applied sciences

MDPI

Article

A Cross-Lingual Media Profiling Model for Detecting Factuality and Political Bias

Chichen Lin [1], Yongbin Wang [1], Chenxin Li [2], Weijian Fan [1], Junhui Xu [2] and Qi Wang [2,*]

[1] State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing 100024, China
[2] School of Computer and Cyber Sciences, Communication University of China, Beijing 100024, China
* Correspondence: vita1982@cuc.edu.cn

Abstract: Media profiling offers valuable insights to enhance the objectivity and reliability of news coverage by providing comprehensive analysis, but the diversity in languages posed significant challenges to our identification of factuality and political bias of non-English sources. The limitation of existing media analysis research is its concentration on a singular high-resource language, and it hardly extends to languages beyond English. To address this, we introduce xMP, a dataset for zero-shot cross-lingual media profiling tasks. xMP's cross-lingual test set encompasses 34 non-English languages and 18 language families, extending media profiling beyond English resources and allowing us to assess cross-lingual media profiling model performance. Additionally, we propose a method, named R-KAT, to enhance the model's zero-shot cross-lingual transfer learning capability by building virtual multilingual embedding. Our experiments illustrate that our method improves the transferability of models in cross-lingual media profiling tasks. Additionally, we further discuss the performance of our method for different target languages. Our dataset and code are publicly available.

Keywords: media profiling; cross-lingual; factuality; political bias

Citation: Lin, C.; Wang, Y.; Li, C.; Fan, W.; Xu, J.; Wang, Q. A Cross-Lingual Media Profiling Model for Detecting Factuality and Political Bias. *Appl. Sci.* **2024**, *14*, 9837. https://doi.org/10.3390/app14219837

Academic Editor: Fabrizio Marozzo

Received: 25 September 2024
Revised: 21 October 2024
Accepted: 22 October 2024
Published: 28 October 2024

1. Introduction

Media profiling stands at the forefront of media studies, offering critical insights into the complexities of news sources in our global information landscape. As media consumption transcends linguistic and national boundaries, there is an increasing demand for multilingual media profiling tools [1]. However, the diversity of languages poses a significant challenge to traditional, manual analysis methods. Verifiers operating primarily in English often struggle to navigate language barriers, cultural differences, and distinct national systems, making it difficult to accurately verify news from other countries. Consequently, the need for automated cross-lingual media analysis has become apparent, particularly for effectively addressing issues like political bias and the spread of misinformation [2,3]. In this context, our work introduces a potential solution by proposing a cross-lingual media profiling dataset and model.

Existing research has shown that media bias plays a pivotal role in shaping public opinion through mechanisms such as selective reporting, framing, and language use [4]. News selection criteria also influence how journalists choose which events to report, making it essential to understand media priorities in different contexts [5,6]. Moreover, biases in news coverage often lead to the underrepresentation of certain demographics and topics, such as gender [7]. Factors like geographic location, political ideology, and commercial interests further contribute to media bias [8–10], which can shape public perceptions and influence election outcomes [11,12]. To analyze these biases, both qualitative and quantitative methods—ranging from content analysis to keyword-based techniques—have been employed [5,13]. However, existing studies predominantly focus on high-resource languages [4,14–17], leaving a significant gap in the analysis of media from low-resource

languages. News is produced globally, often in languages that lack extensive digital resources, making cross-lingual media profiling both crucial and challenging.

In recent years, the development of media profiling datasets such as MBFC [18–21] and NELA-GT [22–25] has advanced the field. These datasets typically focus on textual features [18,19,26–28], audience homogeneity [19,29–31], and media outlet characteristics [32–35]. Yet, they remain limited by their focus on high-resource languages, overlooking the linguistic diversity necessary for global media analysis.

To meet this challenge, our proposed cross-lingual media analysis dataset and model address this gap by extending bias and factuality assessments to low-resource languages worldwide. Specifically, we present xMP, a cross-lingual evaluation dataset for media profiling. xMP covers 242 media outlets and 117.3 K articles in 34 non-English languages from 18 language families, providing comprehensive annotations for both factuality and political bias. We also explore the effects of fine-grained versus coarse-grained labeling, as well as the impact of linguistic differences on media profiling.

Despite advancements in deep learning-based media profiling, the field remains constrained by a reliance on high-resource language training datasets. Zero-shot cross-lingual profiling offers a promising solution by enabling models to generalize across languages that were not part of the training data. However, significant differences among languages present ongoing challenges for cross-lingual transfer [36]. In response to these challenges, we introduce R-KAT, a cross-lingual training method that does not require parallel corpora or supervised data in low-resource languages. Our method employs multistep iterative perturbations to construct a virtual multilingual news source embedding, followed by regularization across both English and multilingual embeddings, thus improving the multilingual pretrained model's cross-lingual transferability and enhancing performance in both monolingual and cross-lingual media profiling tasks. Comprehensive experiments show that while our method outperforms other baselines, zero-shot cross-lingual media profiling tasks still present challenges, particularly with political bias predictions for certain languages.

The main contributions of this paper can be summarized as follows:

- To the best of our knowledge, we present xMP, the first evaluation dataset for media profiling across 34 target languages. It adopts the fine-grained labeling setting that has been analyzed to be more suitable for media profiling tasks.
- We introduce a more powerful training strategy based on virtual multilingual embeddings designed to enhance the model's zero-shot cross-lingual transfer performance. Experimental results indicate that this approach significantly improves the model's cross-lingual capabilities in media profiling tasks.
- We explore the correlation between linguistic similarity and the model's cross-lingual performance. Our findings reveal that the cross-lingual transfer ability of R-KAT-based models is independent of language similarity, suggesting that our approach can effectively generalize existing transformer-based media analysis models to very dissimilar languages.

2. Related Work

In this section, we present the relevant datasets and methodologies for the media profiling tasks, along with an introduction to zero-shot cross-lingual training.

Media Profiling Datasets and Methods. Pioneer studies in media profiling initially focused on simple classifiers using features crafted from articles [37–39]. A significant advancement was made by [18,19], who achieved notable advancements with the introduction of their inaugural large-scale media profiling dataset, MBFC. This dataset surpasses previous ones by 1–2 orders of magnitude, marking significant progress in the field. MBFC is designed to encompass two critical media analysis tasks: factual assessment and political bias assessment. This dataset established a baseline by aggregating features from media sites, social platforms, Wikipedia, and news articles, enabling the assessment of the relative importance of these features in media evaluation. In contrast, Nørregaard et al. [22–25]

introduced a substantial media evaluation dataset with regular updates. While featuring fewer elements than MBFC2018, it specifically includes news articles and corresponding tweets, spanning an entire year's worth of news coverage. This dataset proves valuable for tasks like weakly supervised fake news detection or news political bias detection. Notably, these approaches often focus on content features, neglecting content-agnostic features. Subsequent studies introduce new features such as infrastructure characteristics [32–35] and audience homogeneity [19,40,41].

Recent research has concentrated on integrating audience homogeneity and graph neural networks. Ref. [29] utilized information from the Alexa rank website to model overlapping audience relationships across diverse media, a notable advancement. Similarly, refs. [42,43] used heterogeneous graphs to depict relationships among media, users, and news articles, employing relational inference operators to uncover potential interactions. However, both approaches primarily focus on media profiling in high-resource languages and do not consider the matching order between media and articles. Our goal is to construct a dataset for evaluating cross-lingual capabilities in media profiling models by collecting multilingual media and corresponding articles.

