BIAS PREDICTION IN MULTILINGUAL NEWS REPORTING

Swati

Doctoral Dissertation
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia

Supervisor: Prof. Dr. Dunja Mladenić, Jožef Stefan Institute and Jožef Stefan International Postgraduate School, Ljubljana, Slovenia

Evaluation Board:
Prof. Dr. Nada Lavrač, Chair, IPS and Jožef Stefan Institute, Ljubljana, Slovenia
Prof. Dr. Janez Demšar, Member, University of Ljubljana, Slovenia,
Prof. Dr. Marko Tadić, Member, University of Zagreb, Croatia

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

Swati

BIAS PREDICTION IN MULTILINGUAL NEWS REPORTING

Doctoral Dissertation

NAPOVEDOVANJE PRISTRANSKOSTI V VEČJEZIČNEM POROČANJU NOVIC

Doktorska disertacija

Supervisor: Prof. Dr. Dunja Mladenić

Ljubljana, Slovenia, July 2024

Acknowledgments

I am writing these lines out of deep gratitude for the fruitful collaborations and generous support of many people, without whom I would not have been able to complete my thesis.

I am especially grateful to my advisor, Prof. Dr. Dunja Mladenić, for supporting me in all my research endeavours. I am thankful for her inspiration, which has contributed a lot to becoming the researcher I am today. I thank her for her continuous support, patience, and motivation. Having the opportunity to do research as a member of her research group has been a real privilege.

I am grateful to Marko Grobelnik from the Jožef Stefan Institute in Ljubljana for his insightful comments, valuable ideas, and introduction to the concept of inferential commonsense knowledge, which played a crucial role in shaping this thesis. I am also grateful to my other mentors, Prof. Dr. Richard Rogers from the University of Amsterdam in Amsterdam and Daniel Gomes from Computação Científica Nacional in Lisbon, who supported me during my short-term scientific secondments. A heartfelt thanks to my master's advisor, Dr. Asif Ekbal from the Indian Institute of Technology in Patna, who introduced me to the NLP world. I sincerely thank Prof. Dr. Nada Lavrač, Prof. Dr. Janez Demšar, and Prof. Dr. Marko Tadić for serving as members of my doctoral committee.

I would also like to acknowledge all the projects that supported this work. In particular, it was supported in part by the European Union's Horizon Europe Projects enRichMyData under Agreement 101070284 and AI4Gov under Agreement 101094905, in part by the Slovenian Research Agency under Project J2-1736 Causalify, in part by the Slovenian National Grants CRP V2-2272, V5-2264, and CRP V2-2146, and in part by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant under Agreement 812997.

Special thanks to Adrian Mladenić Grobelnik for his invaluable contribution and comments. His expertise played a crucial role in conducting qualitative analyses for one of the papers that form an integral part of this thesis. My colleague, Dr. Alenka Guče, deserves special thanks for always giving me a shoulder to lean on and helping me refine my Slovenian translations. My deepest gratitude goes out to Klemen Kenda, who always stood by me with advice, support, and an open ear, helping me both professionally and personally. I appreciate that I could always ask for his advice, regardless of the matter. I particularly thank Jasna Franko for helping me complete all the paperwork and for her support on numerous occasions, which really helped me focus on my research. I would also like to extend my thanks to Polona Škraba Stanič, Mateja Škraba, Erik Novak, Dr. Jože Rožanec, and Mojca Kregar for their care and support. You all rock!

Most importantly, I am very thankful to my family. I am incredibly thankful to my loving parents, Binod Kumar and Dharamsheela Kumar, for raising me, understanding me, and supporting me. Without them, I would not have been able to study abroad and complete this thesis. Special thanks to my awesome siblings Sweta and Vishal and my brother-in-law, Naveen Gupta. I thank them for always being there for me and for their unconditional love and encouragement.

Also, a huge thanks to my closest friend, Arjun Roy. My communication with him was always straightforward and provided me with quick and helpful feedback. His belief in me, insightful advice for my research, and encouragement have played an important role in helping me complete my thesis. Without his support, completing this thesis would have been impossible.

Finally, I want to thank myself for being so brave to make the decision to leave the country I had called home for nearly three decades, jump out of my comfort zone, and embark on a journey filled with adventure. Thanks to myself for working so hard and not giving up when facing any frustrations or failures.

Thank you all!

Abstract

Over the past decade, rapid advancements in natural language processing have opened up new avenues for tackling complex issues such as news bias analysis. This progress has empowered researchers to explore innovative approaches to uncovering the complex biases inherent in news production and coverage processes. News bias, a multifaceted reflection of the inherent biases in the news production and coverage workflow, has far-reaching implications for public opinion and decision-making. Analysing it poses a formidable challenge, given its susceptibility to a variety of influencing factors, including but not limited to political affiliation, editorial independence, and writing style.

The thesis at hand delves into this complex landscape, aiming not only to identify and analyse news bias but also to extend these analyses to a multilingual context. By critically evaluating existing research, proposing future strategies, and establishing comprehensive methodologies for bias mitigation, the thesis lays the foundation for comprehensive methodologies for bias mitigation in news reporting. It also presents the design and provides the implementation of adaptable data generation scripts for generating customised datasets for related tasks. Furthermore, it introduces a publicly available, novel event-centric news dataset.

In addition, the thesis introduces a novel learning framework for predicting bias in news headlines, incorporating inferential commonsense knowledge. It demonstrates how such knowledge improves comprehension of short headlines that lack proper semantic and syntactic context. Specifically, it illustrates the crucial role played by commonsense when coupled with proper selection and refinement techniques in simplifying, interpreting, and explaining events not explicitly stated in the headlines, thereby significantly enhancing the task of bias analysis.

This thesis aims to close the gap by broadening its focus to embrace multilingualism in response to the shortcomings of existing approaches, which primarily focus on languages with abundant resources. This is accomplished by extending the previously introduced framework to efficiently harness inferential knowledge in a multilingual context, employing the translate-retrieve-translate strategy. This strategy is then used to propose a multilingual bias prediction framework capable of handling low-resource languages under an imbalanced sample distribution. It assesses the generalisability of the frameworks presented and investigates the effects of knowledge augmentation and attention mechanisms. It also provides a qualitative analysis of bias prediction performance across languages and provides comprehensive explanations.

Povzetek

V zadnjem desetletju je hiter napredek na področju obdelave naravnega jezika odprl nove možnosti za reševanje zapletenih vprašanj, kot je analiza pristranskosti novic. Ta napredek je raziskovalcem omogočil raziskovanje inovativnih pristopov za odkrivanje zapletenih pristranskosti, ki so neločljivo povezane s postopki priprave in poročanja o novicah. Pristranskost novic, ki je večplasten odsev pristranskosti v procesu priprave in pokrivanja novic, ima daljnosežne posledice za javno mnenje in sprejemanje odločitev. Analiziranje te pristranskosti predstavlja velik izziv, saj nanjo lahko vplivajo različni dejavniki, med drugim politična pripadnost, uredniška neodvisnost in slog pisanja.

Pričujoča disertacija se poglablja v to zapleteno področje in si prizadeva ne le opredeliti in analizirati pristranskost novic, temveč tudi razširiti te analize v večjezičnem kontekstu. S kritično oceno obstoječih raziskav, predlogi prihodnjih strategij in vzpostavitvijo celovitih metodologij za zmanjševanje pristranskosti, doktorsko delo postavlja temelje celovite metodologije za zmanjševanje pristranskosti pri poročanju novic. Predstavlja tudi zasnovo in zagotavlja implementacijo prilagodljivih skript za generiranje podatkov za generiranje prilagojenih podatkovnih nizov za sorodne naloge. Poleg tega je predstavljen nov javno dostopen, im na dogodke osredotočen nabor podatkov o novicah.

Delo uvaja tudi nov učni okvir za napovedovanje pristranskosti v naslovih novic, ki vključuje znanje na podlagi inferenčnega zdravo-razumskega sklepanja. Prikazuje, kako takšno znanje izboljša razumevanje kratkih naslovov, ki nimajo ustreznega semantičnega in sintaktičnega konteksta. Natančneje, prikazana je ključna vloga, ki jo ima zdrav razumsko sklepanje v povezavi z ustreznimi tehnikami izbire in izpopolnjevanja pri poenostavljanju, razlagi in pojasnjevanju dogodkov, ki niso izrecno navedeni v naslovih, s čimer se bistveno izboljša naloga analize pristranskosti.

Namen te disertacije je zapolniti vrzel s širitvijo fokusa na večjezičnost kot odgovor na pomanjkljivosti obstoječih pristopov, ki se osredotočajo predvsem na jezike z bogatimi viri. To dosežemo z razširitvijo predhodno predstavljenega okvira za učinkovito izkoriščanje sklepalnega znanja v večjezičnem kontekstu z uporabo strategije prevajanja - zajemanja - prevajanja. Ta strategija se nato uporabi za predlog večjezičnega okvira za napovedovanje pristranskosti, ki lahko ob neuravnoteženi porazdelitvi vzorcev obravnava jezike z malo viri. Ocenjena je posplošljivost predstavljenih okvirov in raziskani so učinki mehanizmov razširitve znanja in pozornosti. Disertacija vključuje tudi kvalitativno analizo uspešnosti napovedovanja pristranskosti v različnih jezikih in podaja izčrpne razlage.

Contents

Abbreviations

AI	...	Artificial Intelligence
IC_Knwl	...	Inferential Commonsense Knowledge
LIME	...	Local Interpretable Model-Agnostic Explanations
ML	...	Machine Learning
NLP	...	Natural Language Processing

Chapter 1

Introduction

The advancement in the field of natural language processing (Khurana et al., 2023; Min et al., 2023) over the last decade has made the solutions to complex machine learning problems more convenient and vice versa. The problems such as machine translation (Ranathunga et al., 2023), text summarization (Widyassari et al., 2022), information retrieval (Zhu et al., 2023), text classification (Q. Li et al., 2022), among others, are being solved much more efficiently than ever before. Consequently, it offered the researchers the opportunity to use these advanced techniques to solve problems in a variety of contexts, such as news bias analysis (Krieger et al., 2022; Naredla & Adedoyin, 2022).

News bias is the product of the inherent bias present in both the production and coverage processes (Hamborg et al., 2019). It occurs when an outlet publishes a story selectively or incorrectly. Multiple factors, such as the ethnic point of view, gender, political affiliation, writing style, and editorial independence of the people involved in the news production process, influence the story (Andrew, 2007; Ecker et al., 2014). It is nearly impossible to produce a bias-free story (Park et al., 2009). Biassed news has the potential to influence the thought process and decision-making of the person listening, watching, and/or reading it (Kahneman & Tversky, 2013). It can have several direct or indirect implications, whether political or social. For example, if the news shows only the positive or negative sides of a political party, it has been observed to influence the public vote (Bernhardt et al., 2008; De Vreese, 2005; Meyrowitz, 1986). Not only politics, but the news about a disaster or the spread of a viral disease also affects the beliefs of the general public.

News bias analysis is concerned with the identification of the inherent bias present in news production and its coverage process (Rodrigo-Ginés et al., 2023). The analysis is not only important for consumers but is also essential for journalists, publishers, and other people involved in the news production process (Dallmann et al., 2015; McCarthy & Dolfsma, 2014). It gives them an insight into whether they have overlooked some significant aspects of the news or exaggerated their reporting. They need this analysis more because at each step of the news production process, starting with story selection, bias can arise. Analysis of such bias in a news story involves various aspects like which story is worth reporting (Bourgeois et al., 2018) and how much importance it should be given (Bucher & Schumacher, 2006; Stovall, 1985), is it fake or not (Bago et al., 2019), are the sources of the story legitimate (Baker et al., 1994; Gentzkow & Shapiro, 2006), does the topic define the story correctly (W.-F. Chen et al., 2018; Gangula et al., 2019), are the words selected for the content writing appropriate (Grefenstette et al., 2004), selection of picture and its captions (Estrin, 2014) and many other factors. Although numerous research works have been published to study news bias, none of them provides its users with a broader perspective in terms of different dimensions, cultures, and languages.

A multitude of ongoing events provides numerous opportunities for biases to manifest in different forms (Bourgeois et al., 2018). Diversity in language and culture makes the identification of biases even more complicated. While most research work seeks to strengthen techniques for monolingual news bias analysis, there is less consideration of multilingual bias in news reporting. As a result, we intend to discuss the strategies and methods that we propose to develop in order to analyse news reporting bias in a multilingual context. We base our work on the identification of political bias in news headlines. It is a critical task since it influences the selection and distribution of news articles (Gangula et al., 2019). Predicting political bias using conventional methods is challenging due to the absence of sufficient syntactic and semantic information in the short headline texts (Gangula et al., 2019; Laban et al., 2021). To compensate for this lack of information, inferential commonsense knowledge can be employed in a way similar to how people use commonsense inferences to perform a variety of tasks. However, without proper emphasis, the additional inferential context is prone to introducing unnecessary noise, preventing models from fully exploiting the acquired knowledge.

To address these bottlenecks, we first introduce a learning framework that makes use of knowledge with proper selection and refinement techniques. Utilising the knowledge via a translate-retrieve-translate strategy, we then propose to extend the introduced framework to deal with multilingual news reporting bias. In addition, we consider investigating news bias in other dimensions, such as event selection, news content, and news sentiment.

1.1 Motivation

With the advent of machine learning and neural networks in recent years, advanced scientific methods for identifying and analysing news reporting bias have been developed (Davis et al., 2022; Spinde, 2021; Spinde, Rudnitckaia, et al., 2021). Despite the fact that the headline is increasingly becoming the only part of a news item that is read (Holmqvist et al., 2003), the related research predominantly focuses on news articles rather than headlines.

Deep learning is assumed to be superior at learning complex and dense text representations by capturing semantic and syntactic information (Patil et al., 2023). However, it is implausible to extend it from bias classification at the article level to the headline level (Gangula et al., 2019). This is due to the fact that headlines are more ambiguous than articles but also lack contextual and conceptual information. This problem becomes even more challenging in a multilingual context, which is a gap this thesis aims to bridge.

Existing research on multilingual news bias primarily focuses on languages with abundant resources (Aksenov et al., 2021; Guo & Zhu, 2022). Even though mitigating the effects of bias is equally important in assisting readers of languages with limited resources (Park et al., 2012), only a few studies have been conducted on these languages (Doan & Gulla, 2022). Moreover, these studies are either limited to a single language or conducted on news in different languages independently.

The framework proposed in this thesis is driven by the motivation to overcome the above-mentioned limitations. It highlights the role of inferential commonsense knowledge in facilitating the comprehension of short news headline text. It further proposes that the knowledge, when used in conjunction with the translate-retrieve-translate strategy (Fang et al., 2022), can effectively aid in the comprehension of narratives in a multilingual context. In effective systems, both implicit and explicit knowledge is expected. The proposed framework supports this viewpoint by combining the implicit knowledge derived from language models with the inferred knowledge in the form of explicit knowledge. A system is also expected to deal with low-resource situations in the real world; it is language-agnostic and thus adaptable to such scenarios.

We believe that the proposed framework is a valuable tool for copy editors responsible for rewriting headlines (Einsohn & Schwartz, 2019). It also has the potential to be useful in practical applications such as e-journalism and manual news-bias prediction portals, where it could be used to automatically classify headlines into different bias types. In addition, it could help to reduce the number of articles that require manual examination, which is a time-consuming process prone to annotator bias (Spinde, Krieger, et al., 2021).

Another driving motivation for this thesis is a lack of annotated data, which occurs frequently in real-world scenarios. The proposed data generation strategy, which is capable of working with low-resource languages with an imbalanced distribution, is intended to address this issue. In addition, it will facilitate future expansion and the creation of custom datasets for related tasks. The generated data would not only be ideal for automated systems like the "bias flipper" (W.-F. Chen et al., 2018; R. Liu, Jia, & Vosoughi, 2021; R. Liu, Wang, et al., 2021) but would also be beneficial for social scientists and researchers interested in analysing news reporting bias.

1.2 Hypotheses and Aims

We summarise our work as a set of hypotheses developed in relation to prior research and with specific goals in mind. We investigate and empirically test these hypotheses, which serve as the foundation for the contributions, as described in Section 1.3.

- **Hypothesis 1:** ***The adaptable data generation method for analysing news bias enables the creation of customised datasets for associated tasks, offering valuable insights into the complex dynamics of bias in news reporting.*** *For instance, its implementation as scripts for generating event-centric news datasets holds promise for facilitating a more nuanced understanding of complex event-outlet relationships.*

  State of the art: Traditional dataset generation methods often lack adaptability for customisation, restricting their suitability across diverse tasks (D. Liu et al., 2020). To address this limitation, the emergence of adaptable data generation methods has shown promising results in enhancing dataset creation for a variety of tasks (El Emam et al., 2020). Nevertheless, when it comes to tasks related to news bias analysis, such adaptable methods are scarce (Wessel et al., 2023). Regarding the availability of datasets, there still remains a gap in publicly accessible event-centric datasets that encompass a diverse array of features. While there are numerous datasets pertaining to news articles (J. Lee et al., 2023; S. Lee et al., 2023), only a few explicitly focus on events (Leetaru & Schrodt, 2013). Furthermore, these event-centric datasets often lack essential attributes for specifying the outlet(s) reporting the event, hindering in-depth analysis to grasp their complex interactions and relationships. Furthermore, the majority of these datasets are tailored to specific categories such as politics, healthcare, and disasters, constraining their applicability to specific research purposes only (Cheng et al., 2020).

  Aim: Our aim is to design an adaptable data generation method for generating customised datasets for tasks related to news bias analysis and to provide its implementation as scripts. We intend to use the scripts to introduce a publicly available, novel, event-centric news dataset with a diverse set of features, making it ideal for studying and analysing complex relationships between events and outlets.

- **Hypothesis 2:** ***The introduction of inferential knowledge, with an emphasis on significant inferences, enhances news headline comprehension and a model's overall ability to predict bias in reporting.***

State of the art: News headlines are inherently short, catchy or appealing, context-deficient, and contain only subtle bias clues (W.-F. Chen et al., 2018; Laban et al., 2021). Understanding the short headlines necessitates understanding the narrative being presented (Bruneau et al., 2012). This can be accomplished by identifying connections between what is explicitly stated and what is implied (Berner, 1983). It is well known that incorporating commonsense reasoning abilities can facilitate the inference of such connections by identifying a set of unstated causes and effects (D. Li et al., 2021; J. Li et al., 2021). Such additional knowledge has been proven to be beneficial for several tasks (Du et al., 2022; Lieto et al., 2021). Nevertheless, existing approaches fail to acknowledge the potential of commonsense reasoning in facilitating headline comprehension to aid in bias prediction. The inclusion of inferential knowledge has demonstrated significant improvements in the model's performance across a wide range of tasks, including emotion inference (D. Li et al., 2021), sarcasm detection (Chowdhury & Chaturvedi, 2021), sarcasm generation (Chakrabarty et al., 2020), and reading comprehension (Huang et al., 2019), among others. Although this additional knowledge has been demonstrated to be promising, existing research fails to acknowledge its potential in aiding text comprehension, particularly in the domain of news bias prediction. Furthermore, existing research lacks the integration of relevant attention mechanisms to emphasise significant inferences suited for news bias prediction.

Aim: Our aim is to leverage inferential commonsense knowledge to enhance the comprehension of news headlines by simplifying, interpreting, and explaining events that are not explicitly stated in their short texts. We intend to use this knowledge to design a knowledge-infused learning framework for enhancing the prediction of news reporting bias. To fully utilise its potential and deal with any unnecessary noise it may introduce, we also intend to emphasise significant inferences using the attention mechanism.

- **Hypothesis 3:** ***Translate-retrieve-translate strategy facilitates the efficient use of inferential knowledge as a source of additional information in a multilingual context for the task of news bias prediction.***

  State of the art: Inferential commonsense knowledge is not tied to any specific language. Although it is language-agnostic, the majority of knowledge sources for generating commonsense are available for languages with abundant resources, such as English (Hwang et al., 2021). To overcome this barrier, related research acknowledges the potential of using the translate-retrieve-translate strategy (Fang et al., 2022). Nonetheless, no attention has been paid to modelling this strategy for comprehending short multilingual headlines in low-resource languages for the task of bias prediction.

  Aim: Our aim is to leverage the translate-retrieve-translate strategy to facilitate the efficient use of inferential knowledge in a multilingual, low-resource language-agnostic setting. We aim to use it as a source of additional information in the proposed multilingual framework for predicting bias in multilingual news headlines. We also intend to evaluate the generalisability of our framework and conduct a qualitative analysis of its performance across languages, accompanied by detailed explanations.

1.3 Scientific Contributions

The key scientific contributions of the thesis are as follows:

- **SC 1. A Novel Data Generation Method:**

We present the design and provide the implementation of adaptable data generation scripts for generating customised datasets for related tasks. Additionally, we introduce a publicly available, novel event-centric news dataset generated with a wide range of features. We expand on it in Chapter 2 and publication (Swati & Mladenić, 2021).

- **SC 2. Commonsense-Infused Bias Prediction Framework:**

  We identify the importance of inferential commonsense knowledge in facilitating the comprehension of short news headlines. Using this knowledge, we introduce a neural network framework for bias prediction and evaluate the impact of selective knowledge augmentation on prediction performance. Our work is the first to harness this knowledge for enhancing bias prediction in short news headlines. We provide additional insights in Chapter 3 and publication (Swati, Mladenić, & Grobelnik, 2023).

- **SC 3. Multilingual Language-Agnostic Bias Prediction Framework:**

  We propose to leverage multilingual commonsense knowledge through a translate-retrieve-translate strategy. Expanding upon our earlier framework, as defined in SC 2, we introduce a knowledge-infused, language-agnostic learning framework. It aims to enhance the prediction of political bias in multilingual news headlines, particularly under the setting of an imbalanced sample distribution for under-resourced languages. We also evaluate the impact of knowledge infusion on prediction performance across languages. We elaborate on it in Chapter 4 and publication (Swati, Grobelnik, et al., 2023).

1.4 Organisation of the Thesis

The thesis begins by introducing the challenges associated with enhancing bias prediction in news reporting. It seamlessly transitions into an in-depth analysis and design of methodologies aimed at addressing these challenges. Shedding light on its complex aspects, Chapter 2 explores the multifaceted realm of news reporting bias. By critically assessing existing research, it lays the foundation for comprehensive methodologies for mitigating bias in news reporting. It is organised into four distinct sections. Section 2.1 begins by navigating through Artificial Intelligence (AI) and Machine Learning (ML) methodologies aimed at uncovering news reporting bias. Section 2.2 introduces a customisable data generation method and provides an event-centric dataset for analysing complex event-outlet relationships. It discusses the findings presented in the paper *EveOut: an event-centric news dataset to analyse an outlet's event selection patterns* (Swati & Mladenić, 2021) and the scientific contribution SC 1. Section 2.3 presents a novel method for addressing the task of predicting outlets, as described in Swati and Mladenić (2020). Section 2.4 analyses the impact of geographical bias on the emotions conveyed in news articles, as described in Swati et al. (2021).

Chapter 3 presents our second scientific contribution (SC 2), also published in Swati, Mladenić, and Grobelnik (2023), and investigates the impact of leveraging inferential commonsense knowledge to address the challenges of predicting political bias in short news headlines. Chapter 4, also published in Swati, Grobelnik, et al. (2023), addresses our third scientific contribution (SC 3) by expanding the challenge introduced in Chapter 3 to a multilingual setting having languages with limited resources. It addresses the challenge by leveraging commonsense knowledge through a translate-retrieve-translate strategy. The thesis concludes with Chapter 5, which reviews the key scientific contributions and outlines future objectives and research directions.

Chapter 2

News Reporting Bias

This chapter explores the multifaceted realm of news reporting bias, focusing on various aspects such as bias in event selection, headline and content, sentiment, political orientation, and news aggregation. It also endeavours to investigate the main challenges by critically assessing the existing research in this field. By acknowledging the limitations, this chapter sets the stage for more comprehensive and sophisticated approaches to mitigate news reporting bias.

The first section navigates through Artificial Intelligence (AI) and Machine Learning (ML) methodologies to uncover news reporting bias. The following sections present the papers authored and published during the course of this thesis. Each of these papers sheds light on distinct yet interconnected concepts. They collectively provide a comprehensive analysis of different aspects of bias in news reporting. In particular, they offer detailed insights into the relationships between events and news outlets, methodologies for predicting outlet biases, and the impact of geographical boundaries on the sentiment of news coverage.

The second section introduces a reproducible event-centric dataset designed to explore and analyse the complex relationships between events and news outlets. The subsequent section presents a novel approach to address the task of outlet prediction. The last section concludes the chapter with an examination of the impact of geographical bias on the sentiments embedded in news articles. Using the legacy articles of the London and Rio Olympics as a case study provides valuable insights into how geographic context influences news sentiment.

2.1 AI/ML Approaches to Analyse Media Bias

The presence of bias in news reporting stems from the crucial phase of event selection, which plays a pivotal role in shaping news coverage (Martin, 2005). It is important for a journalist to know which event is worthy to be published (Lundman, 2003) Even readers would be interested in knowing the factors that affect this selection (Knobloch-Westerwick et al., 2020). However, instead of identifying the differences in topic coverage, the existing methods aggregate the related articles (Hamborg et al., 2020). They then outsource the task of bias identification to users, forcing them to determine the bias on their own. It could be a better approach to provide an overview of the topic and to visualise differences in coverage. Yet it is certainly not an automated approach that exposes the outlet's event selection bias to its readers.

Content-based news-bias analysis tasks still depend on premature text analysis methods such as keyword or frequency analysis (Cho et al., 2021; Nah et al., 2023). They are far from state-of-the-art methods specifically for complex Natural Language Processing

(NLP) tasks (Raza et al., 2024). For example, Word Sense disambiguation and other advanced, task-specific NLP text-analysis techniques are still underutilised (Rakhecha et al., 2023). Additionally, approaches to content-based bias analysis require considerable human expertise (Lazaridou et al., 2020). Manual intervention is needed to screen and annotate the texts systematically, as existing NLP tools are not powerful enough to identify and annotate individual instances of biassed texts. A notable limitation also lies in the lack of investigation from an unsupervised perspective and the development of semi-automated techniques to address these challenges inherent in content-based bias analysis in news reporting (Mowery, 2021; Stefanov et al., 2020).

Tasks that rely on sentiment words often overlook the nuanced sentiment conveyed through sentence structure (W.-F. Chen et al., 2020). Also, a negative quote does not automatically mean that the whole article is negatively biassed (Tan et al., 2023), as authors often choose criticism in a sarcastic way. Besides that, explicitly stated sentiment words used for news bias determination are not task-specific (Balahur et al., 2013). They may not even represent the bias-inducing words. These limitations highlight the challenges that result from the lack of a comprehensive list of news-specific sentiment- and bias-inducing words. The absence of such a resource limits the effectiveness of advanced sentiment analysis tools in detecting subtle biases conveyed through sentence structure and sarcasm (Bestvater & Monroe, 2023).

Studies that aim to identify political orientations are typically based on explicitly defined features, such as quotes from think tanks (Groseclose & Milyo, 2005). However, articles might not cite a specific policy group but still exhibit political favouritism (Dunham, 2013). For instance, one party may be disproportionately criticised for building up another party's reputation. A number of features are therefore required to be considered to determine political leaning. Moreover, the majority of these models classify the articles as liberal/conservative (F. Ribeiro et al., 2018) or right/left/centre (Baly et al., 2020; D'Alessio, 2012). The absence of scores assigned to these classes to quantify the bias of the articles poses yet another constraint. Another limitation lies in the absence of a finer-grained classification model, which could offer deeper insight into the diverse ideologies represented in news articles. The emphasis on identifying biases against certain cultural aspects, such as those specific to Muslims (secular or Islamic), also confines related research. Consequently, it fails to account for potential biases in other cultural dimensions (Knott & Poole, 2016; Mahmood, 2009).

Authors tend to publish facts from their viewpoints, which may result in biassed news (Andrew, 2007; Ecker et al., 2014). It may influence the thought process of the readers (McCarthy & Dolfsma, 2014). For example, any political controversy presented from a specific perspective may alter the voting pattern (Chiang & Knight, 2011). However, only a few of the existing research considers the effect of media bias on the perceptions or actions of readers (Bursztyn et al., 2020; Mastrorocco & Minale, 2018). Furthermore, news stories are transient in nature and evolve over time (Vaca et al., 2014). For instance, some issues, such as natural disasters, may arise during the election period, and some other issues, such as terrorism, may arise during the next election. Factors like the dynamic nature of news narratives and the potential long-term effects of media bias on readers' perceptions make the generation of annotated datasets for news bias analysis challenging. In addition, generating such comprehensive and representative datasets requires extensive annotation and curation, which can be both resource-intensive and time-consuming (Fan et al., 2019). Furthermore, ensuring that such datasets can be generalised across diverse news sources and events adds to the complexities. These limitations underscore the need for adaptable data generation methodologies for generating custom datasets for related tasks to better reflect the multifaceted nature of media bias.

As far as the availability of datasets is concerned, there is a lack of large publicly available annotated data that can be generalised across machine learning models. In fact, the existing datasets are category-dependent (politics, healthcare, disasters, etc.), and that too for a specific time period (weeks, months, or years), which tends to make the generalisation even more difficult (Baly et al., 2020; Cheng et al., 2020). Whereas most of the research work seeks to strengthen mono-lingual news bias analysis techniques (Krieger et al., 2022; Naredla & Adedoyin, 2022), multilingual and cross-lingual bias analysis is given less attention. Although promising, developing such models capable of capturing temporal patterns to determine news reporting bias over time presents difficulties in effectively capturing nuanced shifts in bias across different languages and temporal contexts. Furthermore, developing such models requires addressing complex linguistic and cultural variations, as well as ensuring scalability and reliability across multiple news sources and languages. Therefore, these inherent limitations limit the practical implementation of cross-lingual models, despite their potential.

