Media Bias Everywhere? A Vision for Dealing with the Manipulation of Public Opinion

Michael Färber and Frederic Bartscherer

Karlsruhe Institute of Technology (KIT), Germany
{michael.faerber,frederic.bartscherer}@kit.edu

Abstract. This paper deals with the question of how artificial intelligence can be used to detect media bias in the overarching topic of manipulation and mood-making. We show three fields of actions that result from using machine learning to analyze media bias: the evaluation principles of media bias, the information presentation of media bias, and the transparency of media bias evaluation. Practical applications of our research results arise in the professional environment for journalists and publishers, as well as in the everyday life of citizens. First, automated analysis could be used to analyze text in real-time and promote balanced coverage in reporting. Second, an intuitive web browser application could reveal existing bias in news texts in a way that citizens can understand. Finally, in education, pupils can experience media bias and the use of artificial intelligence in practice, fostering their media literacy.

Keywords: Media Bias · News Bias · Text Mining · Trustworthy AI

1 Introduction

Nowadays, public opinion is shaped in manifold ways. The internet, in particular, is changing citizens' media consumption massively and digital media is increasingly influencing the public's opinions [7]. The distortion in media, also called *media bias*, consists, for example, of reporting through a selected choice of words or topics and is now also forcing large Internet companies, such as Facebook, to act [10].

Media bias is therefore the focus of many research areas [7,4]. In the humanities and social sciences, for example, content and meta analyses are carried out using a holistic approach to reveal different types of bias in news texts [4]. Given the ever-increasing rate of publication in digital media, an extensive manual analysis by experts is not possible [3]. In computer science, in contrast, the focus is on the automatic detection of media bias using machine learning methods, which allow for an automatic analysis over a large number of text documents, such as news texts [1].

In contrast to fake news analysis and detection [15], which are mostly limited to evaluating the content of facts, this paper is devoted to the topics of manipulation and mood-making. In cooperation with computer science, the humanities and social sciences,[1] we develop criteria that can be used to assess media bias in news texts. Furthermore, we investigate whether methods of artificial intelligence (AI) are suitable to analyze media bias in news texts in an understandable manner for citizens and to promote balanced coverage in reporting and media empowerment of citizens.

2 Field Analysis and Recommended Actions

Most computational approaches to assessing media bias use text mining methods, such as the lexical analysis of phrases [12]. AI methods, such as deep neural networks, can recognize complex relationships and extract knowledge from texts. Hence, we assume that, in the future, media bias will be recognized automatically in news texts and a quantitative analysis of media bias in its diverse dimensions (e.g., hidden assumptions, subjectivity, representation tendencies, and overall bias [6]) can be carried out.

In the following, we present three fields of action with respect to the application of AI for detecting media bias.

2.1 Field of Action 1: Evaluation Principles of Media Bias

AI methods, such as deep neural networks, learn relationships and dependencies based on training data, which can vary considerably depending on the application. In the case of news texts, for instance, texts on a specific topic often need to be annotated in a time-consuming manner, according to specified criteria.

We can derive two challenges from this:

1. **Annotation scheme**: Determining suitable criteria for assessing media bias.
2. **Annotation process**: Implementing a scalable annotation process to manage the increased annotation effort of multiple media bias dimensions.

Current research shows that annotated data sets for the fine-grained detection of media bias in news texts are still missing [6,9,14]. As the annotation by experts is time-consuming and expensive, we presented a scalable annotation approach to media bias in news texts based on crowd-sourcing [6]. The approach is applied to news texts about the Ukraine crisis in 2014 and 2015. In this way, we created a new media bias data set based on an annotation scheme at sentence level and the bias dimensions *hidden assumptions*, *subjectivity*, and *framing*.

2.2 Field of Action 2: Information Presentation of Media Bias

Citizens are faced with an ever-increasing rate of published news texts on a multitude of controversial topics, making it difficult to get an overview of hotly debated topics [8,13]. In addition, results must be presented intuitively to promote a balanced coverage in reporting as well as the citizens' media empowerment.

[1] We particularly thank Thomas Fetzer from the University of Mannheim, Jessica Heesen from the University of Tübingen, and Michael Decker from the Karlsruhe Institute of Technology (KIT).

We can derive two challenges from this:

1. **Controversial topics**: Disclosing prevalent positions on an issue.
2. **Comprehensible presentation**: Providing intuitive tools for analyzing media bias.

Media bias in reporting can be revealed through the analysis of controversial topics by extracting prevailing positions using AI methods. To this end, text mining systems can analyze debate portals such as debatepedia.org.[2] In this way, controversial topics can be identified to highlight contrary positions in news texts, for example, by using a label for disputed claims.

Furthermore, disclosing media bias is highly dependent on the way the respective information is presented. Information systems not only have to be intuitive to use, but also need to present the results in a comprehensible and human-centered way. We argue that appropriate tools, which are embedded in the daily news consumption process of citizens, are needed. For instance, a web browser application could highlight relevant text spans (e.g., similar to [2]).

2.3 Field of Action 3: Transparency of Media Bias Evaluation

The use of machine learning processes poses a potential risk that complex relationships are not sufficiently captured by the AI, possibly influencing opinions through incomplete information. Accordingly, the evaluation of AI methods for media bias detection and analysis must be transparent and comprehensible and neither discriminate nor distort information.

We can derive two challenges from this:

1. **Explainable AI methods**: Providing comprehensible explanations.
2. **Fair AI methods**: Making methods free from bias and non-discriminatory.