Zero-Shot Cross-Lingual Training. Pretrained multilingual models, such as XLM-RoBERTa-Large (XLM-R) [44] and mBERT [45], show potential in zero-shot cross-lingual tasks. Improving alignment across languages in a multilingual embedding space through parallel corpora has proven effective [46–49]. However, this approach demands extensive parallel corpora, incurring high data acquisition costs for low-resource languages [50]. Additionally, inherent semantic differences among languages hinder perfect alignment. To enhance zero-shot cross-lingual performance, some studies explore training methods using random smoothing, data augmentation, and perturbation in the training pipeline [51–53]. Inspired by previous work [54–61], we propose a training method for zero-shot cross-lingual media profiling tasks, combining virtual multilingual embedding building and language regularization strategies.

3. Dataset

We build xMP, a crossligual media profiling dataset which consists of 460 K articles from 2862 news media sources with factuality and political bias labels, of which 117.3 K articles from 242 media sources are in 34 non-English languages. In this section, we introduce our data collection process, label settings, and analysis of xMP.

3.1. Data Collection

We describe our media outlets and label collection methods, as well as our article collection and processing methods in this section.

3.1.1. Media Outlets and Label Collection

Similar to the CheckThat! 2023 [21,62] and MBFC datasets [18,19], we collect media outlets in different languages and their factuality and political bias labels.

Firstly, we collect 5398 media outlets from MBFC to obtain a multilingual media outlets list. We filter the list of possible multilingual media outlets by country, sifting out English-only countries, which are the United States, the United Kingdom, Canada, and Australia, and end up with 627 possible multilingual media outlets. Secondly, we use three language detection methods to further filter the multilingual news media list, which are the tags of HTML and two Python packages: langid (v1.1.6) (https://github.com/saffsd/langid.py, accessed on 15 December 2023.) and langdetect (v1.0.9) (https://github.com/Mimino666/langdetect, accessed on 15 December 2023.), and we find that HTML tags are the most accurate. Thus, we utilize the HTML tags and obtain 123 multilingual media outlets. After that, we expand the multilingual media outlets list by finding multilingual versions of media outlets in the English media outlets list, and we finally obtain a multilingual media outlets list containing 291 media outlets. The remaining English news media sites make up our English media outlets list.

In terms of labeling the articles, we used distant supervision based on the media outlet labels from the MBFC website. This method assumes that the overall bias and factuality of a media outlet reflect the articles published by that outlet, a practice commonly adopted in prior research. Therefore, we directly applied the MBFC labels for factuality (ranging from "Very Low" to "Very High") and political bias (ranging from "Extreme Left" to "Extreme Right") to the articles collected from these outlets, without modifying this methodology.

3.1.2. Article Collection and Processing

We crawled data from each media outlet on the list obtained through the previous process. Using the Newspaper3k(v0.2.8) (https://github.com/codelucas/newspaper, accessed on 15 December 2023), we extracted all URLs linked within the media outlets' websites. We applied URL matching to exclude advertisements, as these links often point outside of news websites. After filtering out suspected advertisements, we downloaded the HTML files of the remaining news URLs. The BeautifulSoup library was then employed to parse the HTML and extract the corresponding article titles, texts, image URLs, authors, and publication dates.

However, due to anti-crawling measures on certain websites, we were unable to collect content from them. As a result, our dataset comprises 242 multilingual news media outlets and 2620 monolingual ones. Despite these limitations, we hope that, for academic purposes, these sites may eventually offer open public interfaces to facilitate deeper research in this area.

Our data are stored in JSON format. The JSON file of xMP contains the media outlet URL and name, the factuality label, the political bias label, the language category, and the country of origin for each media outlet. For each article, the JSON file includes the title, content, image URLs, publication date, and the article URL. We believe the image URLs in our dataset will enable future research into multimodal media profiling.

Given our focus on cross-lingual media profiling, we further split the dataset, as detailed in Table 1. The English data were divided into training, validation, and test sets using the standard 8:1:1 ratio. For the cross-lingual media profiling task, the multilingual data were split into cross-lingual test sets, each evaluated separately based on language.

Table 1. Data split of our xMP. The first row shows the split size of media outlets and articles of our English data. The second row shows the split size of media outlets and articles of our cross-lingual data.

Dataset	Train		Val		Test		Total	
	Media	Article	Media	Article	Media	Article	Media	Article
Monolingual (EN)	2096	350,844	262	50,207	262	48,194	2620	449,245
Multilingual	–	–	–	–	242	11,729	242	11,729

3.2. Label Settings

For each media outlet, we have two media profiling tasks for factuality and political bias, so our data have two labels: "factuality" and "political bias". In addition to that, we have two label settings: fine-grained labels and coarse-grained labels.

3.2.1. Fine-Grained Labels

According to MBFC, a media source's "Factuality" is rated on a 6-point scale from "Very High" down to "Very Low". It is based on the evaluation of the factuality of that news media's articles. "Political Bias" ratings are American-centric and include "Extreme Left", "Left", "Left-Center", "Least Biased", "No Rated", "Right-Center", "Right", and "Extreme Right". We found that media sources labeled "No Rated" publish articles that are not politically relevant, so we merged them with "Least Biased". Therefore, we have the factuality label in 6 categories and the political bias label in 7 categories.

Table 2 shows the comparison between our xMP and previous datasets.

It is worth noting that not all articles are politically biased. For example, although sports news is less likely to reflect political bias than political topics in a single language, the opposite is often true in multiple languages. The media may be influenced by nationalism or political system differences and show political bias. We did not choose a specific topic to build the dataset because the previous English media bias dataset mostly collected news in the political field. We hope to cover more languages and events to build a universal media evaluation dataset.

Table 2. Summary of datasets of media profiling.

Dataset	Factuality	Political Bias	Media	Article	Languages
MBFC-2018	3-point scale from MBFC	7-point scale from MBFC	1066	94.2 K	1
MBFC-2020	3-point scale from MBFC	3-point scale from MBFC	865	68.2 K	1
NELA-GT-2018	Multipoint scale from 8 different assessment sites	Multipoint scale from 8 different assessment sites	194	713 K	1
NELA-GT-2019	Multipoint scale from 7 different assessment sites	Multipoint scale from 7 different assessment sites	260	1.12 M	1
NELA-GT-2020	6-point scale from MBFC	10-point scale from MBFC	519	1.78 M	1
NELA-GT-2021	6-point scale from MBFC	10-point scale from MBFC	367	1.85 M	1
NELA-GT-2022	6-point scale from MBFC	10-point scale from MBFC	361	1.77 M	1
Check That! 2023 Task3	–	3-point scale from MBFC	1023	8.7 K	1
Check That! 2023 Task4	3-point scale from MBFC	–	1189	10 K	1
xMP	6-point scale from MBFC	7-point scale from MBFC	2862	460 K	35

3.2.2. Coarse-Grained Labels

In order to be compatible with previous work, we propose a simplified version of the labels. We perform a 3-point scale process for the two labels of xMP. For the factuality label, we simplify the 6 labels into 3 labels: "Very High" and "High" are merged as "High", "Mostly Factual" and "Mixed" are merged as "Mixed", and "Very Low" and "Low" are merged as "Low". For the political bias label, we simplify 7 labels into 3 labels: "Extreme Right" and "Right" are merged as "Right", "Right-Center", "Least Biased" and "Left-Center" are merged as "Center", and "Extreme Left" and "Left" are merged as "Left".

3.3. Data Analysis

The xMP dataset boasts a media source count that is at least 2.3 times larger than its predecessor MBFC, accompanied by a substantial volume of articles. Although it does not provide a complete year of report coverage akin to NELA-GT, ongoing updates are planned until reaching that level. Notably, xMP surpasses other datasets with its inclusion of 35 languages, as highlighted in Table 2.