2.2 Event-Outlet Relationship

This section presents the research published in the article titled *EveOut: an event-centric news dataset to analyze an outlet's event selection patterns* (Swati et al., 2021). The paper was published in the scientific journal *Informatica - an International Journal of Computing and Informatics*, Volume 45, in 2021.

News outlets select events based on their newsworthiness, potentially leading to event selection bias, also known as gatekeeping bias (Pingree et al., 2013). To study and understand this complex relationship between events and news outlets, we introduced EveOut, a novel event-centric news dataset. The methodology involved the creation of flexible data-generation scripts. These scripts were designed not only to facilitate the development of EveOut and its subsequent iterations but also to expedite the process of generating datasets tailored to custom tasks.

The investigation entailed statistical analyses of selected English and Slovenian news outlets, aiming to uncover biases in the event selection process. The investigation examined the impact of geographical, temporal, and categorical preferences on news coverage patterns. Additionally, it highlighted a promising application: the outlet prediction task. It also demonstrated how employing sophisticated conditional probabilistic models can enable the estimation of correlations between outlets. Further investigation encompassed a thorough examination of event selection patterns among the outlets. The findings underscore significant patterns in event selection among the outlets. Specifically, the analysis revealed 'Politics' to stand out as the most extensively covered category, drawing considerable attention. In contrast, 'Environment' received notably minimal coverage, indicating a potential gap in the attention dedicated to such topics. In particular, outlets exhibited a clear preference for geographically relevant events, shedding light on their selective focus. The bias towards events with strong geographical ties also underscores the outlets' preferences in their selection process.

Although the study focused on specific parameters, such as geographical and categorical biases, it provided a foundational framework for understanding inherent biases in event selection. On the other hand, the dataset's availability has empowered researchers to develop tools that can enhance methods for digital journalism concerning bias. The study not only provided valuable insights but also uncovered unexplored avenues for further investigation, including the importance of features such as event descriptions to help identify the underlying biases inherent in the event selection process. In conclusion, the study holds the potential to be used in automated solutions to provide an overview of the events and visualise differences in coverage, as it is important for journalists to know which events are worthy of publication and which factors influence the selection process.

EveOut: an Event-centric News Dataset to Analyze an Outlet's Event Selection Patterns

Swati, Dunja Mladenić and Tomaž Erjavec
Jožef Stefan Institute and Jožef Stefan International Postgraduate School, Ljubljana, Slovenia
E-mail: swati@ijs.si, dunja.mladenic@ijs.si, tomaz.erjavec@ijs.si

Keywords: dataset, news event analysis, event selection bias, news coverage, gatekeeping bias, outlet prediction, digital humanities

Received: January 10, 2021

Automation of computational models to study the structure of events and their value to news outlets is an effective way to understand event-outlet relationships. However, the scarcity of publicly available, comprehensive event-centric news datasets restricts the implementation of such models. To overcome this bottleneck, we collected seventeen months of event data using Event Registry to generate EveOut, a publicly available event-centric news dataset. To conduct statistical analysis, we first select five English-language and three in Slovenian-language news outlets. We then retrieved all the events covered by them and used it to document the prevalence of geographical, temporal, categorical, and several other aspects of the event selection bias by these outlets. We illustrate the significance of our dataset in the field of digital humanities by identifying a motivating use case. The dataset is publicly available from the dedicated website `http://cleopatra.ijs.si/EveOut/`, which provides a detailed description of the fields, usage information, and a link to the GitHub repository.

Povzetek: V prispevku predstavljamo EveOut, javno dostopno množico podatkov, zgrajeno na osnovi dogodkov, o katerih poročajo mediji. EveOut je zasnovana za pomoč pri analizi in razumevanju zapletenega odnosa medijev do poročanja o posameznem dogodku. Zgrajeno množico podatkov smo tudi uporabili za raziskovanje geografskih, časovnih, vsebinskih in drugih vidikov pristranskosti nekaj izbranih poročevalcev pri izbiri dogodkov, o katerih poročajo.

1 Introduction

News outlets are constantly confronted with the task of selecting events to be reported on. This selection is based on the newsworthiness of an event which can be defined by the presence or absence of several news values such as the inclusion of the power elites, the relevance, and popularity of the topic, etc. Determining the news value for an outlet may result in a selection bias, also known as gatekeeping bias [7]. A journalist, for instance, is more likely to report on an event that includes fresh data on an existing and trending event.

Gatekeeping bias can be significantly reduced by studying and analyzing the correlation and impact of different features on the selection of events by the outlets and then using the knowledge to automate the event selection process. Computational models for the study of complex event-outlet relationships can help explore the strategies for selecting publishable events and automating the event selection process. However, to stimulate the development of these models, the availability of data on news events and their relevant details is necessary.

In this paper, we present EveOut, the first, large, publicly available event dataset generated by leveraging the events collected using EventRegistry [4]. The resulting dataset, called "EveOut", consists of $81,562$ news events in English and Slovenian language, with a varied range of features retrieved for the period between January 2019 and May 2020.

We hope that EveOut will serve as a benchmark dataset to study the event-outlet correlation and help mitigate the impact of implicit bias present in the production and reporting process. We also hoped that it will encourage the publishers and others involved in the news production process to develop tools to enhance digital journalism and facilitate research in this field.

The contributions of this paper are as follows:

- We present EveOut, a publicly available novel event-centric news dataset generated with a wide range of features using the EventRegistry platform.
- We provide flexible dataset generation scripts that facilitate the generation of custom versions of EveOut with the required features.
- We identify a potential use-case 'outlet prediction task'. For the task, we then illustrate how conditional probabilistic models can be used to estimate the correlation between outlets.
- We present a detailed statistical analysis with respect to multiple event features to compare, contrast, and infer the coverage pattern of the selected news outlets publishing in English and Slovenian languages.

2 Related work

A number of datasets are based on news articles but, to the best of our knowledge, only a few datasets that explicitly focus on event-centric data have been proposed. GDELT (Global Data on Events, Location, and Tone) [5] is a CAMEO-coded [6], open, and large-scale news dataset that tracks the news media around the world in multiple languages. The articles are then compiled into a list of events, for rich and insightful event analytics. However, there is no attribute in the dataset that would specify the outlet(s) for the event. As a result, there is a lack of information in GDELT that is crucial to the study of the event-outlet relationship that forms the foundation of our dataset.

In terms of availability, publicly available event datasets are scarce. There is some related research on event data [3, 1], but the datasets extracted/generated for the experiments are not publicly available. Besides, the majority of the existing event datasets [2] are category-dependent (*politics, healthcare, disaster, etc.*) which renders them useful for specific research purposes only. EveOut addresses these bottlenecks by introducing a generalized publicly available event-centric news dataset.

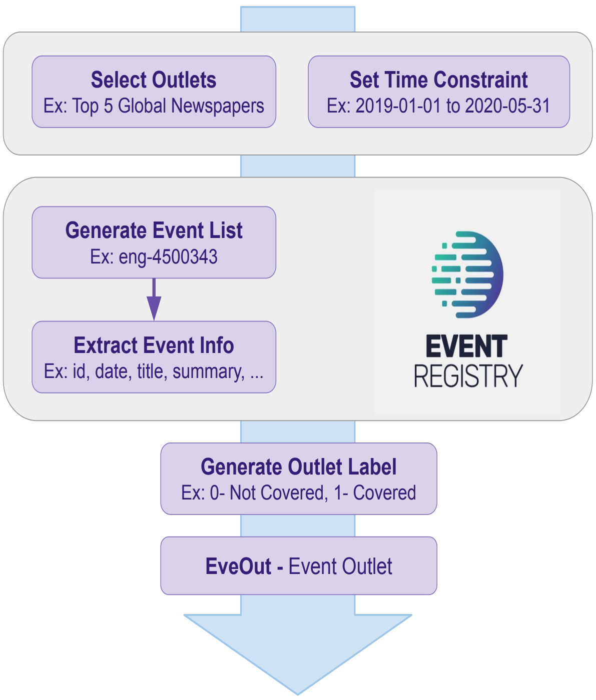
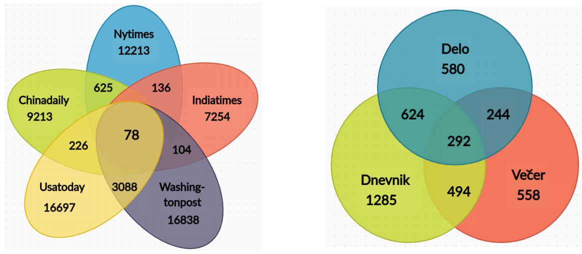


Figure 1: EveOut generation process, composed of user selection of the outlets and the time period, automatic extraction of the data from Event Registry and labeling of the extracted data.



(a) English news outlets. (b) Slovenian news outlets.

Figure 2: Distribution of event coverage by the outlets.

3 Data description

3.1 Raw data source

Event Registry[1] [4] monitors, collects, and delivers news articles from news sources around the world in more than 30 languages. It extracts semantic information from the articles and if the same event is described in multiple articles, it aggregates them into clusters using several clustering algorithms. These article clusters are referred to as events. For instance, *"Trump threatens to shut down social media firms"* is an event recorded internationally in more than 1,220 news articles. Each event is then annotated with various metadata, such as a unique id to track the coverage of the event, topic, categories to which it may belong, geographical location, sentiments, etc. As a result, its large-scale temporal coverage can be used effectively to study the event-outlet relation.

3.2 Data generation process

For the data generation process, as depicted in Figure 1, we first selected five English and three Slovenian news outlets (for the sake of simplicity, we refer news outlets publishing in English/Slovenian language as English/Slovenian news outlets throughout the paper). We selected these outlets following the work in [8] which is based on Alexa Global Rankings of top news outlets.

We then used an explicit temporal query (Q_t) to retrieve all events in all news categories using Event Registry API. $Q_t = \{Q_{text}, Q_{time}\}$ consists of the text component Q_{text} and the time component Q_{time}. Next, we set the time limit $Q_{time} = [Q_{sd}, Q_{ed}]$ for extracting events that occurred within the specified time where, Q_{sd} = '2019--01-01' and Q_{ed} = '2020-05-31' signify the event's start date and end date. Since the outlet's event selection policy may change over time, we selected this time frame as recent data tends to be more reliable in predicting event coverage patterns. We then set $Q_{text} = \{Q_{out}, Q_{lang}, Q_{cat}\}$ where, Q_{out}[2], Q_{cat}[3], and Q_{lang}[4] represents the list of out-

[1] https://eventregistry.org
[2] https://eventregistry.org/documentation?tab=suggSources
[3] https://eventregistry.org/documentation?tab=suggCategories
[4] https://github.com/EventRegistry/

Attribute	Description
uri	a unique event identifier
title	event title in the specified language
event_date	date in yyyy-mm-dd format
sentiment	event sentiment
categories	event categories
loc_country	country where the event occurred
loc_continent	continent where the event occurred
total_article_count	total number of articles published
article_count	total number of articles published in the specified language
summary	summary of the event
outlet_list	list of outlets that reported the event

Table 1: Description of the dataset attributes.

P(O_1|O_2)	**Nytimes**	**Indiatimes**	**Washingtonpost**	**Usatoday**	**Chinadaily**
Nytimes	1.00	0.09	0.28	0.24	0.19
Indiatimes	0.03	1.00	0.03	0.03	0.01
Washingtonpost	0.33	0.09	1.00	0.26	0.19
Usatoday	0.27	0.09	0.25	1.00	0.13
Chinadaily	0.10	0.01	0.08	0.06	1.00

(a) English news outlets.

P(O_1|O_2)	**Delo**	**Dnevnik**	**Večer**
Delo	1.00	0.33	0.33
Dnevnik	0.51	1.00	0.49
Večer	0.30	0.29	1.00

(b) Slovenian news outlets.

Table 2: Conditional probability of an event to be covered by an outlet (in rows), provided it is covered by another outlet (in columns).

lets, categories, and languages respectively.

For English news outlets, we set $Q_{out} = \{$'$nytimes.com$', '$indiatimes.com$', '$washington$-$post.com$', '$usatoday.com$', '$chinadaily.com.cn$'$\}$ and $Q_{lang} = \{$'eng'$\}$ and we set $Q_{out} = \{$'$delo.si$', '$dnevnik.si$', '$vecer.com$'$\}$ and $Q_{lang} = \{$'slv'$\}$ for Slovenian news outlets. We fixed $Q_{cat} = \{$'$news/Politics$', '$news/Business$', '$news/Sports$', '$news/Arts\ and\ Entertainment$', '$news/Science$', '$news/Technology$', '$news/Health$', '$news/Environment$'$\}$ to represent the news categories. If an event falls into more than one category, it is labeled with multiple categories.

We first excluded events from the extracted event list that weren't covered by any of the selected outlets. We then extracted individual outlets from the event's outlet list and generated a column in each dataset to denote individual outlets. We used a binary scalar value to indicate whether the outlets covered the event or not. Table 1 describes the attributes of the generated dataset. From Figure 2, it is apparent that the event coverage by the outlets is not uniform.

4 Availability and reusability

For ease of discovery and preservation, EveOut is archived as an online resource at ***https://doi.org/10.5281/zenodo.3953878***. It is well documented in accordance with the requirements of the *FAIR Data principles*[5] and is freely accessible under the *Creative Commons Attribution 4.0 International license* to make it reusable for nearly any purpose. For dataset regeneration, the GitHub repository at ***https://github.com/Swati17293/EveOut*** gives the source code of the collection process. For an in-depth analysis, a separate web page with detailed statistics and illustrations can be found at ***http://cleopatra.ijs.si/EveOut/***.

The resource is currently being used in several studies within a larger research project[6]. A major part of this project aims to provide a temporal, cross-lingual analysis of concepts around different events, exploring how language impacts the mediatic narratives built by the media. Since EveOut serves as the basis for the study and analysis of events and their attributes, it is ideally suited to the project needs.

5 Potential use case - outlet prediction

Outlet Prediction is the task of estimating the probability that an event will be covered by an outlet. In addition to allowing the publishers of the outlets to evaluate the significance of the event, this task is intended to benefit independent editors who prefer to report on events covered by mainstream outlets. It can be best assessed by calculating the conditional probability P of an event covered by an outlet O_1 given that it is already covered by another outlet O_2 using the following equations.

$$P(O_1|O_2) = \frac{P(O_1 \cap O_2)}{P(O_2)}, \text{ if } P(O_2) > 0 \quad (1)$$
$$= 0, \text{ if } P(O_2) = 0 \quad (2)$$

Table 2a shows that apart from *'Indiatimes'* and *'Chinadaily'*, rest of the outlets tends to overlap each other in terms of event coverage. It is also interesting to note from Table 2, that among the listed outlets, the likelihood of any outlet to cover an event, given that it is already covered by any other outlet is higher (higher P values) for Slovenian outlets.

Unlike the selected English outlets which are supposedly global, the selected Slovenian outlets are the major outlets in Slovenia which is a small country. This difference influences the coverage pattern of the outlets, which reveals how regional priorities affect the event selection process. For instance, $P(Dnevnik|Delo) = 0.51$ and $P(Dnevnik|Večer) = 0.49$ which is quite high as compared to the others which indicates that if an event is covered by either *'Delo'* or *'Večer'* it is highly probable that it will be covered by *'Dnevnik'*.

6 Statistics and analysis

The statistical analysis of our dataset with regard to the distribution of events between the outlets is summarized and visualized in this section.

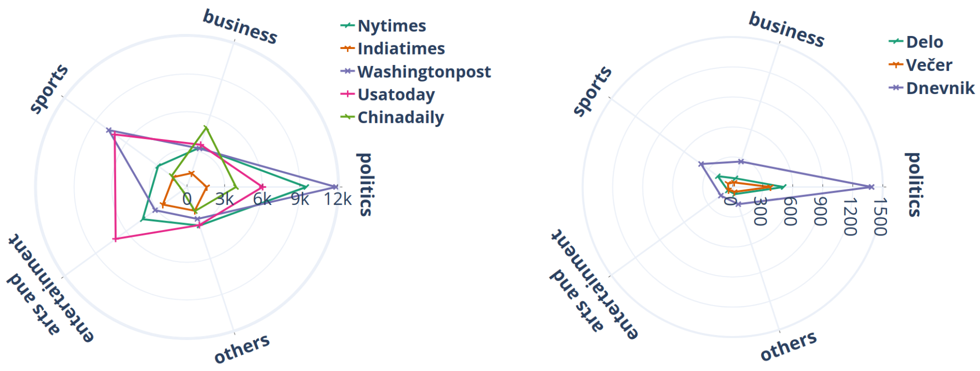
Figure 3a and 3b represents the distribution of event categories covered by the English and the Slovenian news outlets. It is evident from the distribution that each English news outlet focuses on a different event category other than *'Politics'*. For instance, *'Indiatimes'* focuses more on events related to *'Arts and Entertainment'*, whereas *'Chinadaily'* tends to cover more *'Business'* related events. In contrast to English outlets, the event coverage by Slovenian outlets is similar in addition to *'Politics'* focusing on *'Sports'* and to some extent on *'Business'*.

By plotting the proportion of event coverage over time, as shown in Figure 4, the pattern of event coverage by the outlets can be better visualized. In particular, *'May* 2020' contrasts the percentage of event coverage by the English and Slovenian news outlets. Moreover, unlike other news outlets, coverage of events by *'Usatoday'*, *'Washingtonpost'*, and *'Večer'* is somewhat inconsistent. A substantial decline in the coverage of *'Washingtonpost'* in *'May* 2020' is also noteworthy in the graph. It is due to its event preference which is evident from its radial graph in Figure 3a. Its coverage is skewed towards *'Politics'* and *'Sports'* which alone represents around 50% of events in the dataset. However, this percentage dropped to 40% in *'May* 2020', and as

event-registry-python/wiki/Supported-languages

[5] http://www.nature.com/articles/sdata201618/

[6] http://cleopatra-project.eu/



(a) English news outlets. (b) Slovenian news outlets.

Figure 3: Category-wise distribution of event coverage by the outlets. (Category 'others' includes: environment, health, science, and technology)

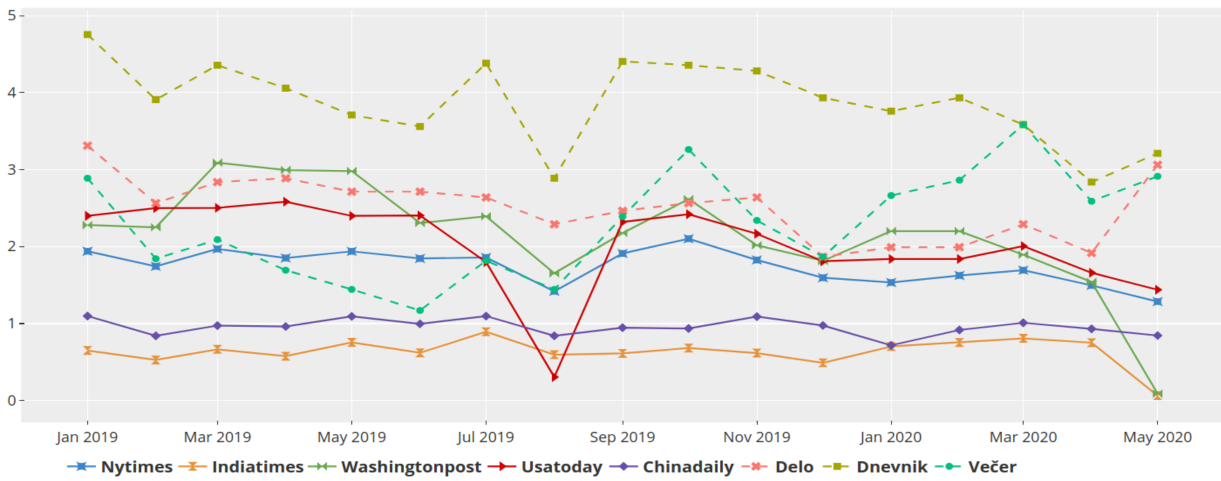


Figure 4: Distribution of the percentage of event coverage by the news outlets over time.

a result, its coverage declined substantially. In a nutshell, if the outlet favors a certain category of events and, in a specific time frame, events of that category are higher/lower than usual, it would be reflected in the outlet's coverage pattern.

Figure 5 reflects the inclination of the news outlets towards geographical bias which indicates that they prefer to cover events relevant to the geographical area in which they are based.

7 Conclusions and future work

In this paper, we presented a novel event-centric dataset EveOut for the study and analysis of complex event-outlet relationships. We also provide flexible data generation scripts, to speed up the development of future versions of EveOut. We also mentioned a potential use case to illustrate how the dataset could be used to study the event coverage patterns of the outlet and to estimate the correlation between the outlets using conditional probabilistic models.

We also conducted a statistical study to compare and contrast five English and three Slovenian outlets to examine their event selection patterns. We found that *'Politics'* is the most popular category, while *'Environment'* is the least popular category covered by the outlets. We also identified that news outlets, as expected, tend to cover geographically relevant events. In particular, we discovered that if the outlet favors a certain category of events and, if in a specific time frame, events collected of that category are higher/lower than usual, then this is reflected in the outlet's coverage pattern.

Although several features, such as event description, have not been analyzed in our study, it is expected that these features will also help to identify the inherent bias present in the event selection process. We hope that our dataset will not only help to discover and interpret event selection bias but will also help researchers to develop tools to enhance digital journalism.

Different news outlets may have different policies for selecting events. For example, some news outlets may want to publish only the top events of the day, while others may want to include exclusive global events. As part of our future work, an automated solution could be developed using

(a) English news outlets.

(b) Slovenian news outlets.

Figure 5: Distribution of country-wise coverage of events by the outlets. Notice the higher coverage density in (a) The USA, India, and China (b) Slovenia.

EveOut to provide an overview of the event and to visualize the differences in coverage, as it is important for journalists to know which event is worthy of publication and which factors influence the selection process.

In the future, it would also be interesting to have a distribution of articles with positive and negative sentiment for specific events and outlets. This would reveal not only the outlet's political orientation but also the editorial's overall attitude.

Acknowledgement

This work was supported by the Slovenian Research Agency under the project J2-1736 Causalify and co-financed by the Republic of Slovenia and the European Union's H2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 812997.

References

[1] Dylan Bourgeois, Jérémie Rappaz, and Karl Aberer. 2018. Selection bias in news coverage: learning it, fighting it. In *Companion Proceedings of the The Web Conference 2018*. International World Wide Web Conferences Steering Committee, 535–543. `https://doi.org/10.1145/3184558.3188724`.

[2] Cindy Cheng, Joan Barceló, Allison Spencer Hartnett, Robert Kubinec, and Luca Messerschmidt. 2020. Covid-19 government response event dataset (coronanet v. 1.0). *Nature Human Behaviour*, 1–13. `https://doi.org/10.1038/s41562-020-0909-7`.

[3] Felix Hamborg, Norman Meuschke, and Bela Gipp. 2018. Bias-aware news analysis using matrix-based news aggregation. *International Journal on Digital Libraries*, 1–19. `https://doi.org/10.1007/s00799-018-0239-9`.

[4] Gregor Leban, Blaž Fortuna, Janez Brank, and Marko Grobelnik. 2014. Event registry: learning about world events from news. In *Proceedings of the 23rd International Conference on World Wide Web*, 107–110. `https://doi.org/10.1145/2567948.2577024`.

[5] Kalev Leetaru and Philip A Schrodt. 2013. Gdelt: global data on events, location, and tone, 1979–2012. In *ISA annual convention*, 1–49.

[6] Philip A Schrodt, Omür Yilmaz, Deborah J Gerner, and Dennis Hermreck. 2008. The cameo (conflict and mediation event observations) actor coding framework. In *2008 Annual Meeting of the International Studies Association*.

[7] Stuart N Soroka. 2016. Gatekeeping and the negativity bias. In *Oxford Research Encyclopedia of Politics*. `https://doi.org/10.1093/acrefore/9780190228637.013.43`.

[8] Swati, Erjavec Tomaž, and Mladenić Dunja. 2020. Eveout: reproducible event dataset for studying and analyzing the complex event-outlet relationship.

2.3 Outlet Prediction

This section presents the research published in the paper titled *Are You Following the Right News-Outlet? A Machine Learning based approach to outlet prediction* (Swati & Mladenić, 2020). The paper was presented in proceedings of the 23rd International Multiconference Information Society – IS 2020, Volume C, in 2020.

Event selection bias encompasses the inherent bias observed in the process of selecting and reporting events by media outlets (Haselmayer et al., 2017). Expanding on the previous research (Swati et al., 2021), which identified the task of outlet prediction as a potential solution to address event selection bias in media outlets, this study introduced a novel neural network model designed to address these aforementioned challenges. The lack of any trustworthy automated method to predict outlets covering particular events of interest acted as the secondary driving force behind this study. The primary objectives were to empower readers to select and engage with diverse event perspectives and to provide media outlets with valuable insights into their reporting patterns. To address these, a model was designed to provide insights into the predictability of event coverage by major outlets while also addressing the issue of passive reliance on waiting for publication.

In addition, the study introduced a benchmark task of outlet prediction along with a tailored dataset of English news events. The evaluation results underscore the superior predictive ability of the proposed model compared to existing rule-based approaches, establishing a foundational baseline for future method evaluations. In particular, the model demonstrated superior performance compared to the uniform and stratified models. The subset accuracy improved by 0.41 points and 0.26 points for both the uniform and stratified models. Correspondingly, the hamming loss showed enhancements of 0.25 points and 0.15 points for the two models.

To further enhance the prediction quality, it is possible to integrate supplementary metadata derived from Wikipedia concepts into our model. Factors such as the speed at which news is reported, the duration of news coverage, the significance of events to media outlets, and the evolution of reporting styles over time are also worth investigating.

Are You Following the Right News-Outlet? A Machine Learning based approach to outlet prediction

Swati
swati@ijs.si
Jožef Stefan Institute
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia

Dunja Mladenić
dunja.mladenic@ijs.si
Jožef Stefan Institute
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia

ABSTRACT

In this work, we propose a benchmark task of outlet prediction and present a dataset of English news events tailored to the proposed task. Addressing this problem would not only allow readers to choose and respond to relevant and broader facets of events but also enable the outlets to examine and report on their work. We also propose a neural network based approach to recommend a list of probable outlets covering an event of interest. Evaluation results reveal that even in its simplest form, our model is capable of predicting the outlet significantly better than the existing rule based approaches. The proposed model will also serve as a baseline for evaluating approaches intended to address the task. Implementation scripts can be found at *https://github.com/Swati17293/outlet-prediction*.

KEYWORDS

News bias, Event Selection bias, News coverage, News Event Analysis, Recommendation System

1 INTRODUCTION

The advancement in the field of Natural Language Processing [9, 10, 5, 4] over the last decade, has made solutions to complex machine learning problems more convenient. The problems such as machine translation, text summarization, and segmentation are being solved much more efficiently than ever before. Consequently, it offered the researchers the opportunity to use these advanced techniques to solve problems in a variety of contexts such as news bias analysis. This analysis task is poised as the identification of the inherent bias present in the news production and its coverage process. It occurs when a news outlet publishes a news story selectively or incorrectly.

If the news is biased, then it can bias the thought process and decision making of the person listening, watching, and/or reading it [12]. It can have several direct or indirect implications whether political or social. For example, if the news shows only the positive or negative side of a political party; it has been observed to influence the public vote [2]. Not only politics but also the news about the disaster or spread of viral disease affects the belief system of the general public.

There are numerous events that happen continuously, and any form of bias can arise in numerous possible ways. It is not possible for any single outlet to capture every event. Thus, an outlet is forced to select a set of reporting events. Several factors, such as the geographical origin of the event, the involvement of an elite person or country, etc. influences such selection. Also the procedure requires rigorous monitoring of current affairs to determine the news value, and may result in event selection bias also known as gatekeeping bias.

However, no well-established automated method reveals to users the outlets that will cover the event of their interest. This drives the motivation of this study. The aim is to predict a list of outlets reporting on a given event. Addressing this problem would not only allow readers to choose and respond to relevant and broader facets of events but also enable the outlets to examine and report on their work. For instance, some outlets tend to publish events covered by well-established outlets. Instead of waiting for the news to be published, the proposed system will help them to get an insight into the degree of predictability of event selection by the major outlets.

1.1 contributions

We make the following contributions in this context:

- We propose a benchmark task of outlet prediction and present a dataset of English news events tailored to the proposed task.
- We provide a neural network model that can serve as a baseline for evaluating approaches intended to address the task.

The GitHub repository containing our code is available at ***https://github.com/Swati17293/outlet-prediction***.

1.2 Problem Statement

The problem is addressed as an outlet prediction task in which the bias is examined by comparing the learning ability of a classifier trained to predict the probability of event coverage by an outlet.

2 LITERATURE REVIEW

During the different stages of news production, various forms of news bias arise as described by Baker et al. [1]. The first stage begins with the selection of events also called gatekeeping, where an outlet selects or rejects an event for reporting. The selection process is driven by a number of factors, such as the geographical origin of the event, the involvement of an elite person or country, etc., and requires rigorous monitoring of current affairs to determine the news value. To our knowledge, only a few methods have been suggested that explicitly attempt to examine this bias.

Saez-Trumper et al. [11] attempted to identify bias in online news sources and social media groups surrounding them. They studied the disparity in the selection of events based on the quantity and exclusivity of stories published by 80 mainstream news

Information Society 2020, 5–9 October 2020, Ljubljana, Slovenia

outlets across the globe over a span of two weeks. From the review, it is found that there is a weak correlation between the quantity and exclusivity of news articles published by the outlets. It is also discovered that both the news and social media follow the same pattern of selection of events in similar geographical areas. However, media in the same region often choose the same events and publish similar-length posts.

