panayotov, P., Shukla, U., Sencar, H.T., Nabeel, M., Nakov, P.: Greener: graph neural networks for news media profiling. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pp. 7470–7480 (2022)
23. Van der Pas, D.J., Aaldering, L.: Gender differences in political media coverage: a meta-analysis. J. Commun. **70**(1), 114–143 (2020)
24. Pope, D.G., Price, J., Wolfers, J.: Awareness reduces racial bias. Manage. Sci. **64**(11), 4988–4995 (2018)
25. Strömbäck, J., et al.: News media trust and its impact on media use: toward a framework for future research. Ann. Int. Commun. Assoc. **44**(2), 139–156 (2020). https://doi.org/10.1080/23808985.2020.1755338
26. Tran, S., Rodrigues, P., Strauss, B., Williams, E.: Accenture at checkthat! 2023: learning to detect factuality levels of news sources. Working Notes of CLEF (2023)
27. Wessel, M., Horych, T., Ruas, T., Aizawa, A., Gipp, B., Spinde, T.: Introducing mbib-the first media bias identification benchmark task and dataset collection. In: Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 2765–2774 (2023)