Figure 1 shows the number of media outlets and articles by language in xMP, and we have 2620 media outlets and 44,925 articles of English data, a much larger number than in other languages. In the cross-lingual test dataset, French has the highest number of news media with 29, Spanish has the highest number of articles with 14,593, some languages such as Norwegian and Vietnamese have the lowest number of news media with only 2, and Macedonian has the lowest number of articles, with only 12. Moreover, the number of media outlets is often not proportional to the number of articles. For example, Persian only has 8 media outlets, but its number of articles is more than that of French, which has the largest number of media outlets, with 13,677 articles. This happens because some languages have fewer resources and limited media reporting on the web, resulting in uneven output rates. It proves that xMP is realistic and that data in some languages are difficult to access, so cross-lingual tasks are appropriate and necessary, and we hope to access more data in our future work. Appendix B shows the languages and corresponding language families.

The label distribution of xMP is shown in Figure 2, where we plot heat maps of the label distribution of the multilingual data and the English data under two label settings. In Figure 2a,b, the distributions of English data and multilingual data under fine-grained labels are very similar by color and trend. In Figure 2c,d, the distributions differ between

English data and multilingual data under the coarse-grained label setting, especially on the label "Extreme Left" of political bias. English data and multilingual data are distributed more closely under fine-grained labels, which is conducive to cross-lingual transfer tasks. Media with a higher degree of extreme political bias are less factual, while media with a more neutral political stance are more factual and, therefore, more trustworthy. This correlation is also consistent with Baly et al.'s conclusion [27], which is in line with our expectations and reality. The specific label distribution is shown in Appendix A. There are more media outlets with factual labels of Mixed and High, and more media with political bias labels of Least-Biased, Left-Center, and Right-Center in xMP.

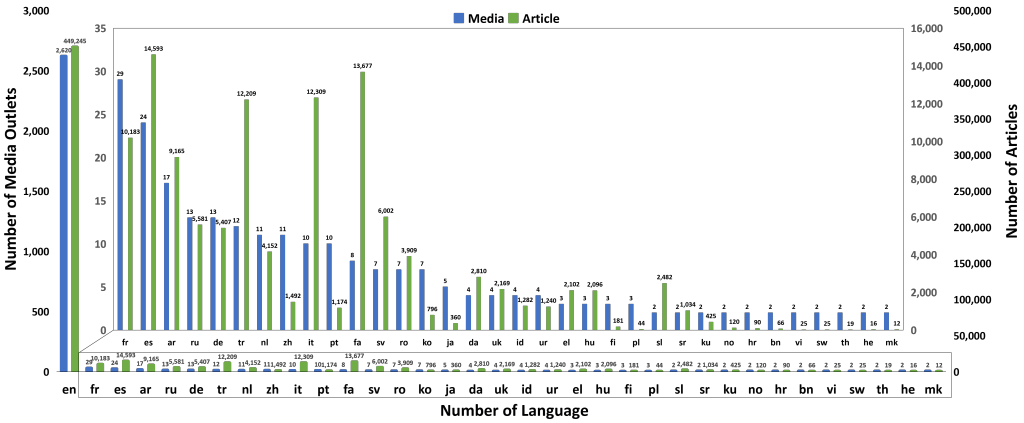

Figure 1. Number of media outlets and articles in our cross-lingual test dataset.

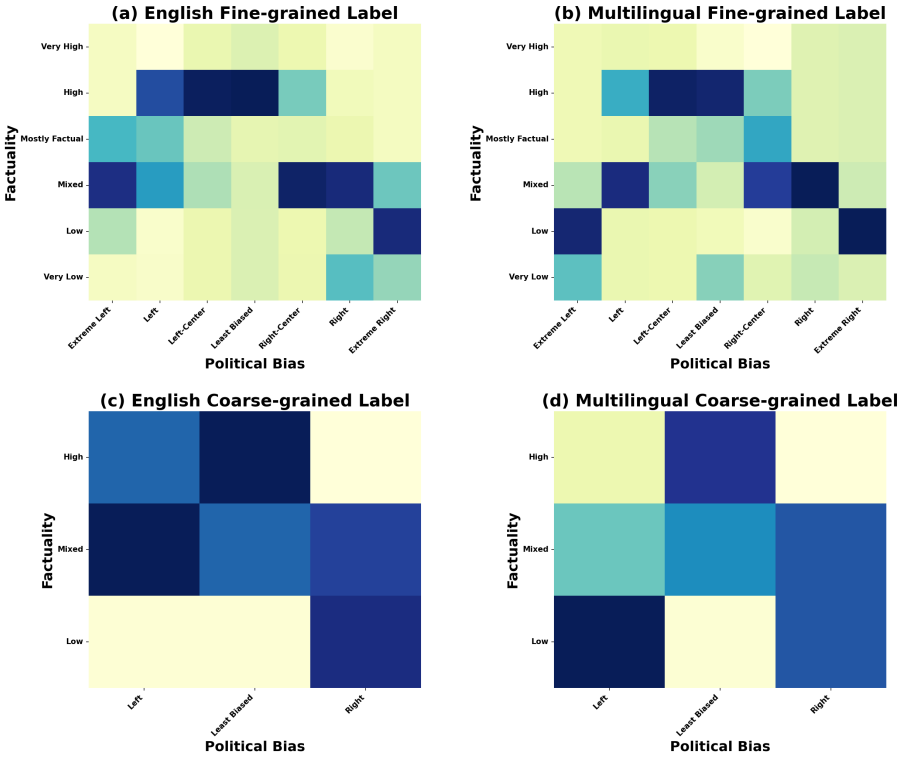

Figure 2. Label distribution heat maps of xMP. (**a**,**b**) are, respectively, the distribution of English and cross-lingual test datasets under the fine-grained label setting; (**c**,**d**) are, respectively, the distribution of English data and multilingual data under the coarse-grained label setting. The color intensity in the charts reflects the relative volume of media sources, with darker shades indicating a higher quantity.

Although there are far more English resources than multilingual resources, it is not enough to conduct media profiling only for English media. In today's global intercommunication environment, media profiling for other languages should also be explored. Therefore, we believe that the task of cross-lingual media profiling is significant.

4. The Proposed Method

This section outlines the media profiling tasks and details our proposed approach.

4.1. Problem Formulation

Following prior work [18,19,63], the media profiling tasks are divided into factuality and political bias rating tasks. Given a set $A_i = a_1, a_2, ..., a_j$ representing all reported articles of a news media outlet i, the corresponding label is denoted as y_i. Our objective is to train a media profiling predictor P, such that $P(A_i) = \hat{y}_i \sim y_i$.

4.2. R-KAT

We propose R-KAT, a training strategy for zero-shot cross-lingual media profiling. R-KAT operates primarily during fine-tuning, where it constructs virtual multilingual news source embeddings and regularizes both the multilingual and English embeddings. The training objective is to find optimal parameters that minimize the classification error between virtual multilingual and English news sources. The min–max optimization formulation is as follows:

$$\min_{\mathcal{L}} \sum_{(A_i, y_i) \in D} \mathcal{L}(P(A_i), P(A_i + \delta_K), y_i) \tag{1}$$

where D is the dataset and y_i is the ground-truth label. A_i denotes the input data, and δ_K represents the difference between the English and other language contextual representations. $\mathcal{L}$ is the loss function. The perturbation δ_K is calculated as follows:

$$\delta_K = \delta_{K-1} + \frac{g(P(A_i + \delta_{K-1}))}{||g(P(A_i + \delta_{K-1}))||_2} \tag{2}$$

where $g(P(A_i))$ is the gradient of the model with respect to the input A_i; when training begins, $\delta_{K=0} = 0$. Each time, we use the inverse of the gradient as a perturbation to align the English representation with the representation of the language with the largest difference. After three iterations, the final perturbation $\delta_{K=0}$ is generated.