Bourgeois et al. [3] used a matrix factorization method to extract latent factors that determine the selection of the event by an outlet. They combined the method with a BPR optimization scheme developed by Rendle et al.[8]. They used the events derived from the GDELT dataset and arranged the outlets in rows and their reported events in columns to form a matrix. Each cell value of the resulting matrix describes the selection/rejection of the event by the outlet.

For the bias analysis, they chose affiliation, ownership, and geographic proximity of the different outlets as the major factors. They suggest that each outlet follows its own latent preferences structure which facilitates the outlet to rank events. They also suggested that events should be selected such that the selected list should be diverse and should include a wide range of actively reported events. They thus adopted the method of Maximum Marginal Relevance which facilitates ranking based on the relevance and diversity of the events. It is discovered that event selection favors the most discussed topics rather than the unique ones.

F. Hamborg et al. [6] uses a matrix similar to the one created by Bourgeois et al.[3] Each cell in the matrix represent the most representative topic of the article reported by one country about the other. By spanning the matrix through outlets and topics in a region, the bias can be examined. They used a collection of 1.6 million articles from more than 100 countries over a two-month span from the Europe Media Monitor (EMM)[1] as their dataset.

Authors in [6] aggregates the related articles and then outsource the task of bias identification to the users, forcing them to determine the bias on their own. While the rest of the existing work analyzes the selection bias, it certainly does not present an automated approach suited to the outlet prediction task, unlike our work.

3 DATA DESCRIPTION

3.1 Raw Data Source

Event Registry[2] [7] monitors, collects, and provides news articles from news outlets around the world. It also aggregates them into clusters that are referred to as events. Each event is then annotated with several metadata such as unique id to track the event coverage, categories to which it may belong, geographical location, sentiment, etc. As a result, its large-scale temporal coverage can be used effectively to study the event selection process of news outlets.

[1]https://ec.europa.eu/knowledge4policy/
[2]https://eventregistry.org

3.2 Dataset

For our experiments, we first selected the top three news outlets based on Alexa Global Rankings [3]. We then used the Event Registry API to collect all news events reported in English between January 2019 and May 2020. We excluded events that were not covered by any of the selected outlets. We ended up with $51,409$ events for which we extracted basic information such as event id, title, summary, and source. Since the event coverage by these outlets is not uniform, which can be visualized in Figure 1, we used a stratified split to mimic this imbalance across the generated train-valid-test sets.

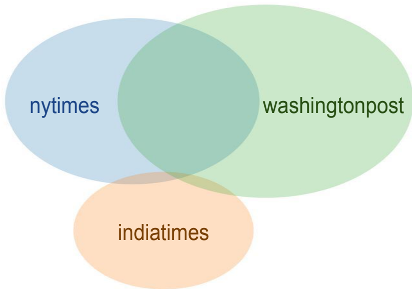


Figure 1: Distribution of event coverage by the outlets.

4 MATERIALS AND METHODS

4.1 Problem Modeling

For an event E and its associated pair (T, S), the task is to generate a list of outlets O expected to cover E. Here T is the event title and S is a short summary of the event as provided by the Event Registry. Mathematically, the task can be formulated as,

$$O = f(T, S, \alpha) \tag{1}$$

where, f is the outlet prediction function and α denotes the model parameters. O can have a well-thought-out variable length response generated from the list unique outlets O^l. For this work, $|O^l| = 3$.

4.2 Methodology

We extract feature vectors from T and S. We fuse them together to create a fused vector which is then passed through several layers to finally generate O. Figure 2 illustrates the entire prediction process. We further outline these tasks with more details in the following subsections.

4.2.1 Feature Extraction and Fusion. We used Google's *Universal Sentence Encoder* [4] *(USE)* to extract 128-dimensional feature vectors T' and S'. For feature fusion, we concatenated T' and S' and applied $tanh$ activation to generate F. We then used batch-normalization to increase the stability of the network and for regularization.

$$F = BN(tanh(T' \oplus S')) \tag{2}$$

In Eq 2, BN and $\oplus$ represents batch-normalization and concatenation respectively.

[3]https://www.alexa.com/topsites/category/Top/News/Newspapers
[4]https://tfhub.dev/google/universal-sentence-encoder/

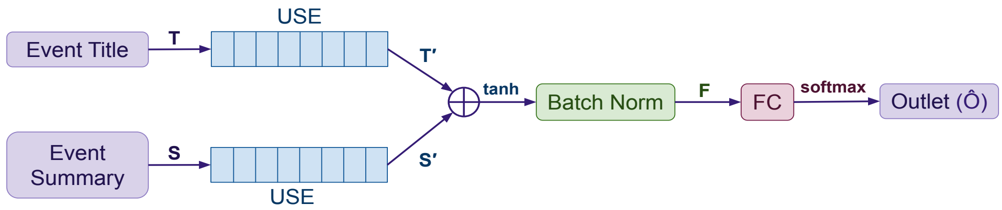


Figure 2: Outlet prediction process.

4.2.2 Outlet Prediction. We solve the problem using a multi-label classification model for which we create a separate outlet-index dictionary for outlets $D = \{o_1 : 1,\ o_2 : 2 \ldots\ o_n : n\}$, where n is the total number of unique outlets in O^l. To predict the list of outlets we pass F to the fully-connected layer (FC) having $softmax$ activation with n output neurons. Since an event can be covered by more than one outlet, we formulate the recursive prediction procedure as,

$$\hat{o} = \mathcal{P}(o_i|F, \hat{o}_{i-1}, b) = softmax(Fw_i + b_i) \tag{3}$$

$$= \frac{e^{Fw_i+b_i}}{\sum_{j=1}^{n} e^{Fw_j+b_j}} \tag{4}$$

where, $\hat{o}$ is the probability of selecting the i^{th} outlet (o_i) given F, bias (b), and the set of probabilities of previously predicted outlets ($\hat{o}_{i-1}$), and w is the weight. We use categorical cross entropy as the loss function as follows:

$$\mathcal{L}(o, \hat{o}) = -\sum_{j=1}^{n}\sum_{i=1}^{x}(o_{ij} * \log(\hat{o}_{ij})) \tag{5}$$

In Eq (5), for i^{th} outlet in the output sequence of length x, o_{ij} and $\hat{o}_{ij}$ denotes the actual and predicted probability of selecting the j^{th} outlet from D.

4.2.3 Hyper-parameters. We used Categorical accuracy[5] as the metrics to calculate the mean accuracy rate for multilabel classification problems across all the predictions. We consider a batch of size 128 and number of epocs as 100 for training. To optimize the weights during training we use Adam optimizer.

5 EXPERIMENTAL EVALUATION

5.1 Baselines

We use the following well-known and simplified methods as our baseline models.

- **Uniform:** Generate predictions randomly using a uniform distribution.
- **Stratified:** Generates predictions by respecting the class distribution of the training set.

5.2 Evaluation Metric

We aim to predict the list of outlets in this work. However, it is not necessary to predict the sequence in which outlets appear on this list. This is explained with an example given in Table 1. In other cases, a combination of correct and incorrect outlets may be predicted by the model.

We used the following metrics to evaluate the effectiveness of our model where, $\hat{o}$ is the predicted outlet, o is the true outlet, and N is the total number of instances.

[5]https://github.com/keras-team/keras/blob/master/keras/metrics.py

Table 1: Multiple correct predictions.

indiatimes nytimes washingtonpost
indiatimes washingtonpost nytimes

- **Subset Accuracy (a):** It measures the percentage of instances in which all of the outlets are correctly classified.

$$\text{Subset Accuracy } (a) = \frac{1}{N}\sum_{i=1}^{N}(\hat{o}_i - o_i) \tag{6}$$

- **Hamming Loss (ℓ):** It measures the fraction of the incorrectly predicted outlet to the total number of outlets. Since it is a loss function, its ideal value is 0.

$$\text{Hamming Loss } (\ell) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\hat{o}_i \cap o_i}{\hat{o}_i \cup o_i}\right| \tag{7}$$

5.3 Results and Analysis

Table 2 shows the comparison of our model with the baseline models in terms of subset accuracy and hamming loss.

Table 2: Comparison between the baseline models and our proposed model.

	Subset Accuracy	Hamming Loss
Uniform	0.140	0.526
Stratified	0.286	0.422
Ours	**0.546**	**0.275**

Quantitative analysis of the experimental results shows that, our model outperforms the Uniform and Stratified models by a margin of 0.41 and 0.26 points for subset accuracy and by 0.25 and 0.15 points for hamming loss respectively. The performance difference is clearly visible in Figure 3.

The intersection that we find among the different outlet pairs differs considerably as evident in Figure 1. This can be best seen by assessing the conditional probability of an event covered by an outlet given that it is covered by another outlet as listed in Table 3. For example, we can note that the $P(washington|nytimes) = 0.492$ which is quite high and indicates that *washingtonpost* tends to cover most of the events covered by *nytimes*. It is also interesting to note that *indiatimes* do not follow *washingtonpost* or *nytimes*, and vice versa.

6 CONCLUSIONS AND FUTURE WORK

It is important for a journalist to know which event is worthy enough to be published. Even readers would be interested to know

Table 3: Conditional probability of an event to be covered by an outlet, provided it is covered by another outlet.

P(x|y)	nytimes	indiatimes	washingtonpost
nytimes	1.000	0.067	0.364
indiatimes	0.034	1.000	0.023
washingtonpost	0.492	0.063	1.000

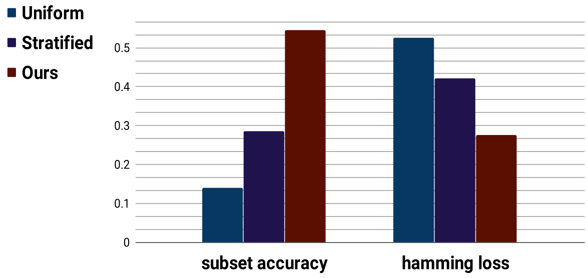


Figure 3: Comparison between the baseline models and our proposed model.

the outlets that are going to cover the event of their interest. Yet it is certainly not an automated approach, therefore in this work, we propose an approach to address the outlet prediction task given the event title and description. We also find that even in its simplest form, our model is capable of predicting the outlet. In the future, we intend to enhance our proposed model to better predict the outlets and to work in a cross-lingual setting. We plan to include a few more metadata provided by Event Registry (refer Section 3.1) along with Wikipedia concepts. We also plan to analyze the speed of reporting, time-span, and importance given to the events by the outlets. In addition, we will also be looking into how the outlets change their coverage style over time.

ACKNOWLEDGMENTS

This work was supported by the Slovenian Research Agency and the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 812997.

REFERENCES

[1] Brent H Baker, Tim Graham, and Steve Kaminsky. 1994. *How to identify, expose & correct liberal media bias.*

[2] Matthew Barnidge, Albert C Gunther, Jinha Kim, Yangsun Hong, Mallory Perryman, Swee Kiat Tay, and Sandra Knisely. 2020. Politically motivated selective exposure and perceived media bias, 82–103.

[3] Dylan Bourgeois, Jérémie Rappaz, and Karl Aberer. 2018. Selection bias in news coverage: learning it, fighting it. In *Companion Proceedings of the The Web Conference 2018*, 535–543.

[4] Li Dong, Nan Yang, Wenhui Wang, Furu Wei, Xiaodong Liu, Yu Wang, Jianfeng Gao, Ming Zhou, and Hsiao-Wuen Hon. 2019. Unified language model pre-training for natural language understanding and generation. In *Advances in Neural Information Processing Systems*, 13042–13054.

[5] Zihao Fu. 2019. An introduction of deep learning based word representation applied to natural language processing. In *2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI).* IEEE, 92–104.

[6] Felix Hamborg, Norman Meuschke, and Bela Gipp. 2018. Bias-aware news analysis using matrix-based news aggregation, 1–19.

[7] Gregor Leban, Blaz Fortuna, Janez Brank, and Marko Grobelnik. 2014. Event registry: learning about world events from news. In *Proceedings of the 23rd International Conference on World Wide Web*, 107–110.

[8] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. Bpr: bayesian personalized ranking from implicit feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence* (UAI '09). AUAI Press, Montreal, Quebec, Canada, 452–461. ISBN: 9780974903958.

[9] Sebastian Ruder. 2019. *Neural transfer learning for natural language processing.* PhD thesis. NUI Galway.

[10] Sebastian Ruder, Matthew E Peters, Swabha Swayamdipta, and Thomas Wolf. 2019. Transfer learning in natural language processing. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials*, 15–18.

[11] Diego Saez-Trumper, Carlos Castillo, and Mounia Lalmas. 2013. Social media news communities: gatekeeping, coverage, and statement bias. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, 1679–1684.

[12] Rune J Sørensen. 2019. The impact of state television on voter turnout. *British Journal of Political Science*, 257–278.

2.4 Geographical Boundaries and News Sentiments

This section presents the research published in the paper titled *Understanding the impact of geographical bias on news sentiment: A case study on London and Rio Olympics.* (Swati & Mladenić, 2021). The paper was presented in proceedings of the 24th International Multiconference Information Society – IS 2021, Volume C, in 2021.

Geography plays a pivotal role in shaping public perceptions of events (Mello et al., 2023), influencing various aspects like political affiliations and editorial independence, which in turn impact the sentiment of news articles (Nisar & Bleich, 2020; Pino et al., 2016). To address this, we explored the influence of geographical boundaries on the sentiments conveyed in news articles, focusing on the legacies of the Olympics in London and Rio as a case study. The study utilised the Event Registry API [1], to generate a comprehensive dataset comprising news articles from these locales published between January 2017 and December 2020. Based on the concept of article similarity, we then computed the cosine similarity scores of all possible article pairs, one from each set (London and Rio), to assess sentiment differences among similar article pairs.

The exploratory analysis unveiled the pivotal role that geographical boundaries play in shaping news sentiment. We extended the analysis to examine the disparity in relation to different news categories. The results highlighted substantial variations in sentiment, notably evident in categories such as politics, business, sports, and technology. Intriguingly, it was observed that political and business articles were significantly influenced by geographical factors, whereas science and technology articles appeared impervious to location-based sentiment shifts.

The descriptors used to portray positive and negative legacies in both Rio and London were also examined, aiming to analyse the cultural biases inherent in these depictions. Although we did not extend this analysis to a multilingual context in this study, it remains an intriguing avenue for future exploration. In essence, the analysis reveals the contrasting expectations between Rio as the first South American Olympic host and London, a representative of a wealthier nation. This exploration unravels the intricate societal and cultural perspectives confined by geographical boundaries, shedding light on how economic disparities contribute to distinct news sentiments.

[1] https://eventregistry.org

Understanding the Impact of Geographical Bias on News Sentiment: A Case Study on London and Rio Olympics

Swati
swati@ijs.si
Jožef Stefan Institute,
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia

Dunja Mladenić
dunja.mladenic@ijs.si
Jožef Stefan Institute,
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia

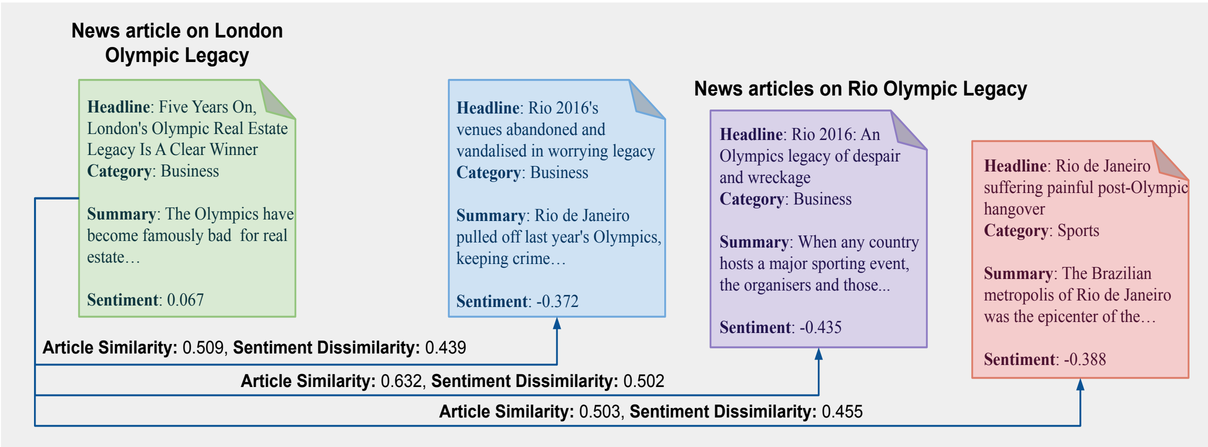


Figure 1: An example to illustrate the impact of geographical location on the sentiment of similar news articles.

ABSTRACT

There are various types of news bias, most of which play an important role in manipulating public perceptions of any event. Researchers frequently question the role of geographical location in attributing such biases. To that end, we intend to investigate the impact of geographical bias on news sentiments in related articles. As our case study, we use news articles collected from the Event Registry over two years about the Olympic legacy in London and Rio. Our experimental analysis reveals that geographical boundaries do have an impact on news sentiment.

KEYWORDS

Bias, News Bias, Geographical Bias, Olympics, Semantic Similarity, Sentiment Analysis, Dataset

1 INTRODUCTION

Claims of bias in news coverage raise questions about the role of geography in shaping public perceptions of similar events. Based on the geographical location, multiple factors, such as political affiliation, editorial independence, etc., can influence the way news articles are generated. Although it is well known that biased news can have more influence on people's thinking and decision-making processes [7, 9], it is nearly impossible to produce an article without any bias. Biased news articles have the potential to induce a variety of political and social implications, both direct and indirect. For instance, any political controversy presented from a specific perspective may alter the voting pattern [4, 1, 6].

There are different forms of news bias, and geographical bias is one of them. It exists if the sentiment polarity of similar articles published in different geographical location is contradictory or varies significantly. Sentiment analysis methods, which are commonly used to determine news bias [3, 14], can be used to examine the shift in sentiment polarity in similar news articles. Now, an intriguing question arises: Is geographical bias a factor affecting news sentiment? This study seeks to answer the above question by identifying and comparing sentiments of similar news articles. In doing so, we demonstrate how geographical location impacts the sentiments of similar articles. We also investigate this impact in relation to several news categories such as politics, business, sports, and so on.

The Olympic Games are a symbol of the greatest sports events in the world. Every edition leaves a number of legacies for the Olympic Movement, as well as unforgettable memories for each host city, whether positive or negative. In this regard, we select news articles about the Olympic legacy in London and Rio as a case study for our analysis.

We use Event Registry[1] [10] to collect English news articles, along with their sentiment and categories, published between January 2017 and December 2020. We use the popular Sentence-BERT (SBERT) [12] embedding to represent the articles and then compute the cosine similarity between them to identify similar article pairs.

Our data and code can be found in the GitHub repository at *https://github.com/Swati17293/geographical-bias.*

[1]https://eventregistry.org

Information Society 2021, 4 October 2021, Ljubljana, Slovenia

1.1 Contributions

The paper's contributions are as follows:

- We propose a task of analyzing the impact of geographical bias on the sentiment of news articles with data on the Olympic legacies of Rio and London as a case study.
- We present a dataset of English news articles customized to the above-mentioned task.
- We present experimental results to demonstrate the aforementioned impact of geographical bias.

2 RELATED WORK

The Majority of the sentiment analysis methods for news bias analysis depend on the sentiment words that are explicitly stated. SentiWordNet[2], which is a publicly available lexical resource used by the researchers for opinion mining to identify the sentiment inducing words that classify them as positive, negative, or neutral.

Melo et al. [5] collected and analyzed articles from Brazil's news media and social media to understand the country's response to the COVID-19 pandemic. They proposed using an enhanced topic model and sentiment analysis method to tackle this task. They identified and applied the main themes under consideration in order to comprehend how their sentiments changed over time. They discovered that certain elements in both media reflected negative attitudes toward political issues.

Quijote et al. [11] used SentiWordNet along with the Inverse Reinforcement Model to analyze the bias present in the news article and to determine whether the outlets are biased or not. The lexicons were first scored for the experiments using SentiWordNet and then fed to the Inverse Reinforcement model as input. To determine the news bias, the model measured the deviation and controversy scores of the articles. The findings lead to the inference that articles from major news outlets in the Philippines are not biased, excluding those from the Manila Times.

Bharathi and Geetha [3] classified the articles published by the UK, US, and India median as positive, negative, or neutral using the content sentiment algorithm [2]. The sentiment scores of the opinion words and their polarities were used as input to the algorithm.

Existing research investigates news bias using sentiment analysis methods, but, unlike our work, it does not provide a suitable automated method for analyzing the impact of geographical bias on news sentiment.

3 DATA DESCRIPTION

3.1 Raw Data Source

We use **Event Registry** [10] as our raw data source which monitors, gathers, and delivers news articles from all around the world. It also annotates articles with numerous metadata such as a unique identifier for article identification, categories to which it may belong, geographical location, sentiment, and so on. Its large-scale coverage can therefore be used effectively to assess the impact of geographical bias on news sentiment.

3.2 Dataset

To generate our dataset, we use a similar data collection process as described in [13]. Using the Event Registry API, we collect all English-language news articles about the Olympic legacy in London and Rio published between January 2017 and December 2020. We consider an article to be about the Olympic Legacy in London/Rio if the headline and/or summary of the article contains the keywords '*London*'/'*Rio*', '*Olympic*', and '*Legacy*'.

For each article, we then extract the summary, category, and sentiment. The article summaries vary in length from 290 to 6,553 words. Sentiment scores ranges from -1 to 1. We select seven major news categories, namely business, politics, technology, environment, health, sports, and arts-and-entertainment, and remove the rest of the categories. After excluding the duplicate articles we end up with 8,690 and 5,120 articles about the Olympic legacy in London and Rio respectively.

[2]http://sentiwordnet.isti.cnr.it/

4 MATERIALS AND METHODS

4.1 Methodology

The primary task is to compute the average difference in sentiment scores between similar news articles about the Olympic legacies in Rio and London. The stated task can be subdivided and mathematically formulated as follows:

(1) Generate two distinct sets of news articles A_1 and A_2, one about the London Olympic legacy and the other about the Rio Olympic legacy. For each $a_i \in A_1$ find a list of $a'_j \in A_2$, where a_i is the i^{th} article in set $A_1 = \{(a_1, s_1), (a_2, s_2)...(a_n, s_n)\}$ and a'_j is the j^{th} article in set $A_2 = \{(a'_1, s'_1), (a'_2, s'_2)...(a'_m, s'_m)\}$ which is the closest match (c.f. Section 4.1.1) to a_i. Here, $n = |A_1|$ and $m = |A_2|$.
(2) For each list, calculate D_{ij} to represents the difference between the sentiment scores s_i and s'_j of the articles a_i and a'_j.
(3) Calculate the average difference D of sentiment scores.
(4) Calculate the percentage of similar article pairs with reversed polarity and those with unchanged polarity.

The secondary task is to assess the primary task with respect to news categories, i.e. to calculate the average difference D of sentiment scores for similar articles in each category.

In the following subsections, we discuss the tasks mentioned above in greater detail.

4.1.1 *Article Similarity.*

We embed the articles in sets A_1 and A_2 to construct sets $F_1 = \{f_1, f_2...f_m\}$ and $F_2 = \{f'_1, f'_2...f'_n\}$. While alternative embedding approaches can be utilized, in this study we select the popular Sentence-BERT (SBERT) [12] embedding to extract 768-dimensional feature vectors to represent the individual articles in F_1 and F_2.

For each article a_i in A_1, we compute the similarity score[3] between a_i and every article a_j in A_2 using the cosine similarity metric $Sim^{cos}(a_i, a'_j)$ (Eq 1). We consider articles a_i and a'_j to be similar only if their similarity score is greater than 0.5.

$$Sim^{cos}(a_i, a'_j) = \frac{f_i \cdot f'_j}{||f_i||\ ||f'_j||} \tag{1}$$

where f_i and f'_j represents the embedded feature vectors of article a_i and a'_j.

The similarity score ranges from -1 to 1, where -1 indicates that the articles are completely unrelated and 1 indicates that they are identical, and in-between scores indicate partial similarity or dissimilarity.

4.1.2 *Average Sentiment Dissimilarity.*

For every pair of similar articles a_i and a'_j, we calculate the difference D_{ij} between their sentiment scores s_i and s'_j. To calculate the average sentiment

[3]https://en.wikipedia.org/wiki/Cosine_similarity

Table 1: Category-wise confusion matrix to show the percentage of similar article pairs with respect to their sentiment polarity.

	Sports		Business		Politics		Environment		Health		Technology		Arts & Entertainment	
	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Pos	77	10	62	28	42	18	55	18	29	12	87	4	59	16
Neg	11	2	7	4	23	16	14	12	12	46	1	0	7	18

Table 2: Confusion matrix to show the percentage of similar article pairs with respect to their sentiment polarity.

	Positive	Negative
Positive	69	15
Negative	11	4

Table 3: Distribution of average sentiment difference across news categories for similar article pairs with identical category.

News category	Average Sentiment Difference
Sports	0.19
Business	0.20
Politics	0.18
Health	0.16
Environment	0.22
Technology	0.14
Arts and Entertainment	0.19

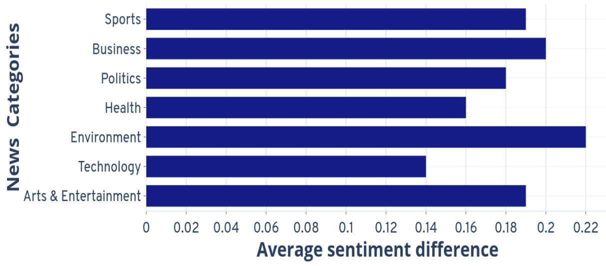


Figure 2: Distribution of average sentiment differences across categories for similar articles in the same category.

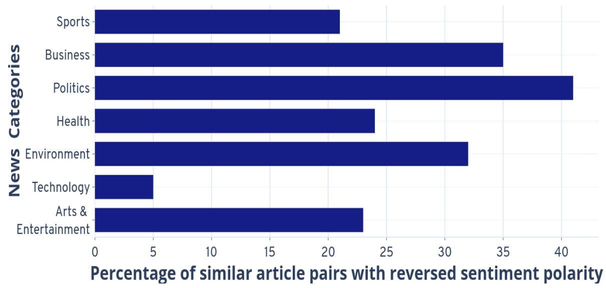


Figure 3: An illustration of the effect of category on sentiment polarity.

dissimilarity score D, we add all D_{ij} and divide it by the total number of similar article pairs.

5 RESULTS AND ANALYSIS

In our experiments, we compare 44,492,800 possible article pairs for similarity and discover 375,008 similar pairs. The comparison in terms of sentiment similarity reveals that if two articles from different geographical regions are similar, in our case Rio and London, the average difference in their sentiment scores is 0.171. In addition, as defined in Table 2, we calculate the percentage of similar article pairs based on their sentiment polarity. It's worth noting that the polarity of the article is completely reversed 27% of the time, indicating the impact of geographic region on sentiments.

It is because the success of mega-events such as the Olympics in a particular host city is heavily influenced by its residents' trust and support for the government [8]. It can be viewed positively as a national event with social and economic benefits, or negatively as a source of money waste. While the Olympics have left an economic and social legacy in London, a series of structural investment demands in Rio raise the question of whether or not the Olympics was worthwhile for the entire country.

5.1 Impact of news categories

The impact of news categories on the sentiments of similar articles with identical categories from different geographical regions is shown in Table 3. It demonstrates that certain news categories have a greater impact than others. Figure 2 depicts this distinction more clearly.

The categorical distribution of the percentage of similar article pairs in terms of sentiment polarity is shown in Table 1. *'Politics'* has the highest percentage of articles with reversed polarity, while *'technology'* has the lowest. Categories such as *'business'* and *'entertainment'*, though not as clearly as *'politics'*, exhibit the same bias.

This disparity arises from the fact that, in contrast to other categories, politics is most influenced by geographical boundaries, whereas science and technology are typically location independent. Since politics has such a large influence on shaping beliefs and public perceptions, it is frequently twisted to fit a particular narrative of a story. It is inherently linked to geographical borders, and it can be extremely polarizing depending on the geographical region.

6 CONCLUSIONS AND FUTURE WORK

In this work, we use news articles about the Olympic Legacy in London and Rio as a case study to understand how geographical boundaries interplay with news sentiments.

We begin by presenting a dataset of news articles collected over two years using the Event Registry API. We compute the cosine similarity scores of all possible embedded article pairs, one from each set of Olympic legacy articles (London and Rio). We use the popular Sentence-BERT for article embedding and then compute the sentiment difference between similar article pairs. From 44,492,800 possible article pairs we end up with 375,008 similar pairs.

In our analysis, we discovered that the sentiment reflected in similar articles from different geographical regions differed significantly. We also investigate this difference in relation to different news categories such as politics, business, sports, and so on. We find a significant difference in news sentiment across geographical boundaries when it comes to political news, while in the case of news in technology, the difference is much smaller. We find that articles in categories such as politics and business can be heavily influenced by geographical location, articles in categories such as science and technology are typically location independent.

In the future, we plan to identify the most frequently mentioned topics in the Olympic legacy corpus to see how they affect the news sentiment of articles about different geographical locations. Since our study is limited to English news articles, we intend to learn more about the role of cultures and languages in this bias analysis. We also intend to broaden our investigation to discover the adjectives used to describe the negative and positive legacies of Rio and London. Such an analysis would aid in understanding the expectations from cities such as Rio (the first in South America to host the Olympics) in comparison to London.