The information must be understandable and presented transparently. Accordingly, trustworthy AI methods (i.e., methods covering human agency and oversight, transparency, as well as diversity, non-discrimination and fairness [5]) are particularly important for media bias evaluation – for instance, through visualization (e.g., similar to [2]). Explanations can be both self-explanatory and make the results comprehensible through background information.

In computer science, fair methods are methods that are free of bias and do not discriminate. Intervening in the annotation and learning process allows outliers to be filtered out and a balanced evaluation to be achieved. In our research paper [6], for instance, we show that the country of origin of the crowdworkers can influence the the perception of media bias in the annotation process.

3 Conclusion

In this paper, we presented how media bias can be analyzed and determined automatically, based on artificial intelligence methods and the extent to which

[2] `http://www.debatepedia.org`

the evaluation of methods for automatic media bias annotation is sensitive to various aspects of computer science, the humanities, and the social sciences. Accordingly, three main fields of action were outlined, in which machine learning processes contribute to media bias evaluation: the evaluation principles of media bias, the information presentation of media bias, and the transparency of media bias evaluation.

Possible use cases of our research findings are the assistance of journalists and publishers for balanced coverage in reporting, as well as fostering citizens' media literacy. For example, an intuitive web browser application could highlight media bias in news texts to better understand distortions, and, applied in education, could allow pupils to experience both media bias and the use of artificial intelligence in practice to remedy weaknesses in their media literacy [11].

Acknowledgements. The project was funded as part of the digilog@bw joint research project by the Baden-Württemberg Ministry for Science, Research and Art with funds from the state digitization strategy digital@bw.

References

1. Bellows, M.: Exploration of Classifying Sentence Bias in News Articles with Machine Learning Models (2018), `https://digitalcommons.uri.edu/theses/1309`
2. Botnevik, B., Sakariassen, E., Setty, V.: BRENDA: Browser Extension for Fake News Detection. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 2117–2120. SIGIR'20, ACM (2020). https://doi.org/10.1145/3397271.3401396
3. Cremisini, A., Aguilar, D., Finlayson, M.A.: A Challenging Dataset for Bias Detection: The Case of the Crisis in the Ukraine. In: Proceedings of the 12th International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation. pp. 173–183. SBP-BRiMS'19 (2019). https://doi.org/10.1007/978-3-030-21741-9_18
4. D'Alessio, D., Allen, M.: Media bias in presidential elections: A meta-analysis. Journal of communication **50**(4), 133–156 (2000). https://doi.org/10.1111/j.1460-2466.2000.tb02866.x
5. European Commission: Ethics guidelines for trustworthy AI (Nov 2020), `https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai`
6. Färber, M., Burkard, V., Jatowt, A., Lim, S.: A Multidimensional Dataset Based on Crowdsourcing for Analyzing and Detecting News Bias. In: Proceedings of the 29th ACM International Conference on Information and Knowledge Management. pp. 3007–3014. CIKM'20, ACM (2020). https://doi.org/10.1145/3340531.3412876
7. Hamborg, F., Donnay, K., Gipp, B.: Automated identification of media bias in news articles: an interdisciplinary literature review. Int. J. Digit. Libr. **20**(4), 391–415 (2019). https://doi.org/10.1007/s00799-018-0261-y
8. Holton, A., Chyi, H.I.: News and the overloaded consumer: Factors influencing information overload among news consumers. Cyberpsychology Behav. Soc. Netw. **15**(11), 619–624 (2012). https://doi.org/10.1089/cyber.2011.0610
9. Lim, S., Jatowt, A., Färber, M., Yoshikawa, M.: Annotating and analyzing biased sentences in news articles using crowdsourcing. In: Proceedings of The 12th

Language Resources and Evaluation Conference. pp. 1478–1484. LREC'20 (2020), https://www.aclweb.org/anthology/2020.lrec-1.184/
10. van der Linden, S., Roozenbeek, J., Compton, J.: Inoculating Against Fake News About COVID-19. Frontiers in Psychology **11**, 2928 (2020). https://doi.org/10.3389/fpsyg.2020.566790
11. Maksl, A., Ashley, S., Craft, S.: Measuring News Media Literacy. Journal of Media Literacy Education **6**(3), 29–45 (2015), https://digitalcommons.uri.edu/jmle/vol6/iss3/3
12. Recasens, M., Danescu-Niculescu-Mizil, C., Jurafsky, D.: Linguistic Models for Analyzing and Detecting Biased Language. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. pp. 1650–1659. ACL'13 (2013)
13. Song, H., Jung, J., Kim, Y.: Perceived News Overload and Its Cognitive and Attitudinal Consequences for News Usage in South Korea. Journalism & Mass Communication Quarterly **94**(4), 1172–1190 (2017). https://doi.org/10.1177/1077699016679975
14. Spinde, T., Rudnitckaia, L., Mitrović, J., Hamborg, F., Granitzer, M., Gipp, B., Donnay, K.: Automated identification of bias inducing words in news articles using linguistic and context-oriented features. Information Processing & Management **58**(3), 102505 (2021). https://doi.org/10.1016/j.ipm.2021.102505
15. Zhou, X., Zafarani, R., Shu, K., Liu, H.: Fake News: Fundamental Theories, Detection Strategies and Challenges. In: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. pp. 836–837. WSDM'19, ACM (2019). https://doi.org/10.1145/3289600.3291382