Our approach simplifies the creation of virtual multilingual embeddings by directly perturbing the embedding vectors, without altering the original text. Through successive perturbations of the Transformer's word embedding matrix, we generate challenging virtual samples, enabling models trained on monolingual data to handle multilingual inputs. For these virtual samples, we compute the corresponding losses using cross-entropy:

$$loss_E = \mathcal{L}(A_i, y) \tag{3}$$

$$loss_M = \mathcal{L}(A_i + \delta_K, y) \tag{4}$$

After constructing the virtual multilingual samples, the model performs forward passes on the original English samples and the virtual multilingual samples, respectively, and calculates the classification losses $loss_E$ and $loss_M$ for each pass. In order to align the media profiling classification effects of different languages, we minimize the bidirectional KL divergence between the two output distributions. The regularization formula is as follows:

$$loss_{KL} = \frac{1}{2}[D_{KL}(P(A_i)||P(A_i + \delta_K)) + D_{KL}(P(A_i + \delta_K)||P(A_i))] \tag{5}$$

The final loss for the input sample A_i is formulated as follows:

$$loss_{total} = \frac{1}{2}\alpha \cdot (loss_E + loss_M) + \beta \cdot loss_{KL} \quad (6)$$

where α and β are coefficient weights that balance the classification loss $loss_E$, $loss_M$, and the KL divergence loss $loss_KL$. The overall training objective is to minimize the total loss $loss_{total}$. After calculating $loss_{total}$, the initial word embedding layer is reset, and final backpropagation is performed.

5. Experiments

In this section, firstly, we introduce our baseline methods. Secondly, we describe the experimental setup. Finally, we conduct experiments to answer the following research questions (RQs):

- **RQ1**: How does label granularity impact the effectiveness of media profiling?
- **RQ2**: How effective is our media profiling approach, considering metrics such as accuracy, F1 score, and MAE?
- **RQ3**: Can English-trained models effectively adapt to diverse languages, demonstrating efficacy in cross-lingual transfer learning for media profiling?

5.1. Baselines

To evaluate the effectiveness of different methods on xMP, we consider cross-lingual transfer learning models based on XLM-R and mBERT trained with the following three approaches:

- **ADV** [54,64–66] utilizes ϵ in the direction of gradient ascent to perturb the model's word embedding layer and generate strong adversarial samples.
- **RS-RP** [51] perturbs sentence embeddings with randomly sampled δ to smooth the classifier and build robust multilingual model.
- **RS-DA** [51] augments training data with English synonym replacement to train a smooth classifier and build robust multilingual model. In our setup, we doubled the size of the original training dataset using RS-DA, i.e., we used RS-DA once for each article.

5.2. Experiment Settings

We choose XLM-R [44] and mBERT [45] as our backbone models for all baselines, but in reality, other models with word embedding layers are equally applicable to this approach. XLM-R is a widely used pretrained language model for cross-lingual tasks, which we implemented using Huggingface's Transformer package. We freeze all layers in XLM-R except for the word embedding layer and build a classifier consisting of an average pooling layer and two fully connected layers. Our corresponding model is trained by fine-tuning this classifier on a single task of media profiling. In addition, we use accuracy, F1-macro, and MAE as evaluation metrics in our experiments for these two media profiling tasks.

We use a batch size of 8 and a learning rate of 2×10^{-5}. Additionally, we set the pretruncation and post-truncation values to 100 and 50, respectively. For cross-lingual media profiling tasks, we designate English as the source language and the remaining 34 languages as the target languages.

For all methods except R-KAT, we do not set dropout to avoid inconsistency in training and inference. For R-KAT, the dropout rate is 0.4, $\alpha = 0.1$, and $\beta = 0.5$ in the monolingual tasks, and the dropout rate is 0.4, $\alpha = 0.1$, and $\beta = 0.9$ in the cross-lingual tasks. The number of adversarial perturbations step K is set to 3.

We used a fixed random seed 42, and all code is run on a GPU, NVIDIA RTX 4090 24 GB, on the Autodl (https://www.autodl.com/home, accessed on 1 October 2023). platform. AutoDL is a cloud-based platform that provides users with access to scalable GPU resources for deep learning tasks. It is designed for researchers and developers who require robust computational power for model training and deployment without the need to manage physical infrastructure.

5.3. RQ1: Label Granularity—Fine vs. Coarse?

In this section, we conducted an investigation into the selection of label granularity within our dataset.

Initially, our approach involved training and testing models using both fine-grained and coarse-grained media evaluation labels, denoted as Fine (6-class/7-class) and Coarse (3-class) in the table, respectively. The coarse-grained labels represent a unified three-class categorization. Subsequently, during training, we employed fine-grained labels and implemented a rule-based methodology to transform fine-grained predictions on the test set into coarse-grained counterparts, identified as Fine2Coarse (3-class) in the table. The specific mapping distribution is meticulously documented in Table A1 within the Appendix A.

As shown in Table 3, fine-tuning a model for coarse-grained media profiling exhibited commendable performance, whereas fine-grained media profiling demonstrated suboptimal efficacy. Under the Fine2Coarse setting, mapping fine-grained predictions to a coarser granularity resulted in performance inferior to models directly trained with coarse-grained labels. The semantic relationships inherent in the labels guided the mapping relationships between labels of different granularity. However, these mapping relationships inadequately link the tasks of fine-grained and coarse-grained media profiling.

Table 3. Experimental results on the relationship between label granularity.

Label Granularity	Factuality			Political Bias		
	ACC	F1-Macro	MAE	ACC	F1-Macro	MAE
Fine (6-class/7-class)	0.221	0.242	0.573	0.226	0.227	0.744
Coarse (3-class)	0.748	0.717	0.263	0.453	0.468	0.229
Fine2Coarse (3-class)	0.460	0.485	0.313	0.254	0.289	0.237

Based on our experimental findings, we posit that one of the contributing factors to the suboptimal performance of existing media profiling models in real-world scenarios is the misalignment between the granularity of existing dataset labels and the demands of real-world environments. Consequently, we introduce the xMP fine-grained media profiling dataset, tailored to more accurately capture the intricacies of real-world media profiling scenarios.

5.4. RQ2: How Effective Is Our Approach?

In this section, we analyze the results of the performance of R-KAT compared to other cross-lingual transfer methods.

Performance of the R-KAT

Table 4 presents the results of XLM-R with multiple training strategies on xMP. We observed significant performance improvements in both monolingual and zero-shot cross-lingual settings for both media profiling tasks. The optimal results are highlighted in bold, while the suboptimal results are underlined.

Conducting media profiling tasks in a cross-lingual context poses significant challenges. Compared to the monolingual setting, the mean absolute error (MAE) for media factuality ratings increased by an average of 0.677. Meanwhile, the F1 score dropped by an average of 0.085, and accuracy decreased by 0.155 on average in the zero-shot cross-lingual setting. Similarly, for media political bias ratings, MAE increased by 0.489, the F1 score decreased by 0.234, and accuracy dropped by an average of 0.326.

These results highlight the need for further research to improve the transfer of knowledge from English-language media profiling to multilingual media profiling. Current approaches face significant performance drops in cross-lingual scenarios, underscoring the complexity of media profiling across different languages and cultural contexts.

Table 4. The experimental results for media profiling tasks on xMP, encompassing both English and zero-shot cross-lingual media profiling, are presented. Bold text denotes the best performance, while underlined text signifies the second-best performance.