7 ACKNOWLEDGMENTS

This work was supported by the Slovenian Research Agency and the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 812997.

REFERENCES

[1] Dan Bernhardt, Stefan Krasa, and Mattias Polborn. 2008. Political polarization and the electoral effects of media bias. *Journal of Public Economics*, 92, 5-6, 1092–1104.

[2] Shri Bharathi and Angelina Geetha. 2017. Sentiment analysis for effective stock market prediction. *International Journal of Intelligent Engineering and Systems*, 10, 3, 146–153.

[3] SV Shri Bharathi and Angelina Geetha. 2019. Determination of news biasedness using content sentiment analysis algorithm. *Indonesian Journal of Electrical Engineering and Computer Science*, 16, 2, 882–889.

[4] Chun-Fang Chiang and Brian Knight. 2011. Media bias and influence: evidence from newspaper endorsements. *The Review of economic studies*, 78, 3, 795–820.

[5] Tiago de Melo and Carlos MS Figueiredo. 2021. Comparing news articles and tweets about covid-19 in brazil: sentiment analysis and topic modeling approach. *JMIR Public Health and Surveillance*, 7, 2, e24585.

[6] Claes H De Vreese. 2005. News framing: theory and typology. *Information Design Journal & Document Design*, 13, 1.

[7] John Duggan and Cesar Martinelli. 2011. A spatial theory of media slant and voter choice. *The Review of Economic Studies*, 78, 2, 640–666.

[8] Dogan Gursoy and KW Kendall. 2006. Hosting mega events: modeling locals' support. *Annals of tourism research*, 33, 3, 603–623.

[9] Daniel Kahneman and Amos Tversky. 2013. Choices, values, and frames. In *Handbook of the fundamentals of financial decision making: Part I*. World Scientific, 269–278.

[10] Gregor Leban, Blaz Fortuna, Janez Brank, and Marko Grobelnik. 2014. Event registry: learning about world events from news. In *Proceedings of the 23rd International Conference on World Wide Web*, 107–110.

[11] TA Quijote, AD Zamoras, and A Ceniza. 2019. Bias detection in philippine political news articles using sentiwordnet and inverse reinforcement model. In *IOP Conference Series: Materials Science and Engineering* number 1. Volume 482. IOP Publishing, 012036.

[12] Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, (November 2019). http://arxiv.org/abs/1908.10084.

[13] Swati, Tomaž Erjavec, and Dunja Mladenić. 2020. Eveout: reproducible event dataset for studying and analyzing the complex event-outlet relationship.

[14] Taylor Thomsen. 2018. Do media companies drive bias? using sentiment analysis to measure media bias in newspaper tweets.

Chapter 3

Knowledge-Infused Bias Prediction Framework

This chapter presents the research published in the article titled *An Inferential Commonsense-Driven Framework for Predicting Political Bias in News Headlines* (Swati, Mladenić, & Grobelnik, 2023). The paper was published in the scientific journal *IEEE Access*, Volume 11, dated July 2023. The journal holds a Q1 quartile ranking with an impact factor of 3.9 and is indexed in the Science Citation Index Expanded.

The task of identifying political bias in news headlines poses a formidable challenge, primarily stemming from their brevity, complexity, and lack of contextual information, which frequently fail to capture the subtle biases embedded within the underlying article (W.-F. Chen et al., 2018). To aid in the comprehension of short headlines by simplifying, interpreting, and explaining events that are not explicitly stated in the headlines, in this study, we proposed to leverage inferential commonsense knowledge (Sap et al., 2019). For the sake of simplicity, we henceforth refer to this knowledge as 'IC_Knwl' throughout this thesis. Despite the proven efficiency of IC_Knwl in numerous tasks (C. Chen et al., 2024; Ravi et al., 2023; Sabour et al., 2022), it remains underutilised for the task of decoding bias. Therefore, to fully exploit its potential in enhancing political bias prediction, we introduced a novel framework called IC-BAIT: Inferential Commonsense aware BiAs IdenTifier. It makes use of inferential knowledge to aid in the task of political bias prediction in news headlines. For a given short headline text, it acquires its associated IC_Knwl to train a classifier that maps this extended feature space of short texts into the political bias label set. The approach emphasises significant inferences using an attentive knowledge selection strategy (Majumder et al., 2020) while mitigating unnecessary noise. It is capable of harnessing both implicit knowledge acquired from pre-trained language models and explicit knowledge in the form of IC_Knwl simultaneously. To ensure its future adaptability, we designed the framework to be model-agnostic, enabling it to be bundled with cutting-edge linguistic models as they emerge. To acquire the necessary knowledge, we employed the neural knowledge model known as COMET (Hwang et al., 2021) trained on ATOMIC$^{20}_{20}$ (Hwang et al., 2021) knowledge graphs.

Addressing the absence of standardised, large-scale annotated datasets tailored for this task, we introduced two bias-annotated datasets: MediaBias and GoodNews. We generated MediaBias by utilising bias labels sourced from the bias rating portal allsides.com. On the other hand, we constructed GoodNews by extracting the headlines and their bias values from GoodNewsEveryone (Oberländer et al., 2020), a dataset based on the bias ratings obtained from adfontesmedia.com.

For a comprehensive evaluation and to determine the framework's generalisability, we conducted experiments on both datasets with multiple state-of-the-art pre-trained language models. The findings suggest that the baseline models exhibited better performance on both datasets when incorporated into our framework. Within our framework, these models exhibited noteworthy enhancements in accuracy ranging from 2.0% to 10.0%, macro-averaged F1 scores ranging from 2.2% to 22.2%, Jaccard scores showing an increase of up to 15.1%, and micro-averaged F1 scores showing an increase of up to 18.6%.

Furthermore, we conducted a qualitative analysis to assess the impact of selective knowledge augmentation on overall performance using the xAI framework LIME (Local Interpretable Model-Agnostic Explanations) (M. T. Ribeiro et al., 2016). The findings reveal the scenarios in which the selected knowledge is beneficial and when it is counterproductive. For instance, when such knowledge enables the model to focus on important entities, events, and explanations for unstated events, it results in enhanced predictions. On the other hand, incorporating it offers little or no enhancement if it lacks valuable information or is generic and insignificant. Additionally, we performed an in-depth error analysis, revealing that pairing the headline with a carefully chosen IC_Knwl enables the model to emphasise essential entities, events, and explanations for unstated events, resulting in enhanced predictive capabilities. In conclusion, our study exhibits potential not only in effectively identifying political bias in news headlines but also in comprehending the complex headline text.

IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 16 June 2023, accepted 17 July 2023, date of publication 25 July 2023, date of current version 7 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3298877

RESEARCH ARTICLE

An Inferential Commonsense-Driven Framework for Predicting Political Bias in News Headlines

SWATI SWATI[1,2], **DUNJA MLADENIĆ**[1,2], **AND MARKO GROBELNIK**[1]

[1]Jožef Stefan Institute, 1000 Ljubljana, Slovenia
[2]Jožef Stefan International Postgraduate School, 1000 Ljubljana, Slovenia

Corresponding author: Swati Swati (swati@ijs.si)

This work was supported in part by the enRichMyData European Union's Horizon Europe Project under Agreement 101070284, in part by the Slovenian Research Agency under Project J2-1736 Causalify, and in part by the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie Grant under Agreement 812997.

ABSTRACT Identifying political bias in news headlines holds significant importance as it influences the dissemination and consumption of news stories. However, employing conventional methods to do so poses a formidable challenge, as the short headline text is often complex and lacks sufficient syntactic and semantic context. Existing approaches fail to acknowledge the potential of commonsense reasoning in facilitating text comprehension, although it has been shown to aid numerous downstream applications. To this end, to facilitate comprehension and compensate for the lack of context, we propose leveraging inferential commonsense knowledge to simplify, interpret, and explain events that are not explicitly stated in headlines. Furthermore, to fully utilise its potential and deal with the unnecessary noise it may introduce, we present a method for emphasising significant inferences. Using this knowledge, we introduce a novel framework, **IC-BAIT**, short for **I**nferential **C**ommonsense aware **B**i**A**s **I**den**T**ifier, which is a flexible neural network framework designed to enhance political bias prediction in news headlines. We also present two bias-annotated datasets: MediaBias and GoodNews. Experiments on both datasets demonstrate that IC-BAIT significantly enhances the performance of the baseline models used in the framework. Experiments on the datasets show that IC-BAIT improves the baseline models in terms of accuracy (2.0-10.0%), macro-averaged F_1 (2.2-22.2%), Jaccard-score (up to 15.1%), and micro-averaged F_1 (up to 18.6%). Our in-depth qualitative analysis reveals the scenarios in which the selected knowledge is beneficial and when it is detrimental. Datasets and scripts are available at https://github.com/Swati17293/IC-BAIT.

INDEX TERMS Deep learning, inferential commonsense knowledge, media bias, news bias, NLP, short text classification.

I. INTRODUCTION

Media outlets often publish news stories that benefit the political party they endorse [1], [2]. Their bias is particularly reflected in news headlines, as readers are more likely to be swayed if the headline is interesting and catchy [3], [4], [5]. However, identifying bias solely based on the headline can pose a challenge. It is because they are typically short and may not contain the context of bias embedded in the story [6]. Furthermore, syntactic and semantic information are difficult to capture in a short headline. For example, consider the three headlines in Fig. 1 from our dataset MediaBias reporting on the same event with conflicting political ideologies. Unless we understand the context of the event and how the bias manifests itself, an automatic bias predictor will perform poorly [7].

To deal with this lack of context, models that use Inferential Commonsense Knowledge (IC_Knwl) have the opportunity to substantially enhance their performance. This enhancement is achieved in a manner nearly similar to the way people use commonsense inferences to carry out their daily routine tasks. For instance, the inclusion of IC_Knwl has been

The associate editor coordinating the review of this manuscript and approving it for publication was Arianna Dulizia.

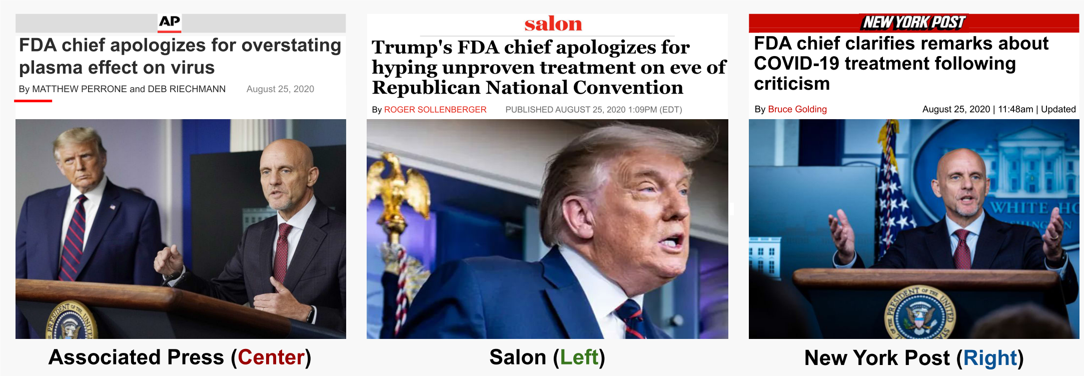


FIGURE 1. **News headlines reporting on the event "*FDA Commissioner Acknowledges Misrepresenting Convalescent Plasma Data*" from opposing political ideologies. (Image source: allsides.com).**

shown to improve the model's performance in a wide range of tasks such as emotion inference [8], sarcasm detection [9], sarcasm generation [10], and reading comprehension [11], among others [12].

Although IC_Knwl has been demonstrated to be promising, existing research neglects to acknowledge its potential in aiding text comprehension for the task of bias prediction. To this end, we propose leveraging it to simplify, interpret, and explain events that are not explicitly stated in headlines. We hypothesise that incorporating such knowledge would ideally help the learning model identify political bias when it is not readily evident from the headlines.

However, without proper emphasis, the additional inferential context is prone to introduce unnecessary noise. This noise can prevent models from fully exploiting the acquired knowledge. We thus present a method that facilitates emphasising significant inferences. We then use this knowledge to introduce a novel framework, **IC-BAIT** (**I**nferential **C**ommonsense aware **B**i**A**s **I**den**T**ifier). It leverages the IC_Knwl acquired using the neural knowledge model *COMET* [13] trained on *ATOMIC*$^{20}_{20}$ [13] knowledge graphs. For example, for the headline "Trump's awful advice on voting twice", the acquired commonsense knowledge provides an inferential context, as shown in Fig. 2.

Additionally, given the scarcity of large-scale datasets that adequately capture the unique challenges posed by the task of bias prediction in news headlines, we present two datasets, *MediaBias* and *GoodNews*, in Section III. We conduct experiments on both datasets to test the robustness of IC-BAIT. Furthermore, to assess its effectiveness, we compare the performance of several state-of-the-art models incorporated into it. We also conduct a comprehensive error analysis on its prediction results to determine when the selective inclusion of IC_Knwl is advantageous and when it is not. To summarise, we make the following contributions:

- Proposing to leverage Inferential Commonsense Knowledge to aid in the comprehension of news headlines by simplifying, interpreting, and explaining events that are not explicitly stated in the headlines.

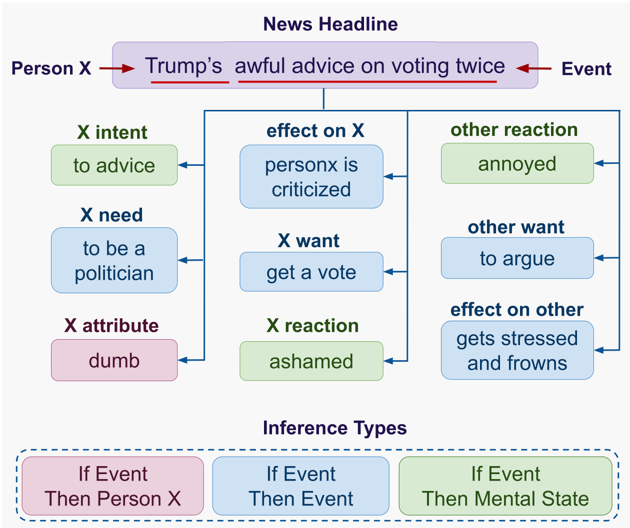


FIGURE 2. **An illustration of IC_Knwl acquired using COMET for a sample news headline. "PersonX is dumb, needed to be a politician, intended to advice, wanted to get a vote, is criticized, feels ashamed. Others want to argue, gets stressed and frowns, feel annoyed" represents the set of inferred causes and effects for the headline "Trump's awful advice on voting twice."**

- Introducing IC-BAIT, a neural network framework designed to enhance political bias prediction in news headlines by incorporating commonsense knowledge with an emphasis on important inferences.
- Presenting two datasets of news headlines annotated with political bias. We conduct experiments on both of them to demonstrate IC-BAIT's reliability.
- Evaluating the effectiveness of IC-BAIT with several state-of-the-art language models.
- Analysing the impact of selective knowledge augmentation on overall performance and conducting an in-depth error analysis.

We believe that our proposed framework could serve as an effective method for copy editors [14]. It could help them avoid any accidental or intentional bias in news headlines written by journalists. It could also be useful in practical applications such as e-journalism and manual news-bias prediction portals (ex: allsides.com, adfontesmedia.com), where it could be used to automatically classify headlines into different bias types. Additionally, it could help reduce the number of articles that require manual examination, which is a time-consuming process that is often susceptible to annotator bias [15]. Furthermore, it could be ideal for social scientists and those interested in the analysis of political bias, as well as for automated systems such as the "bias flipper" [6], [16], [17].

II. RELATED WORK

A news report is deemed politically biassed if it appears to favour one political ideology over another [18]. Such bias can appear in a variety of ways, such as in the selection of news stories, the tone and language of reporting, the sources of citation, etc [19]. Such bias in news reporting can result in an incomplete or inaccurate portrayal of events and issues, as well as the spread of misinformation or propaganda [20]. It also has the potential to influence public opinion on political issues and events [21]. Owing to their immense significance, bias prediction and mitigation have been the focus of research at multiple levels of granularity. In this section, we present a description of the datasets used in the related studies, followed by a discussion of the associated methodologies.

A. DATASETS

Many datasets have been designed to study political bias at different granularity levels: news sources [22], [23], articles [24], paragraphs [25], sentences [26], [27], and headlines [6] with one or more bias types [28].

Political bias annotation at the source level is typically obtained from online platforms, such as allsides.com and adfontesmedia.com. These platforms are dedicated to assessing and rating the bias of media sources to provide balanced news, perspectives, and issues across the political spectrum. They have highly qualified teams that rigorously follow predefined guidelines for rating.

At the sentence level, numerous datasets with manual bias annotations have been developed, with an emphasis on individual sentences of the articles [26], [29]. Manual annotations of individual sentences require substantial time and effort, and they do not scale well [30]. These annotations are even more time-consuming at the article level. As a result, many datasets with article-level annotations are generated by collecting articles from news sources with a known bias, where individual news articles reflect their source's political leanings [31], [32].

For example, authors in [33] used data from one such platform, Media Bias/Fact Check (MBFC), which contains manual annotations of political bias for over $2,000$ news websites. Their dataset includes $1,066$ websites with explicit bias labels ranging from 'extreme-left' to 'extreme-right' on a seven-point scale. Manual inspection of their dataset revealed that the 'left-center' and 'right-center' labels were ill-defined and ambiguous. Therefore, they opted to exclude news websites with these labels. They also merged the labels 'extreme-left' with 'left' and 'extreme-right' with 'right' to mitigate the impact of potentially subjective annotator decisions. As a result, their labels were reduced to a three-point scale (left, center, and right), resulting in a total of 864 websites [34]. However, the political leanings of news articles do not always correspond to the political leanings of the sources [35]. Besides that, the publisher's political leanings may shift while reporting on various topics [36].

Although the majority of these datasets are concerned with news articles or news sources, a few related datasets are focused on news headlines. Authors in [6], compiled a list of article pairs reporting on the same event from sources with opposing political views. They used it for the task of flipping the bias of news headlines. They later started with this dataset and extended it with new article pairs to analyse political bias and unfairness in news articles at different levels of granularity [7]. Event-centred pairing is considered necessary for tasks such as bias flipping and examining how different news outlets cover the same event. Any dataset devoid of such pairings, on the other hand, presents unique challenges for political bias prediction in news headlines and helps determine whether the prediction method is truly generic.

To deal with the scarcity of publicly available large-scale datasets for determining the political bias of news headlines, we present the datasets GoodNews III-B and Media-Bias III-A. GoodNews includes mapping from the corpus GoodNewsEveryone [37] and MediaBias consists of headlines paired as per the events, similar to [6]. We provide data generation frameworks to aid expansion and data customization for both datasets.

B. POLITICAL BIAS PREDICTION

In general, news bias is identified by studying linguistic attributes such as keywords and syntactic features. But, with the advent of machine learning and neural networks [38], [39], [40], [41], advanced algorithmic methods for bias analysis in news texts have been developed [42], [43]. For instance, leveraging recent advancements in deep learning, the authors in [42] developed a sentence-level factuality and bias prediction model fine-tuned on BERT [44]. Authors in [45] used the Gaussian Mixture Model, whereas authors in [43] used BERT and ELMo [46] to classify political bias in news articles, respectively.

There is another set of studies that aim to identify political bias based on explicitly defined features, such as quotes from think tanks. For example, authors in [18] studied news bias by estimating the ideological scores of major outlets. For the study, they made a list of influential policy groups and active think tanks and counted how many times a specific outlet and members of Congress used quotes from those groups.

Based on the pattern of citations, they estimated the bias score. However, articles might not cite a specific policy group but still exhibit political favouritism. For instance, one party may be disproportionately criticised for building up another party's figure. Several features are therefore required to be considered to determine political leaning.

The majority of related research seeks to strengthen methods for determining bias in news content as opposed to the headline. There are only a few methods concerning bias in headlines. For instance, authors in [47] used a headline attention network consisting of a headline encoder, an article encoder, and a headline attention layer to detect political bias in news content. They used headlines, articles, and the bias of the news to generate the dataset. They concluded that headlines alone weren't enough to figure out if an article was biassed or not. To analyse the bias induced at different granularity levels, authors in [6] extracted the most representative, discriminative, and sentiment-inducing words as the features of their classifiers. Their analysis revealed that named entities are very important to discriminate the left or right orientation of the text, whereas sentiment words play a crucial role in subjectivity identification.

Although deep learning is believed to be superior at learning complex and dense text representations through the capture of semantic and syntactic information, extending it from bias classification at the article level to the headline level is not plausible [6]. This is because headlines are more complicated and ambiguous than news articles, and they lack contextual and conceptual information.

To address this bottleneck in similar tasks, prior studies have focused on utilising external knowledge. For instance, Mihaylov and Frank [48] used key-value memory to represent commonsense facts and employ word-to-knowledge attention; Chen et al. [49] used semantic relations present in WordNet [50] to enhance attention and inference capabilities; Bauer et al. [51] proposed a mutual information-based knowledge selection method and integrated knowledge using gated attention; Zhang et al. [52] proposed using triplet-based knowledge to resolve coreferences.

Although these studies incorporate knowledge into pre-trained language models, they prioritise entity-centric facts in knowledge bases over commonsense, limiting their capacity to intuitively reason about events and situations. Empirical evidence suggests that the incorporation of commonsense knowledge is a valuable resource for language inference, leading to significant improvements in a variety of tasks [53], [54]. Consequently, recent research has sought to integrate external commonsense knowledge into pre-trained models to enhance linguistic representation for knowledge-reliant NLP tasks [55].

For generating such knowledge, commonsense knowledge graphs are commonly employed as standard tools to provide models with inferential background knowledge [56], [57]. ATOMIC is an example of such a knowledge graph, with an emphasis on event representation and enhanced relation types. As evidenced by its performance in comparison to human evaluation, it demonstrates competitive performance in the if-then reasoning task. ATOMIC$^{20}_{20}$ [13] which is an extension of ATOMIC [58] is another example. It incorporates additional event-centric relations and ConceptNet [59] facts that are not easily captured by language models, resulting in a comprehensive repository of complex entities.

The commonsense knowledge bases aid in training neural language models to generate commonsense descriptions [60]. In contrast to extractive methods, the models trained on commonsense knowledge bases exhibit a substantial advantage in their ability to generate knowledge about unseen events. This attribute is particularly essential for tasks that necessitate the use of commonsense [9], [12].

Our study diverges from prior research by demonstrating the potential of leveraging implicit knowledge acquired from pre-trained language models and explicit knowledge in the form of inferential commonsense knowledge at the same time. We show that they can be used in conjunction to enhance the task of bias prediction in complex news headlines. We present our framework that utilises the commonsense knowledge acquired using the neural knowledge model COMET [13] trained on ATOMIC$^{20}_{20}$ knowledge graphs. It employs an attentive knowledge selection strategy for emphasising significant inferences in the generated inferential statements. To ensure future adaptability, we design our framework to be model-agnostic, allowing it to be packaged with cutting-edge linguistic models as they become available.

III. DATASET DESCRIPTION

To address the lack of standardised, annotated datasets for the task of bias prediction in news headlines, we present two distinct datasets. We generate the first dataset, MediaBias (Section III-A), by utilising bias labels from the bias rating portal allsides.com. We generate the second dataset, GoodNews (Section III-B), by extracting the headlines and their bias values from the dataset GoodNewsEveryone [37], which is based on the bias rating portal adfontesmedia.com.

In Table 1, we summarise dataset statistics, including the number of headlines for each political ideology, the average number of words in the headline, and so on. We document our datasets following the requirements of the FAIR Data Principles.[1]

A. MediaBias

1) RAW DATA SOURCE: ALLSIDES

To generate the dataset MediaBias, we use the bias rating portal allsides as the raw data source. The team at *allsides* consists of news experts with political leanings from all sides of the political spectrum. Every week, they identify the most important news stories and write about twenty roundups. They cover stories ranging from major national events to niche opinions. Their goal is to cover the most important news stories rather than the most clickable, viral, or interesting ones. For each roundup, they scan over 800 media outlets

[1] https://www.nature.com/articles/sdata201618/

TABLE 1. Dataset Statistics. Avg. length: average number of words in the headline.

			Size of the split				Bias label		
Dataset	**Raw data source**	**Dataset size**	**Train**	**Valid**	**Test**	**Avg. length**	**Left**	**Center**	**Right**
MediaBias	allsides	11,031	8,825	1,102	1,104	13	3,082	4,221	3,728
GoodNews	adfontesmedia	3,058	2,446	306	306	13	877	1,153	1,028

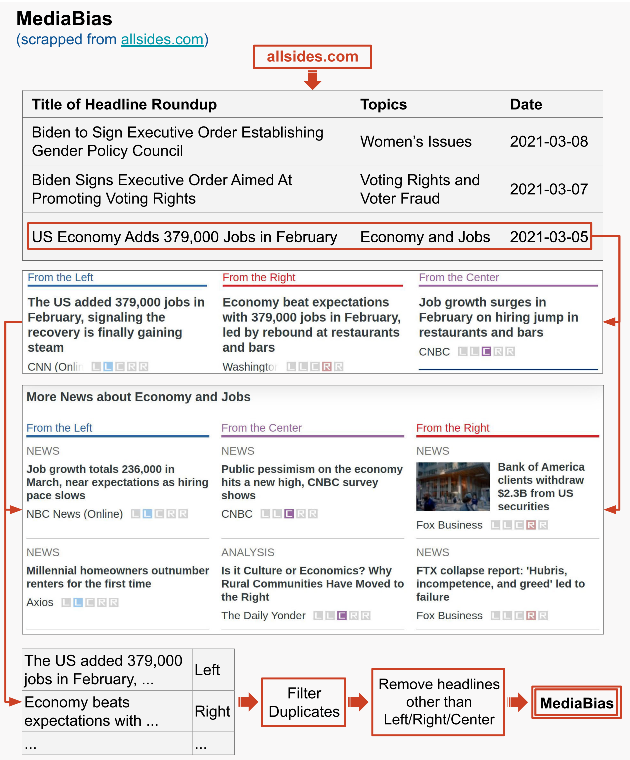


FIGURE 3. Overview of MediaBias generation process.

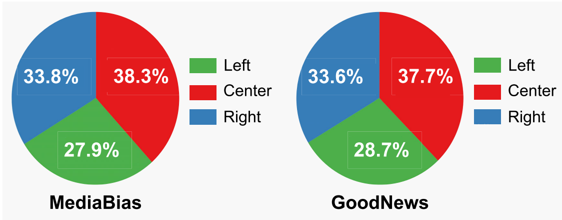


FIGURE 4. Bias label distribution. The graphs demonstrate the imbalance in both datasets.

and release bias ratings for three sources that have published articles about it from different political perspectives. They then compile three main headlines from the *left*, *center*, and *right* sides of the same story and present them side by side to compare news coverage, identify bias, and get the entire story. Since it provides annotations for individual stories based on a rigorous process [61], we use it as a source of high-quality gold standard data for our dataset generation process.

2) MEDIABIAS GENERATION PROCESS

We generate the dataset *MediaBias* by using the existing bias labels from the featured headline roundups of the news aggregator *allsides*. We begin by crawling through all of the featured stories, regardless of their category. For each story, we then crawl through the three major stories and all the stories related to their topic from different political perspectives. Following that, we record the headlines of each story along with their bias values. We remove duplicate headlines as well as those with labels *Lean Left*, *Lean Right*, *Not Rated*, and *Mixed*. The generation process is demonstrated in Fig. 3.

We end up with 11, 031 headlines with an average length of 13 words. Since the bias label distribution is not uniform, as shown in Fig. 4, we use a stratified split to replicate this imbalance across the generated train-valid-test sets.

B. GoodNews

1) RAW DATA SOURCE: ADFONTESMEDIA

We use the bias rating portal Adfontesmedia as the raw data source to generate the dataset, GoodNews. It scores news outlets by using bias and reliability as coordinates on its chart [62]. It assigns ratings to articles based on the opinions of a panel of expert analysts. Before labelling an outlet, the panels read at least three to thirty articles. They calculate the reliability score by adding individual ratings for an article's correctness, use of fact or opinion, and appropriateness of its headline and graphic content. The political bias score is influenced by topic selection or omission, the language used in the article, and the extent to which a political stance from left to right is endorsed. Individual ratings for each reviewed article are then aggregated to calculate an outlet's overall bias and reliability score, with popular articles receiving more weight. This average determines the placement of the outlet on the chart. Markers in the chart range from *Most Extreme Left* to *Most Extreme Right* along the bias axis.

2) GoodNews GENERATION PROCESS

We generate the dataset GoodNews by extracting the headlines and their bias values from the corpus GoodNewsEveryone [37] which is based on adfontesmedia. To do so, we first crawl through all the headlines in the corpus. For each headline, we then extract the headline text and its associated bias values. Following the range of bias values defined by

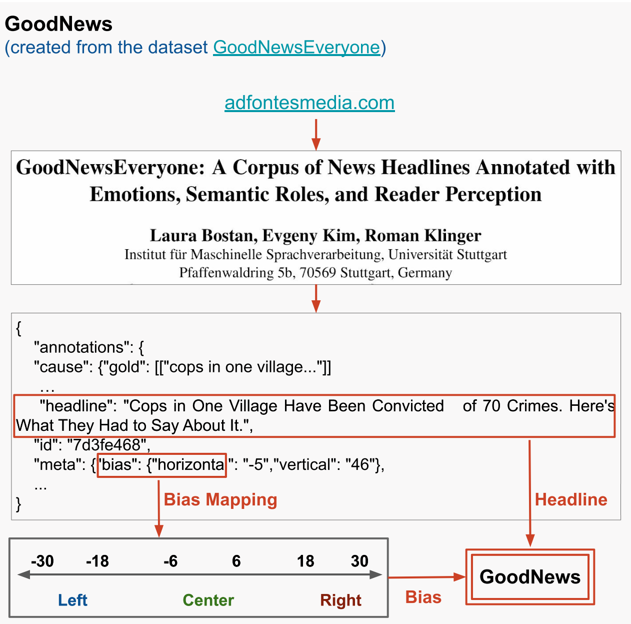


FIGURE 5. Overview of GoodNews generation process.

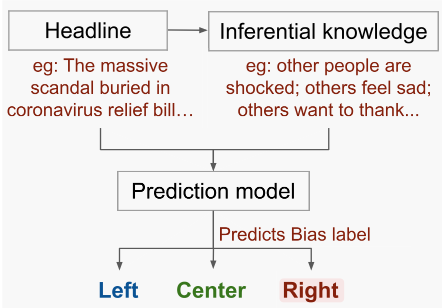


FIGURE 6. Outline of the task of predicting political bias in news headlines.

adfontesmedia.com, we then map the horizontal bias values from -30 to -18 to the *left*, -6 to 6 to the *center*, and 18 to 30 to the *right*. For mapping, we use the horizontal axis, which represents the political orientation of the news. Fig. 5 depicts the dataset generation process.

We end up with 3,058 headlines with an average length of 13 words. Similar to MediaBias, we use a stratified split to mimic the imbalance in the resultant train-valid-test sets, as illustrated in Fig. 4.

IV. MATERIALS AND METHODS

In this section, we first formulate the task of predicting political bias in news headlines. We then present a brief discussion of the baseline models that we use to evaluate IC-BAIT. We then introduce our proposed framework and its key components. Finally, we discuss the metrics that we use for evaluation.

A. PROBLEM FORMULATION

For a given headline text H, the task begins with the acquisition of its associated inferential commonsense knowledge IC_Knwl,

$$IC\_Knwl = c(H, \alpha) \tag{1}$$

Here, c is the commonsense knowledge modelling function, and α denotes the model parameters. Given the pair (H, IC_Knwl), our final task is to train a classifier that maps this extended feature space of short texts into the political bias label set B. Mathematically, it can be formulated as,

$$b = f(H, IC\_Knwl, \theta) \tag{2}$$

where f is the bias prediction function, θ denotes the model parameters, and b indicates the bias labels in the set B. Fig. 6 depicts an outline of the formulated problem.

B. BASELINE MODELS

We examine the following state-of-the-art language models in IC-BAIT for a comprehensive evaluation:

- **ALBERT (A Lite BERT)** [63]: a lite BERT [44] for self-supervised language representation learning. It uses a self-supervised loss that models inter-sentence coherence and the parameter reduction technique to decrease memory usage and accelerate training. Pre-trained model: https://tfhub.dev/google/albert_base/
- **DistilRoBERTa (Distilled Robustly optimized BERT approach)** [64]: a knowledge-distilled version of the robustly optimised BERT-based model, RoBERTa [65]. It significantly outperforms BERT as its model is trained longer with larger batches on longer sequences over more data, the objective of predicting the next sentence is removed, and the masking pattern on training data is modified dynamically.
- **MPNet (Masked and Permuted Net)** [66]: a masked and permuted method that inherits the benefits of BERT and XLNet [67] without suffering their drawbacks. It employs permuted language modelling to take advantage of the dependency among predicted tokens and takes auxiliary position information to present the model with a complete sentence, thereby reducing position discrepancy. Pre-trained model: https://github.com/microsoft/MPNet
- **MiniLMv2 (Mini Language Model v2)** [68]: a generalised and simplified deep self-attention distillation in MiniLM [69]. It introduces multi-head self-attention relation distillation to provide students with a flexible number of attention heads while also improving fine-grained self-attention knowledge. Distilled model: https://aka.ms/minilm
- **CMLM (Conditional Masked Language Modeling)** [70]: a conditional masked language model built on a 12-layer BERT architecture. It combines sentence representation learning with MLM training by using encoded vectors of adjacent

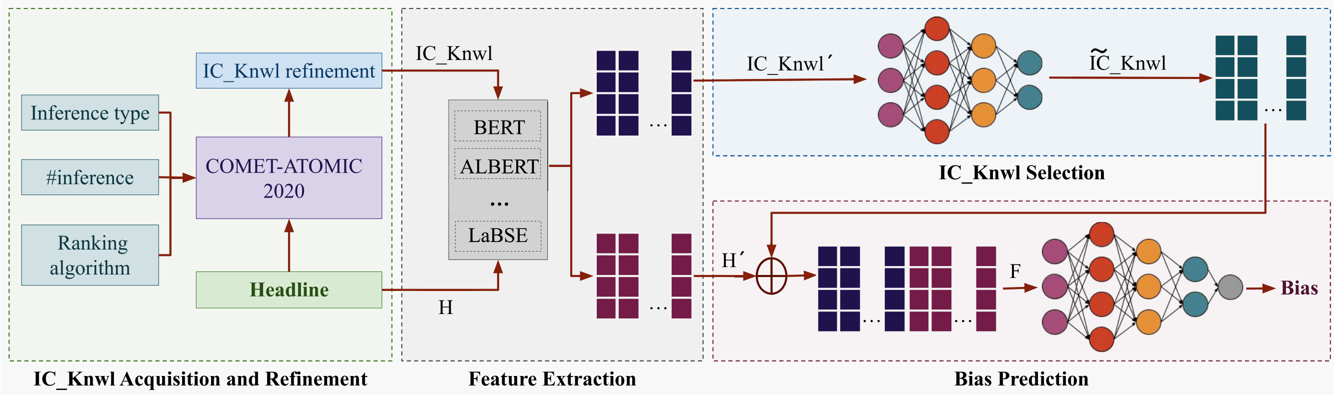


FIGURE 7. **Abstract representation of the bias prediction framework.**

sentences as a conditioning factor. Pre-trained model: https://tfhub.dev/google/universal-sentence-encoder-cmlm/en-base/

- **GPL (Generative Pseudo Labelling)** [71]: a method for training dense retrievers using unsupervised domain adaptation. It relies on pseudo-labelling in conjunction with robust cross-encoders and query generators. Pre-trained model: https://huggingface.co/GPL
- **GPL+BPR (Generative Pseudo Labelling + Binary Passage Retriever)** [72]: a method that optimises BPR [73] using GPL. It achieves memory efficiency by employing compact binary codes instead of continuous vectors, and domain adaptation efficiency without requiring domain-specific annotated training data. Model: https://github.com/NThakur20/income
- **LaBSE (Language-agnostic BERT Sentence Embedding)** [74]: a BERT-based, language-agnostic model. It employs a novel combination of masked language modelling, translation language modelling, dual encoder translation ranking, and additive margin softmax to establish a new benchmark for bi-text mining. Pre-trained model: https://tfhub.dev/google/LaBSE

C. METHODOLOGY

Our proposed framework (IC-BAIT), as illustrated in Fig. 7, is built around the following key components: IC_Knwl acquisition and refinement, feature extraction, IC_Knwl selection, and bias prediction. We sketch the pseudocode for IC-BAIT in Algorithm 1 and present a detailed description of each component in the following subsections.

1) IC_Knwl ACQUISITION AND REFINEMENT

IC_Knwl assists in simplifying, comprehending, and explaining events that are not explicitly stated in the headline. To acquire it for each headline H, we use the neural knowledge model *COMET* (COMmonsensE Transformers)[2] [13] trained on the $ATOMIC_{20}^{20}$ (ATlas Of MachIne Commonsense-2020)[3] [13] knowledge graphs:

$$IC\_Knwl = COMET(H, I_{type}, r, k) \tag{3}$$

where I_{type}, r, and k denote the inference type, ranking algorithm, and the number of returned references respectively. We set I_{type} = {*oEffect*, *oReact*, *oWant*, *xAttr*, *xEffect*, *xIntent*, *xNeed*, *xReact*, and *xWant*}, r = *beam*, k = 3. Fig. 8 defines the inference types.

For each I_{type} inference in *IC_Knwl*, we first determine whether or not the returned list is empty. We then assign *'none'* to *IC_Knwl* if the list is empty or if all of the entries are *'blank'* or *'garbage'*. If the list isn't empty and at least one

---

Algorithm 1 bias prediction algorithm.

Input: headline text H, true bias label b, inference type I_{type}, ranking algorithm r, number of references k
Output: predicted bias label $\hat{b}$

```
while not converge do
    for each headline H do
        extract IC_Knwl using (H, I_type, r, k)
                                            ▷ ref. Eq. 3
        refine IC_Knwl
                                            ▷ ref. Section 3
        extract feature vectors H′ and IC_Knwl′
                                            ▷ ref. Section IV-C2
        filter and select relevant knowledge ĨC_Knwl
                                            ▷ ref. Eq. 4
        generate fused feature vector F using H′ and ĨC_Knwl
                                            ▷ ref Eq. 5
        get predicted label b̂
                                            ▷ ref Eq. 6
        calculate prediction Loss
                                            ▷ ref. Eq. 7
        backpropagate Loss to update IC-BAIT
    end for
end while
```

[2]https://github.com/allenai/comet-atomic-2020/

[3]https://allenai.org/data/atomic-2020

FIGURE 8. Interpretation of ATOMIC$^{20}_{20}$ inference types.

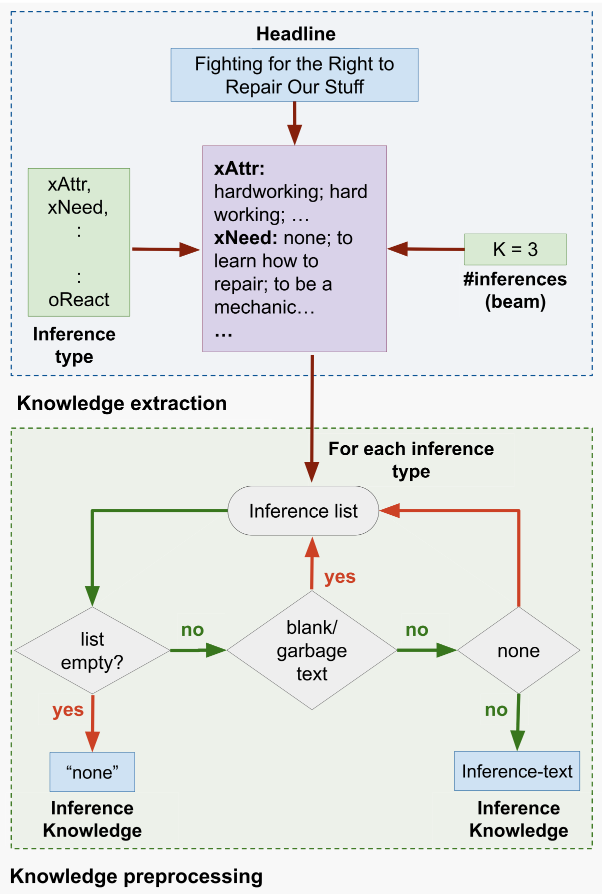


FIGURE 9. Flowchart of inferential knowledge acquisition process.

item isn't *blank* or *garbage*, we add the first such item to the *IC_Knwl* list. Fig. 9 illustrates the flowchart of the IC_Knwl acquisition process.

To render *IC_Knwl* more meaningful, we combine the retrieved inferences into a statement describing the properties of the person involved, for example,

1) **Headline:** Fighting for the Right to Repair Our Stuff
2) **Raw IC_Knwl:** xAttr: *hardworking*, xNeed: *to learn how to repair*, xIntent: *to be able to repair*, xEffect: *fighting for the right to repair*, xWant: *to make sure it's fixed*, xReact: *proud*, oWant: *to help others*, oEffect: *fight for rights*, oReact: *happy*
3) **Processed IC_Knwl:** PersonX is *hardworking*, needed *to learn how to repair*, intended *to be able to repair*, *fighting for the right to repair*, *to make sure it's fixed*, feels *proud*. Others want *to help others*, *fight for rights*, feel *happy*.

2) FEATURE EXTRACTION

To acquire the feature vectors H' and IC_Knwl' from the headline (H) and Inferential Commonsense Knowledge (IC_Knwl), we use the state-of-the-art pre-trained language models that we describe in Section IV-B. IC-BAIT is model-agnostic, and thus models of any depth can be easily integrated.

3) KNOWLEDGE SELECTION

Using only one of the inference types to predict the bias may not be ideal. For instance, if we rely exclusively on one of the inference types, such as Person X's reaction, the resulting inferential knowledge may aid in sentiment prediction but not in comprehending the context of the event for bias prediction. As a result, we use all of the inference types for our task. However, not all of them are equally important. Consequently, we utilise a method that resembles an established and efficacious approach to selecting relevant knowledge [12], [75].

To filter the relevant knowledge, we first apply the Sigmoid function [76] to IC_Knwl', to measure the importance of each inference. Specifically, we employ it as a cumulative density function of the logistic distribution to assign probabilities to the inferences [77]. We then multiply IC_Knwl' by the resulting importance scores and feed the resultant vector to a Multi-Layer Perceptron (MLP) network that learns to mix the inferences of different I_{type} to generate $\widetilde{IC}\_Knwl$:

$$\widetilde{IC}\_Knwl = MLP(sigmoid(IC\_Knwl') \odot IC\_Knwl') \quad (4)$$

where $\odot$ denotes element-wise multiplication.

4) BIAS PREDICTION

For the task of bias prediction, we first generate the fused vector F by concatenating H' and $\widetilde{IC}\_Knwl$:

$$F = H' \oplus \widetilde{IC}\_Knwl \quad (5)$$

where $\oplus$ represents the concatenation operation. We then feed F to the MLP network. To finally predict the bias label $\hat{b}$, we pass the resultant vector to the Fully Connected layer (FC) having Softmax (σ) activation with three output neurons that represent bias labels in B.

$$\hat{b} = FC(\sigma(MLP(F))) \tag{6}$$

For the training process, we use categorical cross-entropy as the loss function:

$$Loss = -\sum_{i=1}^{|B|}(b_i * \log(\hat{b}_i)) \tag{7}$$

where b_i and $\hat{b}_i$ denotes the actual and predicted probability of selecting the i^{th} bias label in B. We release the detailed experimental settings at https://github.com/Swati17293/IC-BAIT

D. EVALUATION METRICS

We use the following well-known metrics [78] for performance evaluation:

- **Accuracy (Acc.)** measures the proportion of correct predictions and overall predictions.
- **F$_1$-score** is the weighted average of Precision (P) and Recall (R) calculated as:

$$\text{F}_1\text{-score} = 2 * ((P * R)/(P + R)) \tag{8}$$

  where P denotes the proportion of correctly predicted instances of a bias category, say (c), to the total number of predicted instances of that category, and R denotes the proportion of correctly predicted instances out of the total number of actual instances that fall under the category c.
- **Jaccard-score (J)** [79] computes the fraction of correctly predicted instances over all instances, excluding true negatives. It disregards true negatives in favour of true positives. For an imbalanced dataset like the ones used in this study, where true negatives outnumber true positives, it helps to gain a deeper understanding of the results.

We report a macro-averaged F$_1$-score ($\mathbf{F_1}$) to ensure that all bias classes are treated equally. In addition, we report a micro-averaged F$_1$-score ($\mathbf{F_{1\mu}}$) and a micro-averaged Jaccard-score ($\mathbf{J_\mu}$) to account for the problem of class imbalance.

V. RESULTS AND ANALYSIS

We begin the section by presenting the experimental results. We then thoroughly examine the experimental results to study the impact of IC-BAIT on the task of bias prediction. We conclude with a comprehensive analysis to determine the scenarios under which IC-BAIT is beneficial and when it is counterproductive.

A. EXPERIMENTAL RESULTS

Table 2 shows the overall performance distribution of the baseline models and our framework. For each metric, it reports the average score across the models. For comparison, we use Fig. 10a and Fig. 10b to illustrate the distributional characteristics of overall performance scores for all reported metrics across the MediaBias and GoodNews datasets. We use boxplots for visualisation as they are reliable and can reveal a lot of statistical information, such as ranges, outliers, and medians. For instance, the line positioned in the middle of the box represents the median, with scores above it indicating the top 50% of results. Similarly, the line at the top part of the box signifies the third quartile (Q3) with scores above it indicating the top 25% of results. From the plots, it can be clearly seen that our framework has superior performance for each metric on both datasets, with a comparatively higher median and Q3.

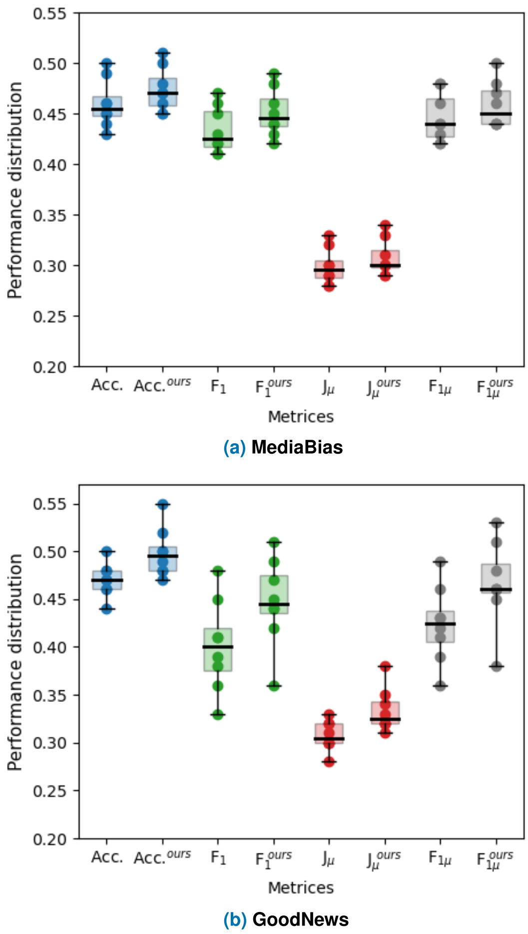


FIGURE 10. Boxplots of overall performance distributions for the datasets (a) MediaBias and (b) GoodNews, where the presence of a median near the third quartile (Q3) and a relatively high Q3 indicates superior performance.

In general, a model performs better when there are more training samples. However, interestingly, the results reveal that the models perform better on the smaller dataset (GoodNews) than on the larger one (MediaBias). The reason for this can be attributed to the dataset's characteristics. Whereas a team of experts selected the headlines in the MediaBias dataset to present the news stories from different political perspectives, the GoodNews dataset was assembled to classify emotions, which may

TABLE 2. Overall performance distribution of the baseline models and our proposed framework. Percentage improvement, denoted by %↑, indicates that our framework significantly improves prediction performance on both datasets.

Dataset	Acc.	Acc.^{ours}	%↑	F_1	$F_1{}^{ours}$	%↑	J_μ	J_μ^{ours}	%↑	$F1_\mu$	$F1_\mu^{ours}$	%↑
MediaBias	0.46	0.47	**2.2**	0.43	0.45	**4.6**	0.30	0.31	**3.3**	0.45	0.46	**2.2**
GoodNews	0.47	0.50	**6.4**	0.40	0.45	**12.5**	0.31	0.33	**6.4**	0.42	0.47	**11.9**

TABLE 3. Detailed performance distribution of the baseline models and our proposed framework for the datasets (a) MediaBias and (b) GoodNews, with %↑ representing the percentage improvement.

(a) MediaBias

	Acc.	Acc.^{ours}	%↑	F_1	$F_1{}^{ours}$	%↑	J_μ	J_μ^{ours}	%↑	$F1_\mu$	$F1_\mu^{ours}$	%↑
ALBERT [63]	0.43	0.45	**4.6**	0.41	0.44	**7.3**	0.28	0.29	**3.6**	0.42	0.44	**4.8**
DistilRoBERTa [64]	0.45	0.46	**2.2**	0.42	0.43	**2.4**	0.29	0.30	**3.4**	0.44	0.44	0.0
MPNet [66]	0.46	0.48	**4.3**	0.45	0.46	**2.2**	0.30	0.31	**3.3**	0.46	0.47	**2.2**
MiniLMv2 [68]	0.44	0.45	**2.3**	0.41	0.42	**2.4**	0.28	0.29	**3.6**	0.42	0.44	**4.8**
CMLM [70]	0.50	0.51	**2.0**	0.47	0.49	**4.2**	0.33	0.34	**3.0**	0.48	0.50	**4.2**
GPL [71]	0.45	0.47	**4.4**	0.43	0.44	**2.3**	0.29	0.30	**3.4**	0.44	0.44	0.0
GPL+BPR [72]	0.46	0.47	**2.2**	0.42	0.45	**7.1**	0.30	0.30	0.0	0.43	0.46	**7.0**
LaBSE [74]	0.49	0.50	**2.0**	0.46	0.48	**4.3**	0.32	0.33	**3.1**	0.48	0.48	0.0

(b) GoodNews

	Acc.	Acc.^{ours}	%↑	F_1	$F_1{}^{ours}$	%↑	J_μ	J_μ^{ours}	%↑	$F1_\mu$	$F1_\mu^{ours}$	%↑
ALBERT [63]	0.46	0.48	**4.3**	0.45	0.47	**4.4**	0.30	0.32	**6.7**	0.46	0.48	**4.3**
DistilRoBERTa [64]	0.47	0.48	**2.1**	0.39	0.45	**15.4**	0.30	0.32	**6.7**	0.42	0.46	**9.5**
MPNet [66]	0.48	0.52	**8.3**	0.41	0.49	**19.5**	0.32	0.35	**9.4**	0.43	0.51	**18.6**
MiniLMv2 [68]	0.46	0.49	**6.5**	0.41	0.42	**2.4**	0.30	0.32	**6.7**	0.43	0.45	**4.6**
CMLM [70]	0.48	0.50	**4.2**	0.38	0.44	**15.8**	0.32	0.34	**6.2**	0.41	0.46	**12.2**
GPL [71]	0.50	0.55	**10.0**	0.48	0.51	**6.2**	0.33	0.38	**15.1**	0.49	0.53	**8.2**
GPL+BPR [72]	0.44	0.47	**6.8**	0.33	0.36	**9.1**	0.28	0.31	**10.7**	0.36	0.38	**5.5**
LaBSE [74]	0.47	0.50	**6.0**	0.36	0.44	**22.2**	0.31	0.33	**6.4**	0.39	0.46	**17.9**

have contributed to improved performance. The fact that MediaBias is event-centric whereas GoodNews is not could also be a contributing factor.

To aid clarity and comprehension of the findings, Table 3 presents a detailed breakdown of the scores for each model. In a nutshell, IC-BAIT improves the baseline models in terms of Acc. (2.0-10.0%), F_1 (2.2-22.2%), J_μ (up to 15.1%), and $F1_\mu$ (up to 18.6%). The results reveal that CMLM for MediaBias and GPL for GoodNews stand out as the best-performing models, with impressive percentage improvements across all metrics. In a nutshell, our findings indicate that the integration of baseline models within our framework has a notable impact on prediction performance.

B. CASE STUDY

We investigate the prediction results of IC-BAIT (with and without IC_Knwl) for Left, Right, and Center oriented headlines and present a representative example from each set in Table 4.

In the study, it can be observed that IC-BAIT without IC_Knwl appears to fail in the majority of cases where the entities and/or terms mentioned in the headline are strongly correlated to an ideology other than that of the headline. In these cases, the model is more likely to learn the unjust correlation and predict it incorrectly. However, by incorporating IC_Knwl, IC-BAIT improves its understanding of the inferred meaning and makes predictions guided by that. As a result, it gains the ability to focus not only on the important entities and events mentioned in the headline but also on the explanation of unstated events for better comprehension.

1) IMPACT of IC_Knwl

To better understand the impact of IC_Knwl, it is necessary to investigate whether or not it guides the model to take a specific decision. For that, we use the LIME (Local Interpretable Model-Agnostic Explanations) [80] xAI framework, a popular qualitative interpretation tool. It provides a set of feature (word) weights as well as colour-coded text to aid in the explanation of a model prediction.

We first use the headline as the framework's sole input, and then we separate the misclassified instances, for which we

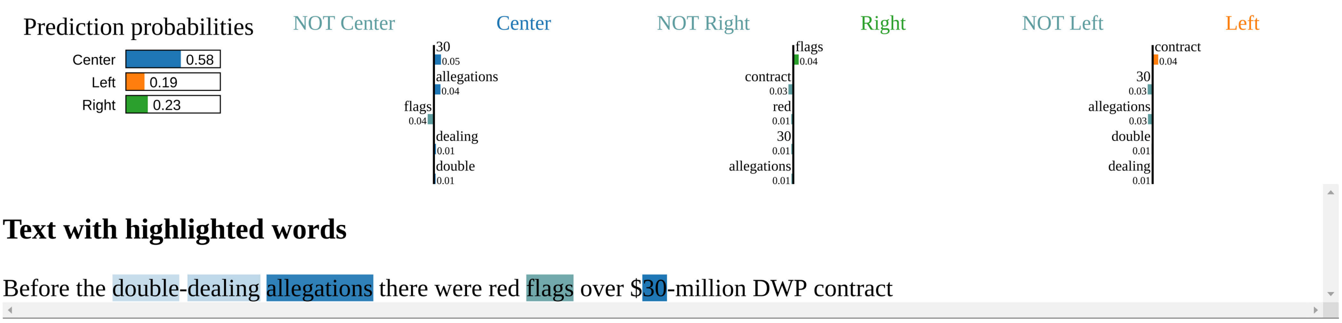
TABLE 4. Case study of bias label predictions by IC-BAIT (with and without IC_Knwl) for Left, Right, and Center oriented headlines.

Headline	**Donald Trump gets ripped to pieces over LGBT tweet**
IC_Knwl	PersonX is intolerant, needed to write a homophobic tweet, intended to be liked, gets yelled at, wants to make amends, feels upset. Others want to defend themselves, gets hurt, feel angry.
Bias Label	True: **Left**, Predicted(without IC_Knwl): **Right**, Predicted(with IC_Knwl): **Left**
Comment	Due to data bias, news related to the named entity "Trump" is heavily skewed to Right-wing media. Furthermore, idioms such as "ripped to pieces" are typically associated with the Right ideology. The model without IC_Knwl tends to learn this unjust correlation and thus ends up predicting it as "Right". However, with the additional commonsense inference, important information such as "personX is seen as intolerant" and "Others get hurt" was passed to the model, allowing it to learn the prediction correctly.
Headline	**Time to Kick the Islamizing Turkey Out of NATO**
IC_Knwl	PersonX is aggressive, needed to be a member of NATO, intended to get rid of terrorism, gets yelled at, wants to to get rid of the Islamists, feels angry. Others want to fight back, gets hurt, feel angry.
Bias Label	True: **Right**, Predicted(without IC_Knwl): **Left**, Predicted(with IC_Knwl): **Right**
Comment	An incorrect correlation between "Islam" and Left-wing media in the collected data causes the model without IC_Knwl to incorrectly predict the label as "Left". However, the acquired commonsense inferences such as, "personX is seen as aggressive and gets yelled at" provide a critical understanding of the statement, allowing the model with IC_Knwl to correctly learn the prediction.
Headline	**The refugee families caught up in a war zone in Libya**
IC_Knwl	PersonX is traumatized, needed to be deployed, intended to escape from war, gets lost in war, wants to find a new home, feels scared. Others want to get out of there, have to find a new home, feel sad.
Bias Label	True: **Center**, Predicted(without IC_Knwl): **Left**, Predicted(with IC_Knwl): **Center**
Comment	The terms "refugee" and "caught" in the headline indicate that the headline is right-oriented. However, with the given commonsense inference, such as "personX wants to find a new home" the model selects the Neutral or Central narration of the headline.

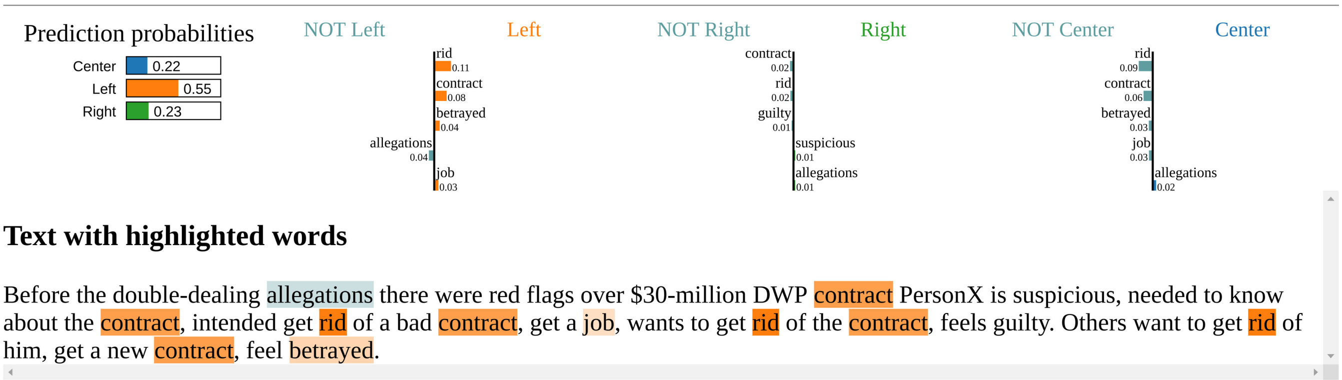
generate explanations to understand the probable reasons for misclassification. We then include IC_Knwl in the framework with the headline and regenerate the explanations for the same instances to see its impact on the prediction. Fig. 11 and Fig. 12 illustrate results for the top five significant features.