	Model	Monolingual (EN)			Cross-Lingual		
		ACC	F1-Macro	MAE	ACC	F1-Macro	MAE
Factuality	XLM-R	0.241	0.244	0.695	0.177	0.206	1.330
	+ADV	0.439	0.367	0.405	0.247	0.268	1.090
	+RS-RP	0.467	0.381	0.477	0.243	0.260	1.154
	+RS-DA	0.224	0.239	0.672	0.173	0.195	1.384
	+R-KAT	**0.544**	**0.437**	**0.385**	**0.302**	**0.315**	**1.062**
Political Bias	XLM-R	0.223	0.214	0.779	0.137	0.180	0.985
	+ADV	0.705	0.534	**0.504**	0.126	0.144	1.146
	+RS-RP	0.575	0.557	0.603	0.186	0.201	1.114
	+RS-DA	0.215	0.235	0.718	0.163	0.201	**0.973**
	+R-KAT	**0.740**	**0.571**	0.508	**0.214**	**0.214**	1.337

Perturbation-based approaches, such as ADV, RS-RP, and R-KAT, consistently exhibit superior performance in cross-lingual media profiling tasks. In one-step perturbation-based training strategies for cross-lingual transfer learning, both ADV and RS-RP show notable performance improvements. However, significant performance gaps remain when compared to our training strategy, which is based on multistep perturbations.

The performance of these strategies, evaluated on mBERT, is presented in Table 5. Our proposed R-KAT methodology consistently demonstrates substantial performance gains across various frameworks and task configurations. As a comparative baseline, RS-RP consistently outperforms ADV in cross-lingual transfer, delivering optimal or near-optimal results across diverse scenarios. This trend aligns with the findings of previous research [51].

While RS-DA has shown success in prior cross-lingual transfer learning studies, its performance in our task fell short. We attribute this to the lack of diversity in RS-DA's sentence-level data augmentation, which hindered the construction of a robust multilingual media profiling embedding space. This limitation compromised its effectiveness in cross-lingual settings. Nonetheless, RS-DA may still be applicable for other cross-lingual media analysis tasks. We recommend exploring article-level data augmentation tailored specifically for cross-lingual transfer learning as a potential avenue for improving performance.

Table 5. The effectiveness of R-KAT on mBERT. Bold text denotes the best performance, while underlined text signifies the second-best performance.

	Model	Monolingual (EN)			Cross-Lingual		
		ACC	F1-Macro	MAE	ACC	F1-Macro	MAE
Factuality	mBERT	0.267	0.275	0.603	0.193	0.216	1.295
	+ADV	0.451	0.423	0.508	0.184	0.210	1.365
	+RS-RP	0.447	0.381	0.531	0.247	**0.258**	1.174
	+RS-DA	0.316	0.314	0.645	0.208	0.228	1.189
	+R-KAT	**0.574**	**0.497**	**0.473**	**0.267**	0.246	**1.063**
Political Bias	mBERT	0.222	0.233	0.740	0.151	0.171	1.067
	+ADV	0.413	0.426	0.607	0.188	0.200	**1.062**
	+RS-RP	0.568	0.524	0.592	0.217	0.230	1.151
	+RS-DA	**0.582**	0.490	0.645	0.205	0.220	1.209
	+R-KAT	0.549	**0.535**	**0.508**	**0.238**	**0.249**	1.070

5.5. RQ3: Can English-Trained Models Effectively Adapt to Divergent Languages?

To address RQ3, we display the experimental results of our method on some languages in Table 6.

Table 6 presents the experimental results across specific languages using XLM-R as the backbone, with the R-KAT training strategy applied. After incorporating R-KAT, performance improvements were observed across nearly all languages on the evaluation metrics. In the factuality assessment task, performance improved for all languages except French. On average, MAE decreased by 0.331 (indicating improved performance), while the F1 score and accuracy increased by 0.101 and 0.136, respectively.

Table 6. Experimental results of our model on some specific languages.

Language	Language Family	Factuality			Political Bias		
		ACC	F1-Macro	MAE	ACC	F1-Macro	MAE
XLM-R							
en	Indo-European (Germanic)	0.241	0.244	0.695	0.217	0.238	0.718
ar	Afro-Asiatic (Center-Semitic)	0.218	0.225	1.118	0.167	0.150	0.941
de	Indo-European (Germanic)	0.083	0.111	1.846	0.050	0.071	0.923
es	Indo-European (Romance)	0.104	0.137	1.083	0.167	0.174	0.833
fr	Indo-European (Romance)	0.103	0.136	1.276	0.146	0.173	0.724
nl	Indo-European (Germanic)	0.136	0.176	1.455	0.080	0.107	0.909
ru	Indo-European (Balto-Slavic)	0.046	0.075	1.769	0.129	0.122	1.077
tr	Altaic (Turkic)	0.042	0.071	1.917	0.033	0.057	1.500
XLM-R + R-KAT							
en	Indo-European (Germanic)	0.449	0.376	0.412	0.628	0.600	0.531
ar	Afro-Asiatic (Center-Semitic)	0.332	0.361	0.529	0.407	0.400	0.824
de	Indo-European (Germanic)	0.343	0.354	1.077	0.333	0.162	1.308
es	Indo-European (Romance)	0.198	0.230	0.833	0.346	0.238	1.417
fr	Indo-European (Romance)	0.103	0.136	1.276	0.230	0.205	1.069
nl	Indo-European (Germanic)	0.300	0.214	1.455	0.225	0.244	1.273
ru	Indo-European (Balto-Slavic)	0.250	0.137	1.615	0.173	0.196	1.385
tr	Altaic (Turkic)	0.186	0.171	1.417	0.190	0.197	2.000

For the political bias assessment task, accuracy and F1 score improved across all languages, with average increases of 0.193 and 0.144. However, some languages showed an increase in MAE, suggesting a decline in performance. As MAE is a comprehensive measure—where lower values indicate better outcomes—it should be regarded as a primary evaluation metric. The fact that some languages experienced increases in MAE despite improvements in accuracy and F1 score highlights that fine-grained news media political bias assessment in a cross-lingual context remains a significant challenge. This suggests that improvements in accuracy and F1 score do not necessarily lead to better MAE performance.

According to our experimental results, while R-KAT enhances the model's cross-lingual transfer capabilities, we observed a decline in zero-shot transfer performance for languages more similar to English, whereas performance improves for those that are more different. This finding aligns closely with previous research [51] and suggests a possible connection to the nature of perturbation-based cross-lingual methods, as it operates as a language-independent approach.

Our approach focuses on constructing virtual multilingual text embeddings through multistep perturbations, independent of language similarity, thereby enhancing overall cross-lingual capabilities. This suggests that incorporating syntactic perturbations could further improve zero-shot transfer performance. While methods utilizing parallel corpora can enhance transfer capabilities for similar languages, our approach aims to boost performance across a broader range of languages. Despite these advancements, overall performance remains modest due to the inherent challenges of the task. We will continue to refine our method in future work, as it represents a language-independent, cross-lingual approach.

In Table 6, we only select some languages with a large amount of data in xMP, because a small amount of data in a language may affect the accuracy of the experimental results, so we only discuss the languages with a relatively sufficient amount of data. In Appendix B, we provide a detailed discussion on the data amount of each language in xMP. Although the data amount of some languages is small, we still retain it in xMP; on the one hand, this will

have higher requirements for the model, which is conducive to improving the cross-lingual ability of the model. On the other hand, these data are also collected according to our standard procedures and are consistent with reality.

6. Conclusions and Future Work

We propose xMP, the first cross-lingual evaluation dataset for media profiling tasks. xMP is accessible in 35 languages and 18 language families, expanding the existing English media profiling dataset in scale and language. In the studies of label granularity, we found that fine-grained labeling would be better suited to the task of media analysis. To accommodate the need for the tasks of zero-shot cross-lingual media profiling, we propose a multistep perturbation-based virtual multilingual news embedding approach as an additional baseline, and experimental results demonstrate that our training method improves zero-shot cross-lingual transfer performance on media profiling tasks. Interestingly, we found no obvious direct correlation between the target languages' similarity to English and zero-shot cross-linguistic mobility after adopting this approach. This preliminary result could be attributed to the strategy of perturbations building.

In the future work, one potential extension is construct more sophisticated syntactic perturbations to enhance cross-lingual media profiling, because the perturbation-based cross-lingual approach presents properties that are inconsistent with the original approach.

7. Limitaion

We use political bias labels based on the US political system to categorize news from other countries. This simplified approach is common in previous datasets, as manually labeling the political stance of media in each country is challenging. Additionally, we observed that some non-English media included translated English news, suggesting possible consistency in bias.

Our labeling relies solely on MBFC for three main reasons: first, it is a recent, publicly available, large-scale source evaluation label set; second, other agencies like NewsGuard and Allsides do not offer free services; third, many previous media evaluations have utilized MBFC's factual and political bias labels, indicating their credibility. However, we recognize that MBFC's labels may not be entirely accurate, as they are primarily derived from labelers in English-speaking countries, which can introduce biases. Furthermore, the evaluation framework is based on the US political spectrum, which may not be applicable to all countries. We aim to address these issues in the future.

Author Contributions: Conceptualization, C.L. (Chichen Lin) and Q.W.; methodology, C.L. (Chichen Lin); software, C.L. (Chichen Lin); validation, C.L. (Chichen Lin); formal analysis, C.L. (Chichen Lin); investigation, C.L. (Chichen Lin) and C.L. (Chenxin Li); resources, C.L. (Chichen Lin) and Q.W.; data curation, C.L. (Chichen Lin) and J.X.; writing—original draft preparation, C.L. (Chichen Lin) and C.L. (Chenxin Li); writing—review and editing, W.F. and Q.W.; visualization, C.L. (Chichen Lin) and C.L. (Chenxin Li); supervision, Q.W. and Y.W.; project administration, Q.W. and Y.W.; funding acquisition, Q.W. and Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China (Grant number 2021YFF0901602) with funding from Qi Wang (10,000 RMB), the Fundamental Research Funds for the Central Universities (Grant number CUC23ZDTJ005) with funding from Qi Wang (10,000 RMB), and the High-quality and Cutting-edge Disciplines Construction Project for Universities in Beijing (Internet Information, Communication University of China) with funding from Yongbin Wang (20,000 RMB).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset is available at https://zenodo.org/record/8340085, accessed on 24 October 2023. Our data should be used only for academic research. Using crawlers to collect data may raise some ethical issues. Before crawling data, we checked the robots.txt file of the target website to understand the restrictions and requirements of the website for crawlers. The news

data we crawled will be provided in a restricted access form, and the source and author of the news will be accurately marked. We will ask users not to use these data for harmful purposes. If there are sensitive data in the news, we will remove these parts.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Label Distribution

Table A1 displays the specific number of English and multilingual news media in xMP under the fine-grained political bias and factuality labels, as well as the number of them mapped to the coarse-grained labels.

Table A1. Label distribution of xMP. Bold text denotes the best performance, while underlined text signifies the second-best performance.

	Label	Fine-Grained		Coarse-Grained	
Monolingual (EN)	Factuality	Very Low	143	Low	303
		Low	160		
		Mixed	956	Mixed	1122
		Mostly Factual	166		
		High	1147	High	1195
		Very High	48		
	Political Bias	Extreme Left	7	Left	334
		Left	265		
		Left-Center	441	Least Biased	1984
		Least Biased	723		
		Right Center	820		
		Right	243	Right	364
		Extreme Right	121		
Multilingual	Factuality	Very Low	25	Low	54
		Low	29		
		Mixed	63	Mixed	97
		Mostly Factual	34		
		High	93	High	92
		Very High	0		
	Political Bias	Extreme Left	8	Left	11
		Left	3		
		Left-Center	48	Least Biased	188
		Least Biased	93		
		Right Center	57		
		Right	24	Right	44
		Extreme Right	20		

Appendix B. Zero-Shot Cross-Lingual Experimental Results of Each Language

The specific results of the zero-shot cross-lingual media profiling tasks of our proposed method on each language are shown in Table A2. We also list in the table the language families to which each language belongs. Our method has different cross-lingual transfer effects for different languages. It performs well in Hebrew and Serbian but performs poorly in Danish, Norwegian, and Swedish. We speculate that this is related to the amount of cross-lingual test datasets. Due to the small number of media in some languages, extreme values appear in some results. However, for media with large amounts of data, cross-lingual performance is more stable. Zero-shot cross-lingual media profiling tasks are still challenging, as evidenced by our experimental results. We strive to obtain more data and propose methods that make it easier for the model to learn more comprehensive and robust cross-lingual transfer related knowledge in our future work.

Table A2. Language-specific experimental results of our model on cross-lingual media profiling tasks.

Language	Language Family	Factuality			Political Bias		
		ACC	F1-Macro	MAE	ACC	F1-Macro	MAE
ar	Afro-Asiatic (Center-Semitic)	0.332	0.361	0.529	0.407	0.400	0.824
bn	Indo-European (Indo-Aryan)	1.000	1.000	0.000	0.000	0.000	3.000
da	Indo-European (Germanic)	0.000	0.000	1.750	0.000	0.000	1.750
de	Indo-European (Germanic)	0.343	0.354	1.077	0.333	0.162	1.308
el	Indo-European (Greek)	0.000	0.000	2.333	0.333	0.333	1.000
es	Indo-European (Romance)	0.198	0.230	0.833	0.346	0.238	1.417
fa	Indo-European (Iranian)	0.422	0.440	0.625	0.438	0.350	0.875
fi	Uralic	0.333	0.400	1.333	0.167	0.250	0.667
fr	Indo-European (Romance)	0.103	0.136	1.276	0.230	0.205	1.069
he	Afro-Asiatic (Northwest-Semitic)	1.000	1.000	0.000	1.000	1.000	0.000
hr	Indo-European (Balto-Slavic)	0.250	0.333	1.500	0.000	0.000	3.000
hu	Uralic	0.111	0.167	1.667	0.000	0.000	1.667
id	Austronesian (Malayo-Polynesian)	0.444	0.389	0.750	0.083	0.125	1.000
it	Indo-European (Romance)	0.111	0.143	1.300	0.150	0.124	1.400
ja	Japonic	0.167	0.222	1.000	0.444	0.467	0.600
ko	Koreanic	0.180	0.214	1.143	0.167	0.213	1.714
ku	Indo-European (Indo-Iranian)	1.000	1.000	0.000	0.000	0.000	2.000
mk	Indo-European (Balto-Slavic)	0.333	0.333	1.500	0.333	0.333	1.500
nl	Indo-European (Germanic)	0.300	0.214	1.455	0.225	0.244	1.273
no	Indo-European (Germanic)	0.000	0.000	2.500	0.000	0.000	2.500
pl	Indo-European (Balto-Slavic)	0.125	0.167	2.000	0.000	0.000	2.333
pt	Indo-European (Romance)	0.320	0.309	1.300	0.250	0.200	0.900
ro	Indo-European (Romance)	0.190	0.242	0.714	0.056	0.095	1.143
ru	Indo-European (Balto-Slavic)	0.250	0.137	1.615	0.173	0.196	1.385
sl	Indo-European (Balto-Slavic)	0.000	0.000	1.000	0.333	0.333	1.500
sr	Indo-European (Balto-Slavic)	1.000	1.000	0.000	1.000	1.000	0.000
sv	Indo-European (Germanic)	0.000	0.000	2.000	0.000	0.000	1.143
sw	Niger-congo (Atlantic-Congo)	0.000	0.000	1.500	0.250	0.333	0.500
th	Sino-Tibetan (Kra-Dai)	1.000	1.000	0.000	0.000	0.000	1.500
tr	Altaic (Turkic)	0.186	0.171	1.417	0.190	0.197	2.000
uk	Indo-European (Balto-Slavic)	0.222	0.267	0.500	0.000	0.000	2.500
ur	Indo-European (Indo-Aryan)	0.375	0.429	0.500	0.111	0.133	1.000
vi	Austroasiatic (Vietic)	0.250	0.333	0.500	0.250	0.333	0.500
zh	Sino-Tibetan (Sinitic)	0.031	0.045	1.545	0.233	0.213	1.818
Mean	-	0.302	0.315	1.062	0.214	0.214	1.337