We follow the same procedure for the rest of the misclassified instances and conduct an in-depth analysis to determine the cases in which IC_Knwl is useful, and we reach the following conclusions:

- **Short headlines**: they lack the necessary contextual information to make an accurate prediction.
  Example: the short headlines, such as "*Brickbat: It's a Gas Gas Gas*" and "*Grit Won*", are insufficiently descriptive.
- **Entity**: whereas entities mentioned in headlines carry considerable weight, in the absence of contextual information, they quite often confuse the model.
  Example: although the entity "*FDA chief*" is the main subject of the headlines from different political ideologies, it has no bearing on definite ideology (ref. Table 5).
- **Metaphor**: while the words used in metaphors are significantly important, their interpretation may or may not be symbolic depending on their context. Typically, their literal and symbolic meanings are quite dissimilar, which complicates prediction.
  Example: consider the following two headlines: "*Trump: Al-Baghdadi Died Like a Dog*" and "*Ari Fleischer Lied and People Died*". The word "*died*" in the metaphor "*died like a dog*" refers to "*died in a painful and humiliating manner*" and thus carries a negative emotional pull. The second headline, on the other hand, uses the same word "*died*", implying an emotion such as grief.
- **Domain-specific/slang words**: these words are typically unimportant and provide little or no context to other significant entities (if present), and may lead to misclassification.
  Example: in the headline "*Journos Hit Hick for Pot Flip-Flop*", "*journos*" is slang for "*Journalists*" which is difficult for a model to comprehend. "*flip-flop*" in turn is a domain-specific word (political news) that refers to "*a sudden reversal of proposals*" in contrast to its general definition "*a type of footwear*".
- **Satire**: it appears to be a straightforward statement, but it contains humorous criticism.
  Example: the headline "*America's Racist Legacy From Slavery to the War on Immigrants*" is a criticism of history being whitewashed.



(a) Prediction framework *without* IC_Knwl, True: *Left*, Predicted: *Center*



(b) Prediction framework *with* IC_Knwl, True: *Left*, Predicted: *Left*

FIGURE 11. LIME visualization result: words like "*allegations*" in the headline 11a strongly indicates that the headline is center-oriented. However, the addition of words such as "*rid*" and "*betrayed*" from IC_Knwl results in a final prediction score of *0.55*, making the model 11b quite confident in its prediction.

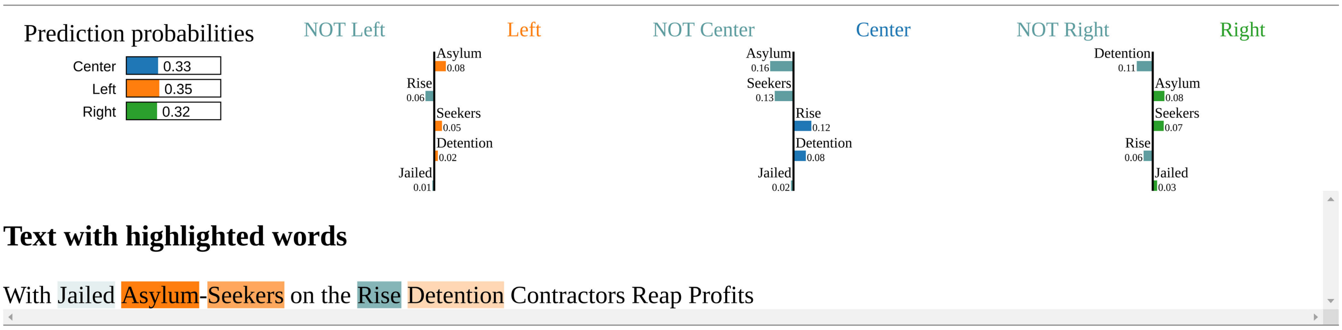
- **Sarcasm**: sarcastic headlines imply the inverse of what is stated plainly. When used sarcastically, words with a larger weight mislead the model.
  Example: the headline "*Trump: Illegal immigrants in sanctuary cities would make the 'Radical Left' happy*" refers to Trump's sarcastic remark about Left-wing supporters. Words like "*Trump*", "*Radical Left*", and "*happy*" convey a positive emotion rather than negative taunting.

Additionally, we identified instances where including IC_Knwl adds little or no value. For example,

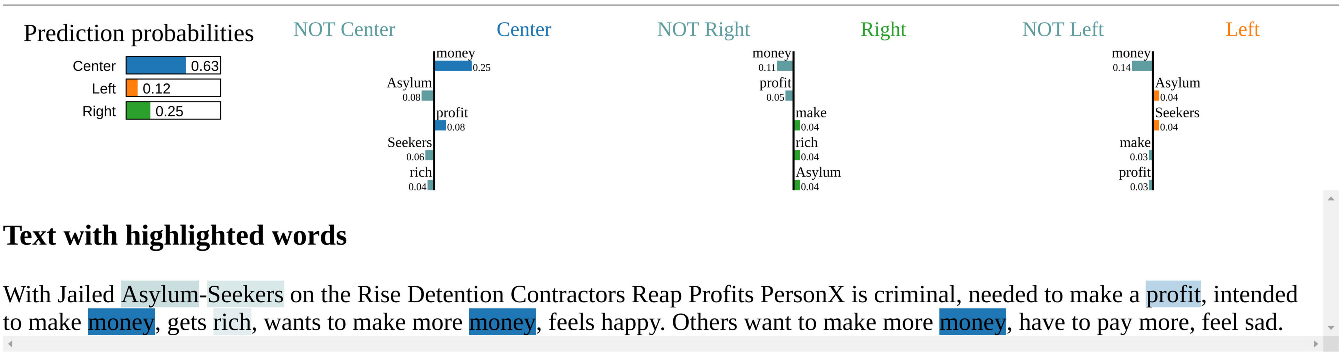
- **When IC_Knwl is comprised of generic information.** There are numerous instances where IC_Knwl returns nearly identical information for different headlines. Consider the following headlines and their IC_Knwl as an example:
  - **Headline**: Bernie Sanders Wins Indiana Democratic Primary. (Bias Label: Left)
    ICKnwl: *PersonX is hopeful, needed to run for office, intended to win the election, wins the election*, wants to get elected, *feels happy*. Others want to congratulate him, *wins the election, feel happy*.
  - **Headline**: Biden Sweeps Dem Primaries in Arizona Florida Illinois. (Bias Label: Right)
    ICKnwl: *PersonX is hopeful, needed to run for office, intended to win the election, wins the election*, wants to win the election, *feels happy*. Others want to win the election, *wins election, feel happy*.
- **When IC_Knwl is devoid of significance.** There are headlines where the inferences drawn by IC_Knwl are insignificant (ref. Fig. 13).
- **When the world knowledge overpowers IC_Knwl.** There are instances where world knowledge is more important than IC_Knwl for the ideology prediction (ref. Fig. 14).

VI. LIMITATIONS AND FUTURE DIRECTIONS

The limitations of our framework originate primarily from its key components. To begin with, including IC_Knwl renders little or no significant improvement when it lacks any useful information (ref. Section V-B1). For instance, when it is comprised of generic information, is insignificant, or when world knowledge overpowers it. Similarly, there are instances where IC_Knwl alone is sufficient for accurate prediction. Identifying when to include and exclude IC_Knwl precisely and quantitatively continues to pose a challenge. Another limitation applies to its reliability. Augmenting explicable justifications for its outcomes would greatly enhance its reliability.



(a) Prediction model *without* IC_Knwl, True: *Center*, Predicted: *Left*



(b) Prediction model *with* IC_Knwl, True: *Center*, Predicted: *Center*

FIGURE 12. LIME visualization result: the headline contains too many important words representing different ideologies, which confuses the model 12a. Using IC_Knwl, the model 12b, on the other hand, correctly predicts the ideology.

TABLE 5. IC_Knwl acquired from headlines of varying political ideologies reporting on the same event. The inclusion of IC_knwl facilitates prediction.

	Center	**Left**	**Right**
Headline	FDA chief apologizes for overstating plasma effect on virus	Trump's FDA chief apologizes for hyping unproven treatment on eve of Republican National Convention	FDA chief clarifies remarks about COVID-19 treatment following criticism
xAttr	careless	irresponsible	helpful
xNeed	to be in charge	to be honest	to make a statement
xIntent	to be honest	to be a better representative	to clear up confusion
xEffect	gets reprimanded	has to apologize	gets asked for clarification
xWant	to make amends	to apologize to the public	to make sure everyone knows the facts
xReact	sorry	ashamed	relieved
oWant	to ask for an explanation	to ask for an apology	to thank the FDA
oEffect	they get sick	gets fired from job	they get better
oReact	angry	scared	informed

Furthermore, given that our framework is model-independent, it can be used to predict bias in headlines that are not in English. However, this scenario is only possible if the language model can process multiple languages and the extracted IC_Knwl is in the target language. For instance, the multilingual language model CMLM [70] has the potential to serve as a multilingual language model. It is a fully unsupervised learning model that can readily be extended to a wide range of languages. For knowledge extraction, MultiCOMET [81] is a viable option. However, when used in this setting, the quality of knowledge it produces will directly influence the framework's performance.

In addition, advanced methods for knowledge selection and its subsequent integration into the framework remain unexplored, despite their potential to significantly improve the framework's performance. Also, the large-scale language models used in the study require considerable resources. Consequently, there remains a need to investigate strategies

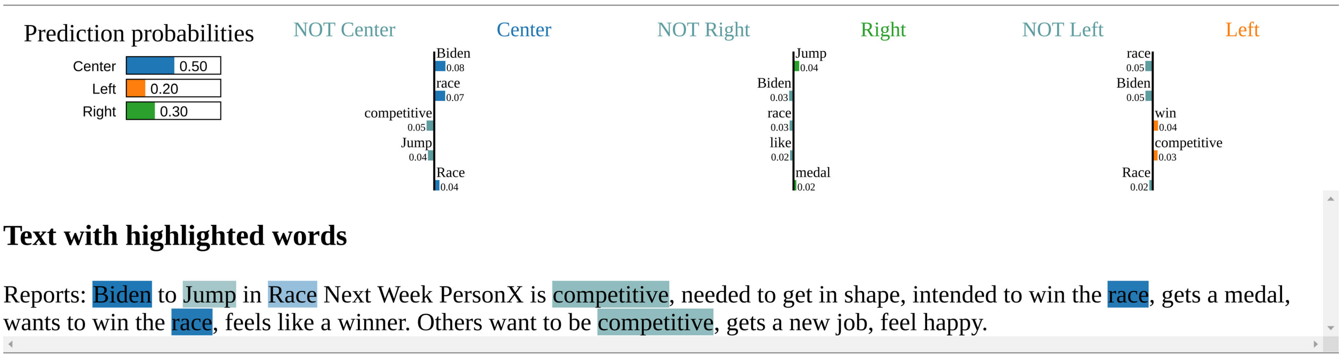


FIGURE 13. True: ***Right***, Predicted: ***Center***. **A flashy headline is typically associated with right-wing ideology. However, including IC_Knwl, which is insignificant for this headline, confuses the model.**

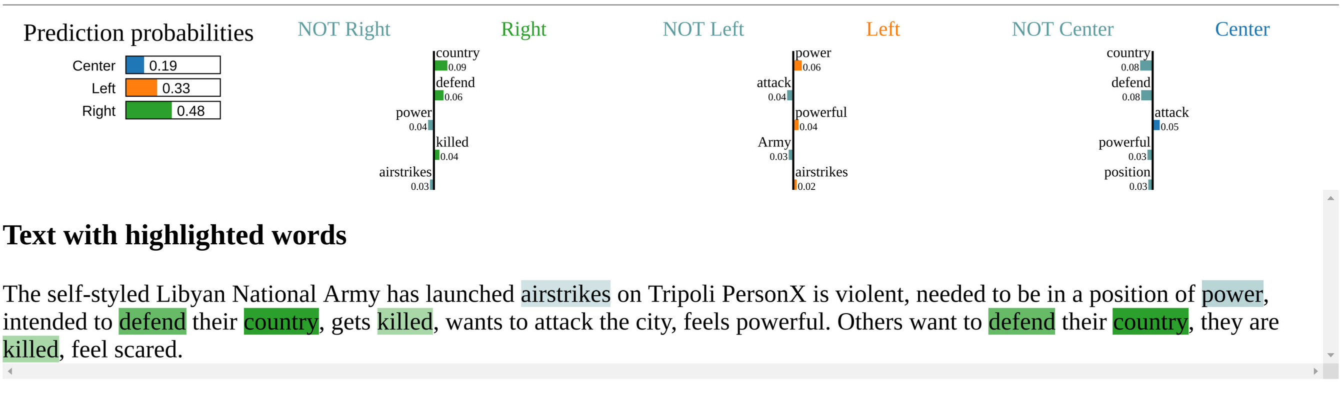


FIGURE 14. True: ***Left***, Predicted: ***Right***. **In order to correctly predict the ideology of this headline, the world's knowledge that the "Libyan National Army" is associated with the "Left-wing" is more important than the IC_Knwl.**

to accelerate training time [82], [83]. Finally, the limitations of the knowledge base and models employed within the framework are also inherent limitations of our proposed framework.

Besides handling the above limitations in the future, additional inferential knowledge bases and their impact on prediction performance can be investigated. Moreover, additional input data, such as information derived from external knowledge bases, can also be pursued.

VII. CONCLUSION

Detecting bias in a news headline can be difficult, as they are typically short and may lack the context of bias embedded in the underlying article. Furthermore, it is challenging to capture syntactic and semantic information in its short text. We propose our framework, *IC-BAIT* to address these challenges. It uses IC_Knwl acquired from *ATOMIC*$^{20}_{20}$ graphs using *COMET* to enhance political bias prediction in news headlines. It employs IC_Knwl in a manner that is very similar to how people use commonsense inferences to carry out their day-to-day tasks.

Due to the scarcity of large-scale datasets for the given task, we also present two datasets: *MediaBias* and *GoodNews*. We conduct experiments on both of them to determine IC-BAIT's generalisability. The results demonstrate that the baseline models, when used within IC-BAIT, had superior performance on both datasets.

We conduct an in-depth, case-by-case investigation using LIME (Local Interpretable Model-Agnostic Explanations), an xAI framework. We discover that, while IC_Knwl can be extremely beneficial in some cases, it can also be counterproductive in others. In a nutshell, when the headline is combined with a carefully chosen IC_Knwl, the model can focus not only on the important entities and events in the headline but also on the explanations for unstated events, resulting in better predictions.

Although the task of classifying political bias in news headlines is still in its early stages, we believe that our method and findings will prove beneficial for copy editors, practical application domains such as e-journalism and manual news bias prediction portals, and automated headline bias-flipping systems, among others.

REFERENCES

[1] N. Hagar, J. Wachs, and E.-Á. Horvát, "Writer movements between news outlets reflect political polarization in media," *New Media Soc.*, vol. 2021, Jan. 2021, Art. no. 14614448211027173, doi: 10.1177/14614448211027173.

[2] P. Maurer and M. Beiler, "Networking and political alignment as strategies to control the news: Interaction between journalists and politicians," *Journalism Stud.*, vol. 19, no. 14, pp. 2024–2041, Oct. 2018, doi: 10.1080/1461670X.2017.1310627.

[3] D. Dor, "On newspaper headlines as relevance optimizers," *J. Pragmatics*, vol. 35, no. 5, pp. 695–721, 2003, doi: 10.1016/S0378-2166(02)00134-0.

[4] N. Hagar, N. Diakopoulos, and B. DeWilde, "Anticipating attention: On the predictability of news headline tests," *Digital Journalism*, vol. 10, no. 4, pp. 1–22, 2021, doi: 10.1080/21670811.2021.1984266.

[5] P. H. Tannenbaum, "The effect of headlines on the interpretation of news stories," *Journalism Quart.*, vol. 30, no. 2, pp. 189–197, Mar. 1953, doi: 10.1177/107769905303000206.

[6] W.-F. Chen, H. Wachsmuth, K. Al-Khatib, and B. Stein, "Learning to flip the bias of news headlines," in *Proc. 11th Int. Conf. Natural Lang. Gener.*, 2018, pp. 79–88, doi: 10.18653/v1/W18-6509.

[7] W.-F. Chen, K. Al Khatib, H. Wachsmuth, and B. Stein, "Analyzing political bias and unfairness in news articles at different levels of granularity," in *Proc. 4th Workshop Natural Lang. Process. Comput. Social Sci.*, 2020, pp. 149–154, doi: 10.18653/v1/2020.nlpcss-1.16.

[8] D. Li, X. Zhu, Y. Li, S. Wang, D. Li, J. Liao, and J. Zheng, "Enhancing emotion inference in conversations with commonsense knowledge," *Knowl.-Based Syst.*, vol. 232, Nov. 2021, Art. no. 107449, doi: 10.1016/j.knosys.2021.107449.

[9] S. B. R. Chowdhury and S. Chaturvedi, "Does commonsense help in detecting sarcasm?" in *Proc. 2nd Workshop Insights Negative Results NLP*, 2021, pp. 9–15, doi: 10.18653/v1/2021.insights-1.2.

[10] T. Chakrabarty, D. Ghosh, S. Muresan, and N. Peng, "R^3: Reverse, retrieve, and rank for sarcasm generation with commonsense knowledge," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 7976–7986, doi: 10.18653/v1/2020.acl-main.711.

[11] L. Huang, R. Le Bras, C. Bhagavatula, and Y. Choi, "Cosmos QA: Machine reading comprehension with contextual commonsense reasoning," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, 2019, pp. 2391–2401, doi: 10.18653/v1/D19-1243.

[12] S. Sabour, C. Zheng, and M. Huang, "CEM: Commonsense-aware empathetic response generation," in *Proc. Conf. Artif. Intell. (AAAI)*, Feb. 2022, vol. 36, no. 10, pp. 11229–11237, doi: 10.1609/aaai.v36i10.21373.

[13] J. D. Hwang, C. Bhagavatula, R. Le Bras, J. Da, K. Sakaguchi, A. Bosselut, and Y. Choi, "(comet-) atomic 2020: On symbolic and neural commonsense knowledge graphs," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, 2021, pp. 6384–6392. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/16792

[14] A. Einsohn and M. Schwartz, *The Copyeditor's Handbook*. Berkeley, CA, USA: Univ. California Press, 2019, doi: 10.1525/9780520972278.

[15] T. Spinde, D. Krieger, M. Plank, and B. Gipp, "Towards a reliable ground-truth for biased language detection," in *Proc. ACM/IEEE Joint Conf. Digit. Libraries (JCDL)*, Sep. 2021, pp. 324–325, doi: 10.1109/JCDL52503.2021.00053.

[16] R. Liu, L. Wang, C. Jia, and S. Vosoughi, "Political depolarization of news articles using attribute-aware word embeddings," in *Proc. Int. AAAI Conf. Web Social Media*, vol. 15, 2021, pp. 385–396. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/18069

[17] R. Liu, C. Jia, and S. Vosoughi, "A transformer-based framework for neutralizing and reversing the political polarity of news articles," *Proc. ACM Hum.-Comput. Interact.*, vol. 5, no. CSCW1, pp. 1–26, Apr. 2021, doi: 10.1145/3449139.

[18] T. Groseclose and J. Milyo, "A measure of media bias," *Quart. J. Econ.*, vol. 120, no. 4, pp. 1191–1237, Nov. 2005, doi: 10.1162/003355305775097542.

[19] F. Hamborg, K. Donnay, and B. Gipp, "Automated identification of media bias in news articles: An interdisciplinary literature review," *Int. J. Digit. Libraries*, vol. 20, no. 4, pp. 391–415, Dec. 2019, doi: 10.1007/s00799-018-0261-y.

[20] P. Navia and R. Osorio, "*El MercurioLies*, and *La TerceraLies* more. Political bias in newspaper headlines in Chile, 1994–2010," *Bull. Latin Amer. Res.*, vol. 34, no. 4, pp. 467–485, Oct. 2015, doi: 10.1111/blar.12364.

[21] M. McCluskey, "A content analysis of 2004 presidential election headlines of The Los Angeles Times and The Washington Times," Ph.D. dissertation, Dept. Commun., College Arts Sci., Univ. Central Florida Orlando, Orlando, FL, USA, 2005. [Online]. Available: https://stars.library.ucf.edu/etd/358

[22] V. P. Aires, F. G. Nakamura, and E. F. Nakamura, "A link-based approach to detect media bias in news websites," in *Proc. Companion World Wide Web Conf.*, May 2019, pp. 1–9. [Online]. Available: https://sentic.net/wisdom2020aires.pdf

[23] Z. Xu, "Examining us newspapers' partisan bias in COVID-19 news using computational methods," *Commun. Stud.*, vol. 2023, pp. 1–19, Jan. 2023, doi: 10.1080/10510974.2023.2169729.

[24] S. D'Alonzo and M. Tegmark, "Machine-learning media bias," *PLoS ONE*, vol. 17, no. 8, Aug. 2022, Art. no. e0271947, doi: 10.1371/journal.pone.0271947.

[25] B. Sinno, B. Oviedo, K. Atwell, M. Alikhani, and J. J. Li, "Political ideology and polarization of policy positions: A multi-dimensional approach," 2021, *arXiv:2106.14387*.

[26] S. Lim, A. Jatowt, M. Färber, and M. Yoshikawa, "Annotating and analyzing biased sentences in news articles using crowdsourcing," in *Proc. 12th Lang. Resour. Eval. Conf.*, 2020, pp. 1478–1484. [Online]. Available: https://www.aclweb.org/anthology/2020.lrec-1.184

[27] T. Spinde, J.-D. Krieger, T. Ruas, J. Mitrović, F. Götz-Hahn, A. Aizawa, and B. Gipp, "Exploiting transformer-based multitask learning for the detection of media bias in news articles," in *Information for a Better World: Shaping the Global Future*. Cham, Switzerland: Springer, 2022, pp. 225–235, doi: 10.1007/978-3-030-96957-8_20.

[28] F. Sakketou, J. Plepi, R. Cervero, H. J. Geiss, P. Rosso, and L. Flek, "FACTOID: A new dataset for identifying misinformation spreaders and political bias," in *Proc. 13th Lang. Resour. Eval. Conf.*, 2022, pp. 3231–3241. [Online]. Available: https://aclanthology.org/2022.lrec-1.345

[29] F. Morstatter, L. Wu, U. Yavanoglu, S. R. Corman, and H. Liu, "Identifying framing bias in online news," *ACM Trans. Social Comput.*, vol. 1, no. 2, pp. 1–18, Jun. 2018, doi: 10.1145/3204948.

[30] L. Fan, M. White, E. Sharma, R. Su, P. K. Choubey, R. Huang, and L. Wang, "In plain sight: Media bias through the lens of factual reporting," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, 2019, pp. 6343–6349, doi: 10.18653/v1/D19-1664.

[31] J. Kulshrestha, M. Eslami, J. Messias, M. B. Zafar, S. Ghosh, K. P. Gummadi, and K. Karahalios, "Search bias quantification: Investigating political bias in social media and web search," *Inf. Retr. J.*, vol. 22, nos. 1–2, pp. 188–227, Apr. 2019, doi: 10.1007/s10791-018-9341-2.

[32] B. Sinno, B. Oviedo, K. Atwell, M. Alikhani, and J. J. Li, "Political ideology and polarization: A multi-dimensional approach," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2022, pp. 231–243, doi: 10.18653/v1/2022.naacl-main.17.

[33] R. Baly, G. Karadzhov, D. Alexandrov, J. Glass, and P. Nakov, "Predicting factuality of reporting and bias of news media sources," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2018, pp. 3528–3539, doi: 10.18653/v1/D18-1389.

[34] R. Baly, G. Karadzhov, J. An, H. Kwak, Y. Dinkov, A. Ali, J. Glass, and P. Nakov, "What was written vs. who read it: News media profiling using text analysis and social media context," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 3364–3374, doi: 10.18653/v1/2020.acl-main.308.

[35] S. Ganguly, J. Kulshrestha, J. An, and H. Kwak, "Empirical evaluation of three common assumptions in building political media bias datasets," in *Proc. Int. AAAI Conf. Web Social Media*, vol. 14, 2020, pp. 939–943. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/7362

[36] R. Baly, G. Da San Martino, J. Glass, and P. Nakov, "We can detect your bias: Predicting the political ideology of news articles," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2020, pp. 4982–4991, doi: 10.18653/v1/2020.emnlp-main.404.

[37] L. A. M. Oberländer, E. Kim, and R. Klinger, "GoodNewsEveryone: A corpus of news headlines annotated with emotions, semantic roles, and reader perception," in *Proc. 12th Lang. Resour. Eval. Conf.*, 2020, pp. 1554–1566. https://www.aclweb.org/anthology/2020.lrec-1.194/

[38] M. A. I. Mahmud, A. A. T. Talukder, A. Sultana, K. I. A. Bhuiyan, M. S. Rahman, T. H. Pranto, and R. M. Rahman, "Toward news authenticity: Synthesizing natural language processing and human expert opinion to evaluate news," *IEEE Access*, vol. 11, pp. 11405–11421, 2023, doi: 10.1109/ACCESS.2023.3241483.

[39] Q. Meng, Y. Song, J. Mu, Y. Lv, J. Yang, L. Xu, J. Zhao, J. Ma, W. Yao, R. Wang, M. Xiao, and Q. Meng, "Electric power audit text classification with multi-grained pre-trained language model," *IEEE Access*, vol. 11, pp. 13510–13518, 2023, doi: 10.1109/ACCESS.2023.3240162.

[40] J. Shin, S.-B. Park, and H.-J. Song, "Token-level fact correction in abstractive summarization," *IEEE Access*, vol. 11, pp. 1934–1943, 2023, doi: 10.1109/ACCESS.2022.3233854.

[41] Y. Sujana and H.-Y. Kao, "LiDA: Language-independent data augmentation for text classification," *IEEE Access*, vol. 11, pp. 10894–10901, 2023, doi: 10.1109/ACCESS.2023.3234019.

[42] Y. Lei, R. Huang, L. Wang, and N. Beauchamp, "Sentence-level media bias analysis informed by discourse structures," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2022, pp. 10040–10050. [Online]. Available: https://aclanthology.org/2022.emnlp-main.682

[43] N. R. Naredla and F. F. Adedoyin, "Detection of hyperpartisan news articles using natural language processing technique," *Int. J. Inf. Manage. Data Insights*, vol. 2, no. 1, Apr. 2022, Art. no. 100064, doi: 10.1016/j.jjimei.2022.100064.

[44] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, vol. 1, 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.

[45] W.-F. Chen, K. Al Khatib, B. Stein, and H. Wachsmuth, "Detecting media bias in news articles using Gaussian bias distributions," in *Proc. Findings Assoc. Comput. Linguistics, EMNLP*, 2020, pp. 4290–4300, doi: 10.18653/v1/2020.findings-emnlp.383.

[46] M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2018, pp. 1–13, doi: 10.18653/v1/N18-1202.

[47] R. R. R. Gangula, S. R. Duggenpudi, and R. Mamidi, "Detecting political bias in news articles using headline attention," in *Proc. ACL Workshop BlackboxNLP, Analyzing Interpreting Neural Netw. NLP*, 2019, pp. 77–84, doi: 10.18653/v1/W19-4809.

[48] T. Mihaylov and A. Frank, "Knowledgeable reader: Enhancing cloze-style reading comprehension with external commonsense knowledge," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, 2018, pp. 821–832, doi: 10.18653/v1/P18-1076.

[49] Q. Chen, X. Zhu, Z.-H. Ling, D. Inkpen, and S. Wei, "Neural natural language inference models enhanced with external knowledge," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, 2018, pp. 2406–2417, doi: 10.18653/v1/P18-1224.

[50] G. A. Miller, "WordNet: A lexical database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, Nov. 1995, doi: 10.1145/219717.219748.

[51] L. Bauer, Y. Wang, and M. Bansal, "Commonsense for generative multi-hop question answering tasks," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2018, pp. 4220–4230, doi: 10.18653/v1/D18-1454.

[52] H. Zhang, Y. Song, Y. Song, and D. Yu, "Knowledge-aware pronoun coreference resolution," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, p. 867, doi: 10.18653/v1/p19-1083.

[53] R. Li, Z. Jiang, L. Wang, X. Lu, M. Zhao, and D. Chen, "Enhancing transformer-based language models with commonsense representations for knowledge-driven machine comprehension," *Knowl.-Based Syst.*, vol. 220, May 2021, Art. no. 106936, doi: 10.1016/j.knosys.2021.106936.

[54] A. Lieto, G. L. Pozzato, S. Zoia, V. Patti, and R. Damiano, "A commonsense reasoning framework for explanatory emotion attribution, generation and re-classification," *Knowl.-Based Syst.*, vol. 227, Sep. 2021, Art. no. 107166, doi: 10.1016/j.knosys.2021.107166.

[55] L. Du, X. Ding, K. Xiong, T. Liu, and B. Qin, "Enhancing pretrained language models with structured commonsense knowledge for textual inference," *Knowl.-Based Syst.*, vol. 254, Oct. 2022, Art. no. 109488, doi: 10.1016/j.knosys.2022.109488.

[56] A. Bosselut, H. Rashkin, M. Sap, C. Malaviya, A. Celikyilmaz, and Y. Choi, "COMET: Commonsense transformers for automatic knowledge graph construction," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 4762–4779, doi: 10.18653/v1/P19-1470.

[57] F. Ilievski, P. Szekely, and B. Zhang, "CSKG: The commonsense knowledge graph," in *Proc. 18th Int. Conf., ESWC*. Cham, Switzerland: Springer, 2021, pp. 680–696, doi: 10.1007/978-3-030-77385-4_41.