References

1. Nakov, P.; Sencar, H.T.; An, J.; Kwak, H. A survey on predicting the factuality and the bias of news media. *arXiv* **2021**, arXiv:2103.12506.
2. Sitaula, N.; Mohan, C.K.; Grygiel, J.; Zhou, X.; Zafarani, R. *Disinformation, Misinformation, and Fake News in Social Media*; Springer: Berlin/Heidelberg, Germany, 2020.
3. Huang, H.; Chen, Z.; Shi, X.; Wang, C.; He, Z.; Jin, H.; Zhang, M.; Li, Z. China in the eyes of news media: A case study under COVID-19 epidemic. *Front. Inf. Technol. Electron. Eng.* **2021**, *22*, 1443–1457. [CrossRef]
4. Gentzkow, M.; Shapiro, J.M.; Stone, D.F. Media bias in the marketplace: Theory. In *Handbook of Media Economics*; Elsevier: Amsterdam, The Netherlands, 2015; Volume 1, pp. 623–645.
5. Hamborg, F.; Donnay, K.; Gipp, B. Automated identification of media bias in news articles: An interdisciplinary literature review. *Int. J. Digit. Libr.* **2019**, *20*, 391–415. [CrossRef]
6. Puglisi, R.; Snyder, J.M., Jr. Empirical studies of media bias. In *Handbook of Media Economics*; Elsevier: Amsterdam, The Netherlands, 2015; Volume 1, pp. 647–667.
7. Haraldsson, A.; Wängnerud, L. The effect of media sexism on women's political ambition: Evidence from a worldwide study. *Fem. Media Stud.* **2019**, *19*, 525–541. [CrossRef]
8. Groseclose, T.; Milyo, J. A measure of media bias. *Q. J. Econ.* **2005**, *120*, 1191–1237. [CrossRef]
9. Merloe, P. Authoritarianism goes global: Election monitoring vs. disinformation. *J. Democr.* **2015**, *26*, 79–93. [CrossRef]
10. Chen, S.; Bruno, W.; Roth, D. Towards Corpus-Scale Discovery of Selection Biases in News Coverage: Comparing What Sources Say About Entities as a Start. *arXiv* **2023**, arXiv:2304.03414.

11. Bovet, A.; Makse, H.A. Influence of fake news in Twitter during the 2016 US presidential election. *Nat. Commun.* **2019**, *10*, 7. [CrossRef] [PubMed]
12. Grossmann, M.; Hopkins, D.A. *Asymmetric Politics: Ideological Republicans and Group Interest Democrats*; Oxford University Press: Oxford, UK, 2016.
13. Lott, J.R.; Hassett, K.A. Is newspaper coverage of economic events politically biased? *Public Choice* **2014**, *160*, 65–108. [CrossRef]
14. 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 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; p. 50.
15. Yang, K.C.; Menczer, F. Large language models can rate news outlet credibility. *arXiv* **2023**, arXiv:2304.00228.
16. Fung, Y.R.; Huang, K.H.; Nakov, P.; Ji, H. The battlefront of combating misinformation and coping with media bias. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 14–18 August 2022; pp. 4790–4791.
17. Lei, Y.; Huang, R.; Wang, L.; Beauchamp, N. Sentence-level media bias analysis informed by discourse structures. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Abu Dhabi, United Arab Emirates, 7–11 December 2022; pp. 10040–10050.
18. Baly, R.; Karadzhov, G.; Alexandrov, D.; Glass, J.; Nakov, P. Predicting Factuality of Reporting and Bias of News Media Sources. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 3528–3539.
19. Baly, R.; Karadzhov, G.; An, J.; Kwak, H.; Dinkov, Y.; Ali, A.; Glass, J.; Nakov, P. What Was Written vs. Who Read It: News Media Profiling Using Text Analysis and Social Media Context. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 3364–3374.
20. 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. In Proceedings of the CLEF 2023: Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 18–21 September 2023.
21. Nakov, P.; Alam, F.; Da San Martino, G.; Hasanain, M.; Nandi, R.; Azizov, D.; Panayotov, P. Overview of the CLEF-2023 CheckThat! lab task 4 on factuality of reporting of news media. In Proceedings of the CLEF 2023: Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 18–21 September 2023.
22. Nørregaard, J.; Horne, B.D.; Adalı, S. NELA-GT-2018: A large multi-labelled news dataset for the study of misinformation in news articles. In Proceedings of the International AAAI Conference on Web and Social Media, Buffalo, NY, USA, 3–6 June 2019; Volume 13, pp. 630–638.
23. Gruppi, M.; Horne, B.D.; Adalı, S. NELA-GT-2019: A Large Multi-Labelled News Dataset for the Study of Misinformation in News Articles. *arXiv* **2020**, arXiv:2003.08444.
24. Gruppi, M.; Horne, B.D.; Adalı, S. NELA-GT-2020: A large multi-labelled news dataset for the study of misinformation in news articles. *arXiv* **2021**, arXiv:2102.04567.
25. Gruppi, M.; Horne, B.D.; Adalı, S. NELA-GT-2022: A Large Multi-Labelled News Dataset for the Study of Misinformation in News Articles. *arXiv* **2022**, arXiv:2203.05659.
26. Horne, B.; Khedr, S.; Adali, S. Sampling the news producers: A large news and feature data set for the study of the complex media landscape. In Proceedings of the International AAAI Conference on Web and Social Media, Palo Alto, CA, USA, 25–28 June 2018; Volume 12.
27. 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 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MN, USA, 2–7 June 2019; pp. 2109–2116.
28. Baly, R.; Da San Martino, G.; Glass, J.; Nakov, P. 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, Online, 16–20 November 2020; pp. 4982–4991.
29. 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, Abu Dhabi, United Arab Emirates, 7–11 December 2022; pp. 7470–7480.
30. Ribeiro, F.; Henrique, L.; Benevenuto, F.; Chakraborty, A.; Kulshrestha, J.; Babaei, M.; Gummadi, K. Media bias monitor: Quantifying biases of social media news outlets at large-scale. In Proceedings of the International AAAI Conference on Web and Social Media, Palo Alto, CA, USA, 25–28 June 2018; Volume 12.
31. Stefanov, P.; Darwish, K.; Atanasov, A.; Nakov, P. Predicting the topical stance and political leaning of media using tweets. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 527–537.
32. 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 Conjuction with 11th International Conference on Web Search and Data Mining, Los Angeles, CA, USA, 6–8 February 2018; ACM: New York, NY, USA, 2018; pp. 799–800.