[58] M. Sap, R. Le Bras, E. Allaway, C. Bhagavatula, N. Lourie, H. Rashkin, B. Roof, N. A. Smith, and Y. Choi, "Atomic: An atlas of machine commonsense for if-then reasoning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 3027–3035, doi: 10.1609/aaai.v33i01.33013027.

[59] R. Speer, J. Chin, and C. Havasi, "ConceptNet 5.5: An open multilingual graph of general knowledge," in *Proc. AAAI Conf. Artif. Intell.*, vol. 31, 2017, pp. 1–10, doi: 10.1609/aaai.v31i1.11164.

[60] B. Y. Lin, W. Zhou, M. Shen, P. Zhou, C. Bhagavatula, Y. Choi, and X. Ren, "CommonGen: A constrained text generation challenge for generative commonsense reasoning," in *Proc. Findings Assoc. Comput. Linguistics, EMNLP*, 2020, pp. 1823–1840, doi: 10.18653/v1/2020.findings-emnlp.165.

[61] H. A. Brechter. (Jun. 2021). *A Step-by-Step Guide to How Allsides Creates balanced News*. AllSides. [Online]. Available: https://www.allsides.com/blog/how-does-allsides-create-balanced-news

[62] V. Otero. (Sep. 2021). *White Paper: Multi-Analyst Content Analysis Methodology September 2021*. Ad Fontes Media. [Online]. Available: https://adfontesmedia.com/white-paper-2021/

[63] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A lite BERT for self-supervised learning of language representations," 2019, *arXiv:1909.11942*.

[64] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "Distilbert, a distilled version of bert: Smaller, faster, cheaper and lighter," in *Proc. NeurIPS*, 2019, pp. 1–16. [Online]. Available: https://www.emc2-ai.org/assets/docs/neurips-19/emc2-neurips19-paper-33.pdf

[65] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," 2019, *arXiv:1907.11692*.

[66] K. Song, X. Tan, T. Qin, J. Lu, and T.-Y. Liu, "MPNet: Masked and permuted pre-training for language understanding," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 16857–16867. https://proceedings.neurips.cc/paper/2020/file/c3a690be93aa602ee2dc0ccab5b7b67e-Paper.pdf

[67] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–4. [Online]. Available: https://proceedings.neurips.cc/paper/2019/file/dc6a7e655d7e5840e66733e9ee67cc69-Paper.pdf

[68] W. Wang, H. Bao, S. Huang, L. Dong, and F. Wei, "MiniLMv2: Multi-head self-attention relation distillation for compressing pretrained transformers," in *Proc. Findings Assoc. Comput. Linguistics*, 2021, pp. 2140–2151, doi: 10.18653/v1/2021.findings-acl.188.

[69] W. Wang, F. Wei, L. Dong, H. Bao, N. Yang, and M. Zhou, "MiniLM: Deep self-attention distillation for task-agnostic compression of pre-trained transformers," in *Proc. 34th Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 33, 2020, pp. 5776–5788. [Online]. Available: https://proceedings.neurips.cc/paper/2020/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[70] Z. Yang, Y. Yang, D. Cer, J. Law, and E. Darve, "Universal sentence representation learning with conditional masked language model," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2021, pp. 6216–6228, doi: 10.18653/v1/2021.emnlp-main.502.

[71] K. Wang, N. Thakur, N. Reimers, and I. Gurevych, "GPL: Generative pseudo labeling for unsupervised domain adaptation of dense retrieval," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2022, pp. 2345–2360, doi: 10.18653/v1/2022.naacl-main.168.

[72] N. Thakur, N. Reimers, and J. Lin, "Injecting domain adaptation with learning-to-hash for effective and efficient zero-shot dense retrieval," 2022, *arXiv:2205.11498*.

[73] I. Yamada, A. Asai, and H. Hajishirzi, "Efficient passage retrieval with hashing for open-domain question answering," in *Proc. 59th Annu. Meeting Assoc. Comput. Linguistics 11th Int. Joint Conf. Natural Lang. Process.*, 2021, pp. 979–986, doi: 10.18653/v1/2021.acl-short.123.

[74] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, and W. Wang, "Language-agnostic BERT sentence embedding," in *Proc. 60th Annu. Meeting Assoc. Comput. Linguistics*, 2022, pp. 878–891, doi: 10.18653/v1/2022.acl-long.62.

[75] N. Majumder, P. Hong, S. Peng, J. Lu, D. Ghosal, A. Gelbukh, R. Mihalcea, and S. Poria, "MIME: MIMicking emotions for empathetic response generation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2020, pp. 8968–8979, doi: 10.18653/v1/2020.emnlp-main.721.

[76] K. Nantomah. (2019). *On Some Properties of the Sigmoid Function*. [Online]. Available: Asia Mathematika. [Online]. Available: https://hal.archives-ouvertes.fr/hal-02635089/.

[77] N. Balakrishnan, *Handbook Logistic Distribution*. Boca Raton, FL, USA: CRC Press, 2013.

[78] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey," *Information*, vol. 10, no. 4, p. 150, Apr. 2019, doi: 10.3390/info10040150.

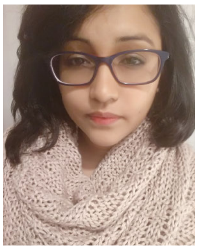
[79] R. Real, J. M. Vargas, and R. Olmstead, "The probabilistic basis of Jaccard's index of similarity," *Syst. Biol.*, vol. 45, no. 3, pp. 380–385, Sep. 1996.

[80] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[81] A. M. Grobelnik, M. Grobelnik, and D. Mladenic, "Multicomet–multilingual commonsense description," in *Proc. 23rd Int. Multiconf. Inf. Soc.*, 2020, pp. 37–40.

[82] J. Zhang, Y. Li, W. Xiao, and Z. Zhang, "Non-iterative and fast deep learning: Multilayer extreme learning machines," *J. Franklin Inst.*, vol. 357, no. 13, pp. 8925–8955, Sep. 2020, doi: 10.1016/j.jfranklin.2020.04.033.

[83] J. Zhang, Y. Zhao, F. Shone, Z. Li, A. F. Frangi, S. Q. Xie, and Z.-Q. Zhang, "Physics-informed deep learning for musculoskeletal modeling: Predicting muscle forces and joint kinematics from surface EMG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 484–493, 2023, doi: 10.1109/TNSRE.2022.3226860.



SWATI SWATI received the M.Tech. degree in mathematics and computing from the Indian Institute of Technology Patna, India. She is currently pursuing the Ph.D. degree with the Jožef Stefan International Postgraduate School, Ljubljana. She is also an Early Stage Researcher (ESR) of the EU-Funded Horizon 2020 Project CLEOPATRA with Jožef Stefan Institute, Ljubljana. Her current research work focuses on the analysis and development of deep learning-based solutions for cross-lingual news reporting bias. It includes domains such as news-value determination and computational journalism. She works on the study and development of machine learning, deep learning, data/text mining, and semantic technology techniques and their applications.

DUNJA MLADENIĆ is leading the Artificial Intelligence Laboratory, Jožef Stefan Institute, Ljubljana. From 2013 to 2017, she was with the Institute's Scientific Council, where she was the Vice President, from 2015 to 2017, and working on several research projects mainly related to machine learning, data and text mining, big data analytic, semantic technologies, and their application to real-world problems. She was with the Jožef Stefan International Postgraduate School, the University of Ljubljana, the University of Primorska, and the University of Zagreb (FERI and FOI), teaching classes related to data analytics and artificial intelligence.

MARKO GROBELNIK is currently an expert researcher in the field of artificial intelligence (AI). He co-leads the Department for Artificial Intelligence, Jožef Stefan Institute Ljubljana; and has co-founded the UNESCO International Research Center on AI (IRCAI). He is also the CEO of Quintelligence.com. He collaborates with major European academic institutions and major industries. He is the coauthor of several books and the co-founder of several start-ups. He is/was involved in over 50 EU-funded research projects in various fields of artificial intelligence. His research interests include machine learning, data/text/web mining, network analysis, semantic technologies, deep text understanding, and data visualization.

Chapter 4

Framework Extension for Multilingual Context

This chapter presents the research published in the article titled *A commonsense-infused language-agnostic learning framework for enhancing prediction of political bias in multilingual news headlines* (Swati, Grobelnik, et al., 2023). The paper was published in the scientific journal *Knowledge-Based Systems*, Volume 277, dated October 2023. The journal holds a Q1 quartile ranking with an impact factor of 8.8 and is indexed in the Science Citation Index Expanded.

In our previous study (Swati, Mladenić, & Grobelnik, 2023), we delved into the challenges of predicting political bias in short news headlines. By utilising Inferential Commonsense Knowledge (IC_Knwl), our objective was to enhance the comprehension of English-language news headlines for the task of predicting political bias. This study expands upon the prior research by incorporating low-resource languages in multilingual settings. The proposed approach involved leveraging IC_Knwl through a translate-retrieve-translate strategy using the Google Translate API [1] for translation. The strategy aimed to uncover contextual features for comprehension of the overall narrative of multilingual headlines. Additionally, it employed an attention mechanism to emphasise relevant inferences within the acquired knowledge. Using this knowledge, we introduced a language-agnostic learning framework for enhancing political bias prediction in low-resource multilingual news headlines, even under imbalanced sample distributions. Ensuring its future adaptability, we maintained a model-agnostic approach, enabling seamless integration with forthcoming advanced multilingual pre-trained language models.

To evaluate the effectiveness of the framework, we introduced a bias-annotated dataset of over 62.6K multilingual news headlines in five low-resource European languages (Czech, Finnish, Romanian, Slovenian, and Swedish). We generated the dataset through a systematic process of crawling headlines from selected media outlets using the Event Registry [2] (Leban et al., 2014) and incorporating bias ratings obtained from the Media Bias/-Fact Check portal [3].

We extensively evaluated our framework using several state-of-the-art multilingual pre-trained language models since their performance tends to vary across languages (low- or high-resource). The results illustrate the effectiveness of the framework across models and languages. Overall, the top-performing model, trained solely with headlines, demonstrated

[1]https://cloud.google.com/translate
[2]https://eventregistry.org
[3]https://mediabiasfactcheck.com

an accuracy of 0.90, an F1 score of 0.83, and a Jaccard score of 0.83. The integration of attended knowledge led to a notable enhancement of 2.2% in accuracy and F1, along with a 3.6% improvement in the Jaccard score. We extended the experiments to individual languages, which also demonstrated significant performance improvements when IC_Knwl was incorporated. In terms of relative performance to models trained solely on headlines, Swedish and Romanian showed the most plausible improvement of 1.43 and 1.42 times in accuracy and F1, and 1.84 and 1.76 times in the Jaccard score, respectively. We also conducted a focused analysis using the Slovenian language as a case study due to its relatively low performance within the dataset. The findings offer valuable insights, uncovering multiple instances of incorrect translation and suggesting a possible connection between the performance gap and the language's inferior translation quality. Extensive analysis helped us uncover multiple instances of incorrect translation, which we grouped into five types: entity detection error, comprehension error, improper sentence formation, inversion of meaning, and miscellaneous error. Our findings offered valuable insights, suggesting a possible connection between the performance gap and the language's inferior translation quality. In conclusion, the method and findings in the study exhibit promise in the task of classifying political bias in multilingual news headlines, even in low-resource, imbalanced settings.

Knowledge-Based Systems 277 (2023) 110838

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A commonsense-infused language-agnostic learning framework for enhancing prediction of political bias in multilingual news headlines

Swati Swati [a,b,*], Adrian Mladenić Grobelnik [a], Dunja Mladenić [a,b], Marko Grobelnik [a]

[a] Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia
[b] Jožef Stefan International Postgraduate School, Jamova cesta 39, 1000 Ljubljana, Slovenia

A R T I C L E I N F O

Article history:
Received 1 December 2022
Received in revised form 20 July 2023
Accepted 21 July 2023
Available online 25 July 2023

Dataset link: https://github.com/Swati17293/KG-Multi-Bias

Keywords:
News
Bias
NLP
Commonsense
Inferential commonsense knowledge
Multilingual
Headline
Low-resource
Imbalanced sample distribution
Pre-trained language models

A B S T R A C T

Predicting the political bias of news headlines is a challenging task that becomes even more challenging in a multilingual setting with low-resource languages. To deal with this, we propose to utilise Inferential Commonsense Knowledge via a Translate–Retrieve–Translate strategy to introduce a learning framework. To begin with, we use the translate–retrieve–translate strategy to acquire inferential knowledge in the target language. We then employ an attention mechanism to emphasise important inferences. We finally integrate the attended inferences into a multilingual, pre-trained language model for the task of bias prediction. To evaluate the effectiveness of our framework, we present a dataset of over 62.6K multilingual news headlines annotated with their respective political biases in five low-resource European languages. We evaluate several state-of-the-art multilingual pre-trained language models since their performance tends to vary across languages (low or high resource). Evaluation results demonstrate that our proposed framework is effective regardless of the models employed. Overall, the best-performing model trained with only headlines shows 0.90 accuracy and F1 and a 0.83 Jaccard score. With attended knowledge in our framework, the same model shows an increase in 2.2% accuracy and F1 and a 3.6% Jaccard score. Extending our experiments to individual languages reveals that the models we analyse for Slovenian perform significantly worse than other languages in our dataset. To investigate this, we assess the effect of translation quality on prediction performance. It indicates that the disparity in performance is most likely due to poor translation quality. We release our dataset and scripts at https://github.com/Swati17293/KG-Multi-Bias for future research. Our framework has the potential to benefit journalists, social scientists, news producers, and consumers.

1. Introduction

News plays a significant role in the functioning of a democratic society [1,2]. Even though it is presumed to be a reliable source of information [3], bias is inevitable [4]. As a result, research communities devote a great deal of attention to the study of news bias [5–7]. However, the first step in conducting such a study is to identify the bias [8,9]. Although the task may appear trivial, it is challenging as bias can manifest itself at different levels in complex ways [10]. When it comes to news headlines, this task becomes even more challenging as headlines are inherently short, catchy or appealing, context-deficient, and contain only subtle bias clues [11,12].

With the rise of digital journalism and microblogging, the headline is becoming the only part of a news item that people read [13]. Furthermore, since it serves as an entry point for an article, people are more likely to form an opinion by simply reading it without reading the rest of the article [14,15]. They seem to be swayed more by its creativity than its clarity [16]. Journalists often use this to their advantage by fabricating facts in a way that expresses their intended point of view, which captures the readers' emotions and interests [14,17].

Such biassed reporting has a direct impact on how the public perceives events such as elections [18], protests [19], terrorism [20], and so on [21,22]. Therefore, it is important to identify bias to help people form an unbiased and well-informed opinion [23,24]. Some studies deal with news bias, but most of them are for High-Resource Languages such as English and German [25,26]. Such research is especially scarce for Low-Resource Languages (LRLs) [27], even though mitigating the effects of bias is equally important in assisting readers of these languages [28].

With a scarcity of standard labelled data, existing studies, and external knowledge to draw from, the task of news bias

* Corresponding author at: Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia.
E-mail addresses: swati@ijs.si (S. Swati), adrian.m.grobelnik@ijs.si (A.M. Grobelnik), dunja.mladenic@ijs.si (D. Mladenić), marko.grobelnik@ijs.si (M. Grobelnik).

https://doi.org/10.1016/j.knosys.2023.110838

(a) **Novinky.cz (Czech)**: Hackeři vyhlásili Rusku válku, vyřazují z provozu jeden cíl za druhým (Hackers have declared war on Russia, decommissioning one target after another)
IC_Knwl: Hackeři jsou vidět jako 'agresivní', který 'chce zničit nepřítele' (Hackers are *seen as* 'aggressive' who *wants to* 'to take revenge on Russia')
Bias: Left Center

(b) **24ur.com (Slovenian)**: Hekerska skupina Anonymous trdi, da je vdrla v rusko centralno banko (The hacker group Anonymous claims to have hacked into Russia's central bank)
IC_Knwl: Hekerji veljajo za 'zlonamerne', ki želijo 'dati izjavo' (Hackers are *seen as* 'malicious' who *wants to* 'make a statement')
Bias: Least Biased

Fig. 1. News headlines from (a) Czech and (b) Slovenian news outlets on the "*hacker attacks on Russia*" with varying political ideologies. IC_Knwl can help improve prediction accuracy by facilitating the acquisition of additional bias cues.
(Note: this example shows only a subset of IC_Knwl relations).
Source: 24ur.com, novinky.cz, Translation: translate.google.com

identification in these LRLs becomes even more challenging [27,29]. As a result, resolving these issues necessitates understanding the narrative being presented [30]. This can be accomplished by identifying connections between what is explicitly stated and what is implied [31].

It is well known that incorporating commonsense reasoning abilities can facilitate the inference of such connections by identifying a set of unstated causes and effects [32,33]. Such additional knowledge has been proven to be beneficial for several tasks [34–36], including the prediction of bias in English news headlines [37]. To this end, we use the popular neural knowledge model COMET [38] trained on ATOMIC$^{20}_{20}$ [38] to generate the Inferential Commonsense knowledge (IC_Knwl). Since the textual descriptions of commonsense in the ATOMIC$^{20}_{20}$ knowledge repository are composed in English, it creates a language barrier.

Thus, to extend its capability beyond this barrier, we propose to leverage the Translate-Retrieve-Translate (TRT) approach [39]. Specifically, given a headline in the target language, TRT first translates it into English and then acquires the associated knowledge in English. It then translates the knowledge back into the target language. As illustrated in Fig. 1, IC_Knwl in the target language can help enhance prediction accuracy.

To finally predict the political bias of multilingual news headlines, we present a learning framework in Section 4.2.1. Given a multilingual headline, we first use COMET with TRT to acquire IC_Knwl in the target language. Next, we employ an attention mechanism to emphasise important inferences. We finally integrate the attended IC_Knwl into a multilingual, pre-trained language model for bias prediction.

However, there are no standard labelled datasets available for evaluating our framework [27]. Prior studies either restricted their scope to news in a single language [11] or analysed news in different languages separately [40]. Even the overall ratings for news outlets that publish in these languages are unavailable on popular bias rating platforms such as allsides.com and adfontesmedia.com.

Given the limited number of news outlets publishing in these LRLs for each bias class [41], imbalanced data distribution poses another challenge. Furthermore, no labelled data may exist for some LRLs. Especially for European LRLs, data and knowledge resources are extremely scarce [29]. To this end, we present our dataset of news headlines in five European LRLs annotated with their respective political leanings (ref. Section 3). It is constructed to mimic the challenges encountered by LRLs.

For a model to overcome the aforementioned challenges, cross-lingual transfer learning is crucial [42–44]. It can be achieved with the help of multilingual Pre-trained Language Models (PLMs) [45–47]. These models can generate vector embeddings of texts in different languages that are aligned in a single vector space, enabling few-shot or zero-shot learning. Advances in multilingual PLMs have shown promise in numerous NLP tasks [48,49]. However, to use them effectively, systems must be fine-tuned to the task at hand [50]. Unfortunately, as stated previously, the majority of these LRLs lack large enough data sets for such fine-tuning. They also suffer from the problem of specificity in their vocabulary, which focuses on their cultural heritage, which further hinders the performance of these models [51]. Therefore, in this study, we also evaluate several state-of-the-art multilingual PLMs for their effectiveness (ref. Section 4.2.4).

1.1. Contributions

The key contributions of our work are summarised as follows:

- Proposing to leverage IC_Knwl through a TRT strategy to facilitate comprehension of the overall narrative of the multilingual headlines.
- Introducing an IC_Knwl-infused language-agnostic learning framework for enhancing the prediction of political bias in multilingual news headlines under imbalanced sample distribution.
- Presenting a dataset of multilingual news headlines annotated with their respective political bias in five low-resource European languages.
- Thorough experiments with several state-of-the-art multilingual pre-trained language models to assess their effectiveness.
- Analysing the impact of IC_Knwl infusion on overall performance and across languages with and without attention mechanisms.

The remainder of this paper is structured as follows: After a brief review of the key related works in Section 2, we introduce our dataset and provide an overview of its data collection framework in Section 3. We then present the materials and methods utilised in this study in Section 4. In Section 5, we present the results and analysis of our experiments, followed by research implications in Section 6. Finally, in Section 7, we present the concluding remarks and potential directions for future research.

2. Literature review

Researchers have long been interested in studying news articles and headlines to address problems such as fake news detection [52–54], sentiment analysis [55,56], topic modelling [57,58], and so on [59,60]. While predicting the bias (e.g., left, right, centre) of news articles is not a new problem [61–63], modelling it at the headline level has received less attention [37]. In this study, we predict the political bias of multilingual news headlines by incorporating commonsense knowledge into a pre-trained multilingual language model. Consequently, we organise the related work in this section from these three perspectives as follows:

2.1. Prediction of bias in multilingual news headlines

Related studies on bias prediction relied on predefined linguistic feature sets [64,65] and standard machine learning techniques [66]. Recent studies, on the other hand, have advanced to deep-learning techniques [11,67,68]. In particular, Transformers-based models have demonstrated remarkable performance enhancements [69,70]. However, the majority of these studies focus on languages with abundant resources, with only a few exceptions studying languages with limited resources [11]. Moreover, these studies are either limited to a single language [22,71] or analyse news in different languages independently [40].

The lack of large-scale annotated gold-standard datasets for these languages further complicates the task [27,51]. Most existing datasets were generated manually [11]. Manual annotation requires a substantial amount of time and effort. Moreover, these small-scale datasets are not suitable for training deep learning models [72]. There are also datasets generated using an approach in the form of distant supervision, in which the ideology of a news outlet is mapped to each of its articles [64,73]. The bias score is typically obtained from prominent bias rating platforms, such as allsides.com and adfontesmedia.com where a team of domain experts employs specialised guidelines for annotations. Even though distant supervision facilitates the creation of large datasets, bias ratings are typically not available for all outlets, especially those that publish in languages with limited resources [41]. Another possibility is to combine the datasets available in different languages. However, this strategy would result in an uneven distribution of topics and events across biassed classes and languages.

To mitigate the aforementioned issues of data scarcity, we present a diverse and scalable multilingual news headline dataset in five low-resource languages to predict political leanings (ref. Section 3). Inspired by but distinct from these related works, we then introduce our learning framework (ref. Section 4.2.1). We infuse it with inferential commonsense knowledge and explore its application for the task of bias prediction. Furthermore, we propose a language-agnostic learning framework that we use to evaluate the effectiveness of several state-of-the-art multilingual pre-trained language models.

2.2. Commonsense knowledge

Multiple studies have revealed that large-scale pre-trained language models are implicitly capable of encoding some commonsense and factual knowledge [74,75]. However, these models hardly acquire inferential commonsense knowledge, especially in context-deficient settings [76,77]. Consequently, recent studies have investigated the application of such knowledge in a number of NLP-related tasks [78–80]. It has been demonstrated that injecting such knowledge improves output performance on a variety of tasks, including reading comprehension [81], question answering [82], and story generation [83], among others [84–86].

There exist several widely used commonsense knowledge resources such as ConceptNet [87], SentiNet [88], GLUCOSE [89], ATOMIC$^{20}_{20}$ [38], etc. [90–92]. ConceptNet is a semantic network containing concept-level relational commonsense knowledge as phrases and words in natural language. SentiNet is a well-known resource used for sentiment analysis at the concept level. GLUCOSE is a large-scale resource used for capturing implicit casual knowledge in narrative contexts. Structured as if-then relations with an emphasis on inferential knowledge, ATOMIC$^{20}_{20}$ is a resource composed of everyday commonsense knowledge.

These knowledge resources are used to train generative models such as COMET [38] and ParaCOMET [93]. Trained on ConceptNet and ATOMIC$^{20}_{20}$, COMET is capable of generating a diverse range of context-relevant commonsense descriptions. Motivated by the related studies, we thus use COMET trained on the ATOMIC$^{20}_{20}$ knowledge base. However, unlike these studies, we use it to identify unstated causes and effects in context-deficient headlines.

2.3. Multilingual pre-trained language models

A number of language representation models, such as BERT [94], ELECTRA [95], XLNet [96], etc. [97,98], have emerged in recent years. The majority of them are based on transformers, a non-sequential deep learning approach that provides positional embeddings via a multi-headed attention technique [99]. Due to their many advantages [100], they are popular not only for solving a wide range of NLP-related tasks [101–104] but also for a variety of other practical applications [105–107].

A number of their multilingual variants, such as Multilingual BERT (mBERT) [94], XLM-RoBERTa (XLM-R) [108], and Multilingual Bidirectional Auto-Regressive Transformers (mBART) [109], have shown promising results for text processing across multiple languages [42,110,111]. If followed by task-specific fine-tuning, they have proven to be effective [112]. However, they are ineffective at generating sentence-level representations [113].

Several models designed to generate semantically meaningful sentence representations, such as Sentence BERT (SBERT) [114], Universal Sentence Encoder (USE) [113], and Language-Agnostic Sentence Representations (LASER) [115], were proposed to address this limitation. They have proven useful in a variety of NLP applications [116,117]. Over the past few years, several similar frameworks have been extended to support over 100 languages [113,118]. Some even support low-resource languages such as Slovenian, Romanian, and so on [47,115].

Despite having millions of parameters and being trained on diverse datasets, these models are not guaranteed to generalise to all tasks and domains [112]. As a result, we investigate and compare several state-of-the-art PLMs in this study for their effectiveness.

3. Dataset

We introduce our dataset and describe its data collection framework in this section. To begin with, we introduce two primary data sources that serve as the foundation for our dataset. We then present a detailed description of our framework for data collection, followed by a description of our dataset.

3.1. Primary data sources

In this section, we present two primary data sources: Media Bias/Fact Check (MBFC) and Event Registry (ER). We use the bias rating portal MBFC to select media outlets and retrieve their associated bias labels. We use ER to crawl the headlines of articles published by these selected media outlets.

3.1.1. Media bias fact/check

Several well-known platforms, such as allsides.com, adfontesmedia.com, and mediabiasfactcheck.com [119], publish bias ratings for media outlets. However, due to the scarcity of such ratings for outlets in low-resource languages, we chose to acquire labels exclusively from MBFC. It is a trustworthy bias rating and fact-checking platform with extensive coverage and regular updates. It has been employed to predict and assess media bias in a number of studies [120]. In addition, it has also been utilised to develop tools such as 'Iffy Quotient' [121], which monitors the prevalence of fake news and questionable sources on social media.

To assign bias ratings to media sources, it establishes five levels of political bias: 'left', 'left-center', 'center', 'right-center', and 'right' [122]. It also assigns ratings based on their credibility and factual accuracy. These ratings are assigned by a group of paid contractors and volunteers who are instructed to adhere to a predetermined methodology [123]. Based on a quantifiable system, its methodology includes both objective and subjective measures.

3.1.2. Event registry

To scrape news headlines, we use the Event Registry [124] platform. It has a custom collection of over 150,000 diverse sources from around the world in over 50 languages. It is widely used in studies involving news event analysis [125–127]. Its primary objective is to cluster contents as events, but it also facilitates the collection of news stories and articles. It offers a Python API[1] for accessing news content minutes after it has been published online. It has several search options for filtering out the desired content, such as searching by any news outlet, keyword, language, and others. Using this API, it is possible to extract news content as well as metadata published by different publishers in different languages.

3.2. Data collection framework

As illustrated in the data collection framework in Fig. 2, we begin the process by compiling a list of low-resource European languages (L). $L = \{l_1, l_2, \ldots, l_n\}$, with n representing the total number of languages in the list. $\forall l \in L$, we then compile a list of media outlets (O) publishing in l ranked by MBFC (ref. Section 3.1.1). We define $O = \{o_1, o_2, \ldots, o_m\}$, with m as the total number of outlets in the list. $\forall o \in O$, we then check whether o is ranked as a questionable source or not. Since questionable sources are prone to promote unfounded claims or theories as facts and offer little or no references to credible sources of information, they may turn out to be untrustworthy. Therefore, we discard such sources. $\forall$ unquestionable o, we extract the political bias label b assigned by MBFC.

We then define an explicit temporal query (Q_t):

$$Q_t = \{Q_o,\ Q_l,\ Q_{cat},\ Q_{dt}\} \tag{1}$$

where, Q_o, Q_l, and Q_{cat} defines the query o, l, and categories[2] respectively, and Q_{dt} defines the time constraint with Q_{sd} and Q_{ed} representing the start and end dates:

$$Q_{dt} = [Q_{sd}, Q_{ed}] \tag{2}$$

To scrape all the article headlines (H) published by each unquestionable o, we utilise Q_t to query ER (ref. Section 3.1.2):

$$H = ER\,(Q_t) \tag{3}$$

Finally, we assign the previously extracted bias label b to the headlines in H to construct the dataset. To generate the train/valid/test splits, we adopt a stratified split to simulate the imbalance in the collected data across the languages.

3.3. Dataset description

Our dataset consists of news headlines annotated with their respective political leanings. We construct it to mimic the challenges encountered by LRLs. We begin by selecting five low-resource European languages: *Czech*, *Finnish*, *Romanian*, *Slovenian*, and *Swedish*. We then compile a list of media outlets ranked by MBFC in these selected languages. We end up with seven news outlets: *24ur*, *Dagens Nyheter*, *Delo*, *Digi24*, *Helsingin Sanomat*, *Hotnews*, and *Novinky* with bias labels *Left Centre*, *Least Biased*, and *Right centre*. In the end, we manage to generate 62,689 news headlines with an average length of 10.2 words.

In Table 1, we list the statistics for each language in the dataset. It is carefully documented and adheres to the requirements of the FAIR Data Principles[3].

4. Materials and methods

In this section, we begin by stating the research objectives, followed by formally defining the task of predicting the political bias of multilingual news headlines. We then present our learning framework and its key components, followed by a brief discussion of baseline models and the evaluation metrics used in this study.