33. Castelo, S.; Almeida, T.; Elghafari, A.; Santos, A.; Pham, K.; Nakamura, E.; Freire, J. A topic-agnostic approach for identifying fake news pages. In Proceedings of the Companion Proceedings of the 2019 World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 975–980.
34. Hounsel, A.; Holland, J.; Kaiser, B.; Borgolte, K.; Feamster, N.; Mayer, J. Identifying disinformation websites using infrastructure features. In Proceedings of the 10th USENIX Workshop on Free and Open Communications on the Internet (FOCI 20), Online, 11 August 2020.
35. Papadogiannakis, E.; Papadopoulos, P.; P. Markatos, E.; Kourtellis, N. Who funds misinformation? A systematic analysis of the ad-related profit routines of fake news sites. In Proceedings of the ACM Web Conference 2023, Austin, TX, USA, 30 April–4 May 2023; pp. 2765–2776.
36. Ahmad, W.; Zhang, Z.; Ma, X.; Hovy, E.; Chang, K.W.; Peng, N. On Difficulties of Cross-Lingual Transfer with Order Differences: A Case Study on Dependency Parsing. 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), Minneapolis, MN, USA, 2–7 June 2019; pp. 2440–2452.
37. Rashkin, H.; Choi, E.; Jang, J.Y.; Volkova, S.; Choi, Y. Truth of varying shades: Analyzing language in fake news and political fact-checking. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 9–11 September 2017; pp. 2931–2937.
38. Potthast, M.; Kiesel, J.; Reinartz, K.; Bevendorff, J.; Stein, B. A Stylometric Inquiry into Hyperpartisan and Fake News. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, 15–20 July 2018; pp. 231–240.
39. Pérez-Rosas, V.; Kleinberg, B.; Lefevre, A.; Mihalcea, R. Automatic Detection of Fake News. In Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, NM, USA, 20–26 August 2018; pp. 3391–3401.
40. Ye, J.; Skiena, S. MediaRank: Computational ranking of online news sources. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2469–2477.
41. Gruppi, M.; Horne, B.D.; Adalı, S. Tell me who your friends are: Using content sharing behavior for news source veracity detection. *arXiv* **2021**, arXiv:2101.10973.
42. Nguyen, V.H.; Sugiyama, K.; Nakov, P.; Kan, M.Y. Fang: Leveraging social context for fake news detection using graph representation. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, Online, 19–23 October 2020; pp. 1165–1174.
43. Mehta, N.; Pacheco, M.L.; Goldwasser, D. Tackling fake news detection by continually improving social context representations using graph neural networks. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Dublin, Ireland, 22–27 May 2022; pp. 1363–1380.
44. Conneau, A.; Khandelwal, K.; Goyal, N.; Chaudhary, V.; Wenzek, G.; Guzmán, F.; Grave, É.; Ott, M.; Zettlemoyer, L.; Stoyanov, V. Unsupervised Cross-lingual Representation Learning at Scale. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 8440–8451.
45. Kenton, J.D.M.W.C.; Toutanova, L.K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.
46. Wei, X.; Weng, R.; Hu, Y.; Xing, L.; Yu, H.; Luo, W. On Learning Universal Representations Across Languages. In Proceedings of the International Conference on Learning Representations, Addis Ababa, Ethiopia, 30 April 2020.
47. Chi, Z.; Dong, L.; Wei, F.; Yang, N.; Singhal, S.; Wang, W.; Song, X.; Mao, X.L.; Huang, H.Y.; Zhou, M. InfoXLM: An Information-Theoretic Framework for Cross-Lingual Language Model Pre-Training. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Online, 6–11 June 2021; pp. 3576–3588.
48. Pan, L.; Hang, C.W.; Qi, H.; Shah, A.; Potdar, S.; Yu, M. Multilingual BERT Post-Pretraining Alignment. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Online, 6–11 June 2021; pp. 210–219.
49. Dou, Z.Y.; Neubig, G. Word Alignment by Fine-tuning Embeddings on Parallel Corpora. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Online, 19–23 April 2021; pp. 2112–2128.
50. Qiang, J.; Zhang, F.; Li, Y.; Yuan, Y.; Zhu, Y.; Wu, X. Unsupervised statistical text simplification using pre-trained language modeling for initialization. *Front. Comput. Sci.* **2023**, *17*, 171303. [CrossRef]
51. Huang, K.H.; Ahmad, W.; Peng, N.; Chang, K.W. Improving Zero-Shot Cross-Lingual Transfer Learning via Robust Training. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Virtual Event, 7–11 November 2021; pp. 1684–1697.
52. Ding, K.; Liu, W.; Fang, Y.; Mao, W.; Zhao, Z.; Zhu, T.; Liu, H.; Tian, R.; Chen, Y. A Simple and Effective Method to Improve Zero-Shot Cross-Lingual Transfer Learning. In Proceedings of the 29th International Conference on Computational Linguistics, Gyeongju, Republic of Korea, 12–17 October 2022; pp. 4372–4380.
53. Wang, Y.; Wu, A.; Neubig, G. English Contrastive Learning Can Learn Universal Cross-lingual Sentence Embeddings. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Abu Dhabi, United Arab Emirates, 7–11 December 2022; pp. 9122–9133.

54. Madry, A.; Makelov, A.; Schmidt, L.; Tsipras, D.; Vladu, A. Towards Deep Learning Models Resistant to Adversarial Attacks. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
55. Zhang, D.; Zhang, T.; Lu, Y.; Zhu, Z.; Dong, B. You only propagate once: Accelerating adversarial training via maximal principle. *Adv. Neural Inf. Process. Syst.* **2019**, *32*.
56. Zhu, C.; Cheng, Y.; Gan, Z.; Sun, S.; Goldstein, T.; Liu, J. FreeLB: Enhanced Adversarial Training for Natural Language Understanding. In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
57. Jiang, H.; He, P.; Chen, W.; Liu, X.; Gao, J.; Zhao, T. SMART: Robust and Efficient Fine-Tuning for Pre-trained Natural Language Models through Principled Regularized Optimization. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 2177–2190.
58. Wu, L.; Li, J.; Wang, Y.; Meng, Q.; Qin, T.; Chen, W.; Zhang, M.; Liu, T.Y. R-Drop: Regularized Dropout for Neural Networks. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 10890–10905.
59. Ni, S.; Li, J.; Kao, H.Y. R-AT: Regularized Adversarial Training for Natural Language Understanding. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2022, Abu Dhabi, United Arab Emirates, 7–11 December 2022; pp. 6427–6440.
60. Gao, T.; Yao, X.; Chen, D. SimCSE: Simple Contrastive Learning of Sentence Embeddings. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Virtual Event, 7–11 November 2021; pp. 6894–6910.
61. Xu, H.; Mannor, S. Robustness and generalization. *Mach. Learn.* **2012**, *86*, 391–423. [CrossRef]
62. 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. In Proceedings of the CLEF 2023: Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 18–21 September 2023.
63. Barrón-Cedeño, A.; Alam, F.; Caselli, T.; Da San Martino, G.; Elsayed, T.; Galassi, A.; Haouari, F.; Ruggeri, F.; Struß, J.M.; Nandi, R.N.; et al. The CLEF-2023 CheckThat! Lab: Checkworthiness, Subjectivity, Political Bias, Factuality, and Authority. In Proceedings of the European Conference on Information Retrieval, Dublin, Ireland, 2–6 April 2023; pp. 506–517.
64. Goodfellow, I.J.; Shlens, J.; Szegedy, C. Explaining and harnessing adversarial examples. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
65. Miyato, T.; Dai, A.M.; Goodfellow, I. Adversarial Training Methods for Semi-Supervised Text Classification. In Proceedings of the International Conference on Learning Representations, San Juan, PR, USA, 2–4 May 2016.
66. Shafahi, A.; Najibi, M.; Ghiasi, M.A.; Xu, Z.; Dickerson, J.; Studer, C.; Davis, L.S.; Taylor, G.; Goldstein, T. Adversarial training for free! *Adv. Neural Inf. Process. Syst.* **2019**, *32*.