4.1. Research objectives

The primary objective of this study is to investigate the impact of our proposed framework for predicting political bias in multilingual news headlines. It takes advantage of state-of-the-art pre-trained language models and inferential commonsense knowledge in a multilingual setting. In this context, we define the following research objectives:

- **RO1:** Introduce a knowledge-infused and language-agnostic learning framework.
- **RO2:** Evaluate the impact of using inferential commonsense knowledge as a source of additional information in a multilingual setting.
- **RO3:** Compare the effectiveness of several state-of-the-art multilingual pre-trained language models.
- **RO4:** Investigate the influence of knowledge attention on prediction performance.

1 https://eventregistry.org

2 https://eventregistry.org/documentation?tab=suggCategories Note: For our dataset, we only use the categories defined by ER as 'news'.

3 https://www.nature.com/articles/sdata201618/

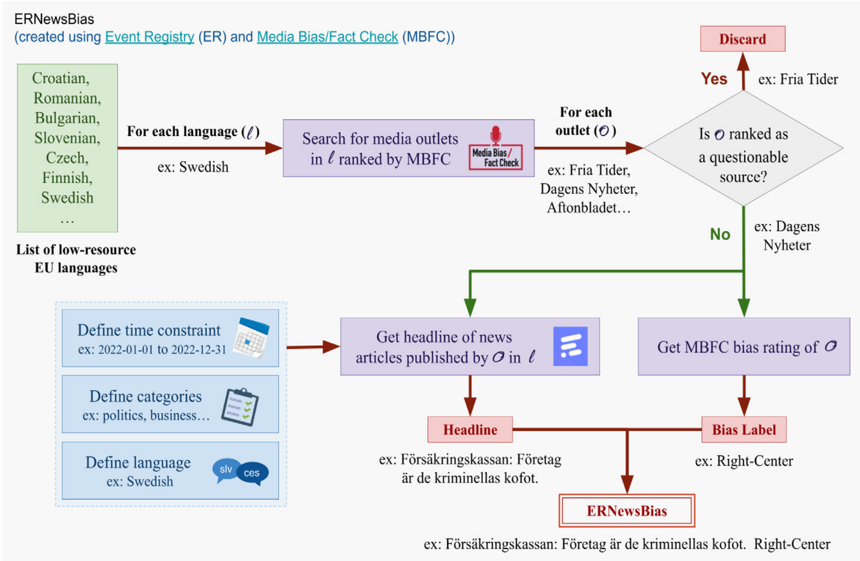


Fig. 2. Data Collection Framework. We use MBFC and ER as the primary data sources in the framework.

Table 1
Dataset Statistics. Len: average number of words in the headline.

	All	Czech	Finnish	Romanian	Slovenian	Swedish
Train	50,157	9,992	7,120	5,829	15,557	11,659
Test	6,269	1,237	940	756	1,879	1,457
Valid	6,263	1,310	880	764	1,853	1,456
Total	62,689	12,539	8,940	7,349	19,289	14,572
Len.	10.2	9.4	10.2	12.8	8.8	8.9

4.2. Task definition

We denote a language by $l \in L$, a short news headline text by H, an auxiliary piece of information as inferential commonsense knowledge by IC_Knwl, a H in l as H^l, an IC_Knwl in l as IC_Knwl^l, and a political bias label by $b \in B$. We define the sets $L = \{l_1, l_2, \ldots, l_n\}$ and $B = \{b_1, b_2, \ldots, b_N\}$, where n and N represent the number of languages and bias labels in the respective sets L and B. Given H^l, its corresponding IC_Knwl^l can be acquired using the commonsense knowledge modelling function C with the appropriate model parameters α, as shown in Eq. (4).

$$IC\_Knwl^l = C(H^l, \alpha) \tag{4}$$

H^l can then be fused with the acquired IC_Knwl^l to represent its extended feature space (H^l, IC_Knwl^l). Given H^l, the task aims to train a classifier that maps its extended feature space to the bias set B. It can be mathematically formulated using Eq. (5) with f as the bias prediction function and θ as the model parameters.

$$b = f((H^l, IC\_Knwl^l), \theta) \tag{5}$$

4.2.1. Methodology

To fulfil **RO1**, we propose a framework that is primarily based on inferential commonsense knowledge. It helps uncover contextual features that, in turn, can help predict the bias of multilingual news headlines. To facilitate generalisation, our framework is compatible with any multilingual pre-trained language model. Fig. 3 depicts its overall architecture. Its key components include knowledge acquisition, knowledge translation, feature encoding, knowledge attention, and bias prediction. Each of these components is described in detail in the following subsections.

4.2.2. Knowledge acquisition

The ATOMIC$^{20}_{20}$ (ATlas Of MachIne Commonsense 2020)[4] [38] is a well-known, publicly available commonsense knowledge resource that is "*able to cover more correct facts about more diverse types of commonsense knowledge than any existing, publicly-available commonsense knowledge resource*". Its relations are composed of textual descriptions containing more than one million tuples of everyday inferential knowledge about entities and events. It is coded into different relation types, which are categorised into different sub-types, such as nine commonsense relations for social interaction, seven for physical entities, and seven for events. Fig. 4 illustrates a subset of these relations generated in response to a sample news headline.

Relations of this type of social interaction provide an insight into socially triggered states and behavioural patterns. As demonstrated by the examples in Table 2, it is valuable for predicting people's reactions and behaviour in a given situation by assessing their intentions and goals. Motivated by its effectiveness in enhancing the performance of models designed to handle short news headlines in English language [37], we use it as the sole relation type for IC_Knwl in our work.

To retrieve IC_Knwl, we use COMmonsensE Transformers (COMET)[5] [38,128] trained on the ATOMIC$^{20}_{20}$ knowledge graphs. COMET is a large, pre-trained neural network model that generates IC_Knwl in response to a query text. Given H, Inference type (I_{type}), and number of returned references (k), IC_Knwl can be retrieved using the following equation,

$$IC\_Knwl = COMET(H, I_{type}, k) \tag{6}$$

4 https://allenai.org/data/atomic-2020
5 https://github.com/allenai/comet-atomic-2020/

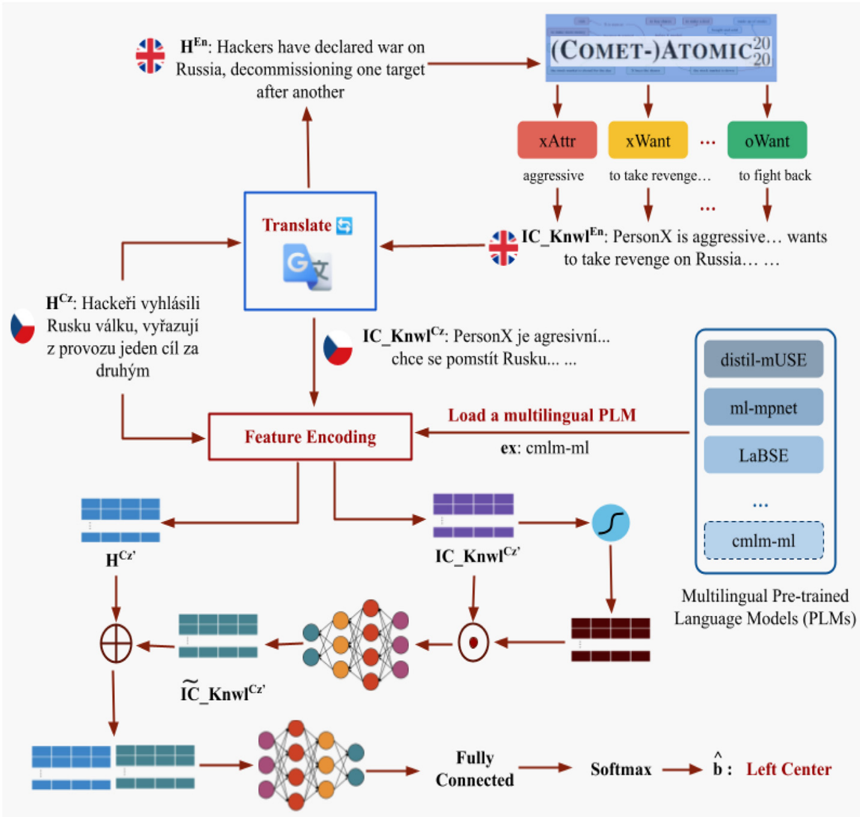


Fig. 3. An overview of our proposed learning framework. To predict the political bias of multilingual news headlines, it combines Inferential Commonsense Knowledge retrieved via the Translate–Retrieve–Translate strategy with multilingual pre-trained language models.

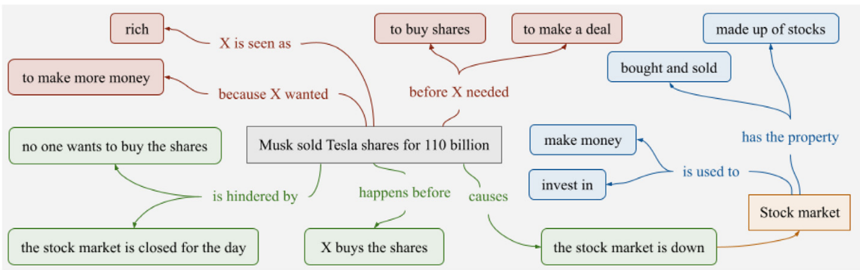


Fig. 4. A small subset of *IC_Knwl* relations generated using ATOMIC$^{20}_{20}$ as the knowledge base in response to the news headline '*Musk sold Tesla shares for 110 billion*'. Nodes in the colours red, green, blue, and orange represent relations depicting social interactions, events, physical entities, and category intersections, respectively.

Table 2
Examples of social interaction relations retrieved using ATOMIC$^{20}_{20}$ as the knowledge base for the short news headline 'Grit Won'. Each relation type is interpreted using the human-readable template provided in [38].

Relation	Interpretation	Examples
xAttr	X is seen as	Lucky; competitive
xEffect	As a result, X	Wins the game; personx wins the race
xIntent	Because X wanted	To win; to be the best
xNeed	But before, X needed	To train hard; to enter the contest
xReact	As a result, X feels	Happy; excited
xWant	As a result, X wants	To celebrate; to win
oEffect	As a result, others	Loses the game; loses money
oReact	As a result, others feel	Disappointed; sad
oWant	As a result, others want	To congratulate X, to win the game

where $I_{type} = [i_1, i_2, \ldots, i_x]$ with i as the inference type defined in Table 2 and x as the total relations in the set. Since COMET returns the *IC_Knwl* as a list of inference results $\forall i \in I_{type}$, we set $k = 1$ to return only one inference result per I_{type}. Furthermore, while retrieving *IC_Knwl*, we combine the returned pieces of inferences of each I_{type} to make it more meaningful. For example,

> **Headline:** *Grit Won*
> **IC_Knwl:** xAttr: *lucky*, xIntent: *to win*, xEffect: *wins the game*, xWant: *to celebrate*, xReact: *happy*, oWant: *to congratulate X*, oEffect: *looses the game*, oReact: *disappointed*
> **Processed IC_Knwl:** PersonX is *lucky*, needed *to train hard*, intended *to win*, *wins the game*, wants *to celebrate*, feels *happy*. Others want *to congratulate X*, *looses the game*, feel *disappointed*.

4.2.3. Knowledge translation

To generate IC_Knwl^l for H^l, we employ the Translate-Retrieve-Translate approach inspired by the work of Fang et al. [39]. Specifically, given a H^l, we first translate it into English and retrieve its associated *IC_Knwl* in English. We then translate the retrieved *IC_Knwl* into the target language l to finally get the IC_Knwl^l. We use the Google Translate API[6] for our translations.

4.2.4. Feature encoding

To acquire feature vectors $H^{l'}$ and $IC\_Knwl^{l'}$, we use multilingual PLMs. For their optimal performance, they are required to map embedding vectors of text written in different languages into a single vector space. As a result, the degree of vector alignment influences their performance. In this regard, we explore the state-of-the-art multilingual PLMs defined in Section 4.3. These PLMs differ from word-embedding models as they are trained on a wide range of tasks that require modelling the meaning of word sequences as opposed to individual words.

4.2.5. Knowledge attention

Ideally, not all retrieved inferences are expected to be of the same relevance. To address this, we employ a strategy similar to one that has been shown to be effective in knowledge selection [129] and other related problems [130,131].

We apply the Sigmoid function [132] to transform the feature vector of the retrieved inferences ($IC\_Knwl^{l'}$). In particular, we use the Sigmoid as a cumulative density function of the logistic distribution that assigns a probability to the inferences [133]. If fed to a Multi-Layer Perceptron (MLP) trained to determine the relevance of each transformed inference, it can aid in obtaining the relevance vector. However, we require significant inferences, not just a score indicating how relevant they are.

As a result, we multiply the relevance vector by $IC\_Knwl^{l'}$ and then feed the resulting vector to an MLP. We then train the MLP to highlight the most significant inferences by mixing different inferences from different I_{type} to finally generate the attended vector $\widetilde{IC}\_Knwl^{l'}$:

$$\widetilde{IC}\_Knwl^{l'} = MLP(Sigmoid(IC\_Knwl^{l'}) \odot IC\_Knwl^{l'}) \tag{7}$$

where $\odot$ denotes element-wise multiplication.

4.2.6. Bias prediction

To predict the bias label $\hat{b}$, we first fuse the vectors $H^{l'}$ and $\widetilde{IC}\_Knwl^{l'}$ to generate F:

$$F = H^{l'} \oplus \widetilde{IC}\_Knwl^{l'} \tag{8}$$

where $\oplus$ represents the concatenation operation.

We then feed the fused vector F to an MLP network and forward the resultant vector to a Fully Connected layer (FC) having Softmax (σ) activation to finally predict $\hat{b}$:

$$\hat{b} = FC(\sigma(MLP(F))) \tag{9}$$

We train our network using AdaMax [134] as the optimiser with its default parameters. We use the Categorical cross-entropy as the loss function, which is defined as follows:

$$Loss = -\sum_{i=1}^{|B|}(b_i * \log(\hat{b}_i)) \tag{10}$$

where b_i and $\hat{b}_i$ are the actual and predicted probabilities of selecting the ith bias label in B.

4.3. Baseline models

Based on their superior performance in a variety of related tasks in multilingual settings [49,135], we chose the following state-of-the-art baseline models for a comprehensive evaluation of our proposed framework.

- **ml-MiniLM** [114] [paraphrase-multilingual-MiniLM-L12-v2]: a multilingual version of the sentence transformer, paraphrase-MiniLM-L12-v2 [46]. It generates 384-dimensional aligned dense vectors. It is pre-trained on parallel data for more than 50 languages. It trades accuracy for speed, and its reduced dimensions result in lower memory requirements.
- **distil-mUSE** [114] [distiluse-base-multilingual-cased-v2]: multilingual Universal Sentence Encoder (mUSE) [136] is based on the transformer architecture [99], which uses a multi-task trained dual encoder to embed texts into a single vector space. The multilingual knowledge-distilled version of mUSE (distil-mUSE) supports over 50 languages. It maps text to a 512-dimensional dense vector space.
- **ml-mpnet** [114] [paraphrase-multilingual-mpnet-base-v2]: a multilingual version of the sentence transformer, paraphrase-mpnet-base-v2 [46]. It is pre-trained on parallel data for over 50 languages and generates 768-dimensional aligned dense vectors. It outperforms other multilingual models based on sentence transformers. However, its increased computational complexity makes it time-intensive.
- **LaBSE** [45][LaBSE/2]: a language-agnostic BERT-based model that maps text into a 768-dimensional dense vector space. To map single plain-text segments to encoder inputs, it requires a separate preprocessor API build for the universal-sentence-encoder-cmlm multilingual models[7]. It is trained and optimised to generate aligned vectors for bilingual sentence pairs, and it currently supports over 109 languages. Although the model, like other BERT models, can be fine-tuned, the authors recommend that it be used as it is.
- **cmlm-ml** [47] [cmlm/multilingual-base/1]: a multilingual model trained with a conditional masked language model (cmlm-ml). Its architecture is based on a 12-layer BERT transformer [137], but it is far more complex. Similar to LaBSE, it also requires an additional preprocessor to map plain-text inputs to encoder inputs. It transforms text into 768-dimensional aligned vectors and supports more than 100 languages. Although its inference speed is significantly slower than that of other comparable models, its performance is far superior.

6 https://cloud.google.com/translate

7 https://tfhub.dev/google/universal-sentence-encoder-cmlm/multilingual-preprocess/2

Table 3
Description of the values of the confusion matrix.

True Positive (TP):	label is present and is predicted.
True Negative (TN):	Label is not present and is not predicted.
False Positive (FP):	Label is not present but is predicted.
False Negative (FN):	Label is present but is not predicted.

4.4. Evaluation metrics

To assess the performance of our proposed framework, we employ well-known metrics used to evaluate prediction models [138], such as Accuracy and F_1-score. However, in the case of an imbalanced dataset like ours, where true negative instances outnumber true positive instances for several languages, they are not a reliable indicator. The Jaccard [139] score is a reliable metric for evaluating models where no examples exist for each class. It disregards true negatives in favour of true positives, facilitating the interpretation of the results. It is even more reliable when evaluating models for individual languages since the imbalance is more apparent. As a result, we also employ the Jaccard score to gain a deeper understanding. We compute these metrics using the values of the confusion matrix defined in Table 3.

The metrics we use are defined as follows:

- **Accuracy (A)**: fraction of true prediction over the total.

$$A = (TP + TN)/(TP + TN + FP + FN) \tag{11}$$

- **F_1-score** (F_1): harmonic mean of Precision (P) and Recall (R), where P is the fraction of relevant instances among the retrieved instances and R represents the fraction of relevant instances that were retrieved:

$$F_1 = 2TP/(2TP + FP + FN) \tag{12}$$

- **Jaccard (J)**: fraction of correctly predicted instances over all instances except those where a label is not present and is not predicted.

$$J = (TP)/(TP + FP + FN) \tag{13}$$

To ensure all bias classes are treated equally, we use the macro-averaged F_1 and macro-averaged J (J_m) scores to evaluate the overall performance of the models.

To evaluate the performance of the models for each language, we use the micro-averaged J (J_μ) score which accounts for the problem of class imbalance. Inspired by Nagle [140], we also report the Relative Performance (RP) of the models for each language used in our study. RP is defined as the ratio of the absolute performance of the models under consideration. To compute it, any underlying evaluation metric (e.g. accuracy, Jaccard, etc.) can be used. In particular, we report on the relative performance of models trained with only headlines compared to those trained with or without additional knowledge and attention mechanisms.

5. Results and discussion

We begin this section by analysing the experimental results of the models trained across all reported languages. Following that, we examine the performance of the models evaluated for individual languages. Finally, we present the findings of a case study investigating the effect of translation quality on prediction accuracy.

5.1. Overall performance

We evaluate the baseline models and our proposed framework across all the reported languages and present their performance in terms of accuracy(A), macro-averaged-F_1 (F_1), and macro-averaged-Jaccard(J_m) scores in Table 4. As the results indicate, with 0.92 A and F_1, and 0.86 J_m, our proposed framework trained with headlines and attended IC_Knwl using cmlm-ml outperforms other models trained with headlines only. It surpasses the performance of the best model (cmlm-ml) in terms of A and F_1, and J_m respectively by 2.2% and 3.6%.

To determine whether IC_Knwl contains knowledge useful for bias prediction, we train the models with IC_Knwl as the only input. We reported the results in the first column of Table 5. We observe that models trained exclusively with IC_Knwl achieve comparable results to models trained only with headlines. In terms of A, F_1, and J_m scores, models other than cmlm-ml show an average improvement of 22%, 28%, and 47%, while cmlm-ml shows a slight decrease in performance of 5%, 2%, and 5% respectively. The findings demonstrate that the IC_Knwl provide useful inferential information for the task of bias prediction (**RO2**).

Furthermore, as evident in column two of Table 5, integrating IC_Knwl with the headlines can significantly improve the performance of all models by enhancing their reasoning abilities. In terms of A, F_1, and J_m scores, these models exhibit average performance improvements of 4%, 4%, and 7%, respectively, over models trained exclusively with IC_Knwl.

Integration of IC_Knwl, on the other hand, may not always function as expected and may introduce unwanted noises. Given the fact that they are generated automatically rather than manually, noise is inevitable, which may weaken their role in bias prediction. To minimise the impact of this noise, we integrate IC_Knwl with an attention mechanism and reported the results in column three of Table 5. The introduction of attention results in an average performance gain of 5%, 4%, and 9% in terms of A, F_1, and J_m scores, respectively.

The high performance of the models can be attributed to their deep network architectures, which enable them to learn rich universal text representations. Furthermore, it demonstrates that integrating IC_Knwl significantly improves their performance, while the introduction of attention improves it even further (**RO4**). To summarise, the results indicate that our proposed framework for bias prediction is effective regardless of the models used (**RO3**).

5.2. Language-wise performance

The models evaluated for individual languages present plausible results, as shown in Table 6. However, the performance of models across languages varies significantly due to an imbalanced number of samples in each class.

Among all the low-resource languages present in the dataset used for this study, the models analysed for Czech demonstrate the most impressive performance, with an average A, $F_{1\mu}$, and J_μ of 0.88, 0.87, and 0.79 respectively, for the models trained with headlines only. Since it leaves little room for performance improvement, models trained with additional IC_Knwl with or without attention contribute an average of only 1.01 times more to the calculated scores.

Following that, we have the models analysed for Finnish with the next best averages of A, $F_{1\mu}$, and J_μ of 0.85, 0.84, and 0.79 respectively, for the models trained with headlines only. With the additional IC_Knwl, the average scores for A and $F_{1\mu}$ increase by 1.15 times and J_μ by 1.30 times. Nonetheless, the benefits of employing attention are negligible.

The impressive performance of the languages Czech and Finnish can be attributed to the fact that all of their samples

Table 4
Comparison between the baseline models and our proposed framework in terms of A, F_1, and J_m scores across all the reported languages. Trained with headlines and attended IC_Knwl using cmlm-ml, our framework outperforms the baseline models trained with headlines only.

	ml-MiniLM	distil-mUSE	ml-mpnet	LaBSE	cmlm-ml	ours
A	0.62	0.64	0.66	0.75	0.90	**0.92**
F_1	0.57	0.61	0.63	0.74	0.90	**0.92**
J_m	0.40	0.44	0.46	0.59	0.83	**0.86**

Table 5
A, F_1, and J_m scores of the analysed models for all the reported languages. Each model is trained using IC_Knwl, headlines with IC_Knwl (Headline+IC_Knwl), and headlines with attended IC_Knwl (Headline+Attn(IC_Knwl)) respectively.

	IC_Knwl			Headline+IC_Knwl			Headline+ Attn(IC_Knwl)		
	A	F_1	J_m	A	F_1	J_m	A	F_1	J_m
ml-MiniLM	0.78	0.78	0.64	0.81	0.81	0.68	0.86	0.87	0.77
distil-mUSE	0.78	0.78	0.64	0.83	0.83	0.71	0.90	0.90	0.83
ml-mpnet	0.81	0.81	0.69	0.83	0.84	0.72	0.89	0.89	0.81
LaBSE	0.86	0.87	0.77	0.89	0.90	0.82	0.90	0.91	0.83
cmlm-ml	0.86	0.88	0.79	0.91	0.92	0.85	0.92	0.92	0.86

Table 6
A, $F_{1\mu}$, and J_μ scores of the analysed models for each language used in the study. Each model is trained using headlines, headlines with IC_Knwl (Headline+IC_Knwl), and headlines with attended IC_Knwl (Headline+Attn(IC_Knwl)) respectively. For Headline+IC_Knwl, we report its relative performance to the models trained with headlines only. For Headline+Attn(IC_Knwl), we report its relative performance to the models trained with headlines and IC_Knwl.

		Headline			Headline+IC_Knwl			Headline+ Attn(IC_Knwl)		
		A	F_{1_μ}	J_μ	A	F_{1_μ}	J_μ	A	F_{1_μ}	J_μ
Slovenian	**ml-MiniLM**	0.53	0.53	0.36	1.05	1.03	1.05	1.67	1.16	1.23
	distil-mUSE	0.53	0.53	0.36	1.15	1.15	1.19	1.16	1.16	1.27
	ml-mpnet	0.55	0.54	0.37	1.05	1.07	1.10	1.12	1.10	1.14
	LaBSE	0.56	0.55	0.38	1.19	1.21	1.31	1.02	1.02	1.06
	cmlm-ml	0.54	0.70	0.71	1.01	1.01	1.01	1.02	1.04	1.05
Romanian	**ml-MiniLM**	0.47	0.47	0.31	1.87	1.87	2.54	1.06	1.05	1.11
	distil-mUSE	0.55	0.55	0.38	1.50	1.50	1.86	1.14	1.13	1.25
	ml-mpnet	0.56	0.56	0.38	1.60	1.58	2.13	1.05	1.05	1.09
	LaBSE	0.81	0.81	0.68	1.14	1.14	1.27	1.02	1.02	1.03
	cmlm-ml	0.95	0.94	0.89	1.00	1.01	1.01	1.01	1.00	1.01
Swedish	**ml-MiniLM**	0.52	0.51	0.34	1.71	1.74	2.35	1.08	1.08	1.17
	distil-mUSE	0.56	0.56	0.38	1.60	1.58	2.23	1.10	1.11	1.20
	ml-mpnet	0.58	0.58	0.41	1.58	1.58	2.07	1.07	1.07	1.15
	LaBSE	0.78	0.78	0.64	1.26	1.26	1.54	1.00	1.00	1.00
	cmlm-ml	0.98	0.98	0.96	1.01	1.01	1.03	1.00	1.00	1.00
Finnish	**ml-MiniLM**	0.81	0.81	0.68	1.17	1.17	1.33	1.00	1.00	1.00
	distil-mUSE	0.79	0.78	0.64	1.24	1.24	1.48	1.01	1.02	1.03
	ml-mpnet	0.82	0.82	0.69	1.18	1.18	1.36	1.02	1.02	1.04
	LaBSE	0.84	0.84	0.72	1.17	1.17	1.36	1.00	1.00	1.00
	cmlm-ml	0.99	0.99	0.98	1.00	1.00	1.01	1.00	1.00	1.00
Czech	**ml-MiniLM**	0.80	0.80	0.67	1.17	1.16	1.31	1.01	1.02	1.02
	distil-mUSE	0.87	0.86	0.76	1.10	1.11	1.21	1.02	1.02	1.04
	ml-mpnet	0.84	0.83	0.72	1.15	1.16	1.30	1.02	1.02	1.04
	LaBSE	0.92	0.91	0.84	1.07	1.08	1.17	1.00	1.00	1.00
	cmlm-ml	0.99	0.98	0.97	1.00	1.01	1.02	1.00	1.00	1.00

belong to the class 'Left-Centre.' Since all of their bias labels are from the same class, it is possible that the classifiers may end up modelling the language specifics and writing style of the outlet in addition to the bias embedded in the headlines.

The models evaluated for Swedish and Romanian produce the next-best results that are nearly identical to each other, differing only by a small margin. For Swedish, models trained with only headlines show an average $A/F_{1\mu}$ of 0.68 and J_μ of 0.54. The score differs by only 0.02 points for Romanian. IC_Knwl provides a substantial performance boost for both languages. Swedish and Romanian have $A/F_{1\mu}$ boosts of 1.43 and 1.42 times, and J_μ boosts of 1.84 and 1.76 times, respectively. They clearly benefit from the attention as well. Both the languages exhibit a 1.05 times boost in $A/F_{1\mu}$ and a 1.09 times boost in J_μ.

In the case of the models analysed for Slovenian, one can notice a significant performance gap when compared to others. It demonstrates the lowest performance with an average A, $F_{1\mu}$, and J_μ of 0.54, 0.57, and 0.43 respectively for the PLMs trained with headlines only. With the additional IC_Knwl, the average scores for $A/F_{1\mu}$ increase by 1.09 times and J_μ by 1.13 times. Moreover, the benefits of employing attention can be noticed by a performance increase of 1.19, 1.09, and 1.15 times in terms of A, $F_{1\mu}$, and J_μ scores. Even with the highest number of examples (30% ref. Table 1), its performance is low. To some extent, this could be attributed to the lower average headline length and language complexities that hinder the models' ability to comprehend the text for the task of bias prediction. Alternatively, it could be due to the limited embedding coverage of the Slovenian language.

Models trained with IC_Knwl indicate low A and high J_μ. This implies that with more true negatives than true positives, performance evaluation using A as the evaluation metric can turn out to be misleading. For instance, since there are no Slovenian examples that exist for the Right Centre class, considering instances where they are not present and are not predicted (true negatives), would not be credible. In such cases, considering J_μ is more reliable since it disregards true negatives in favour of true positives.

To sum up, across all languages analysed, cmlm-ml performed the best among models, whereas ml-MiniLM performed the worst. Overall, the results indicate the models trained with only headlines are capable of predicting the bias inherent in them, even for low-resource languages like the ones used in this study. Moreover, IC_Knwl significantly enhances model performance, especially when attention is employed.

5.3. Qualitative analysis

In this section, we assess the effect of translation quality on prediction performance by analysing translation errors. We use the Slovenian language as a case study since the models analysed for it exhibit a significant performance gap when compared to other languages in our dataset. With the help of native Slovenian speakers in our research group, we discovered several translation errors, which we classify as follows:

1. **Entity Detection Error:** occurs when the translation engine misinterprets the entities referenced in the headline.
2. **Comprehension Error:** arises when the translation engine fails to comprehend the meaning of a headline, resulting in an unintelligible translation.
3. **Improper Sentence Formation:** when the translated headline grasps the basic idea of the original headline, but fails to form a coherent translated sentence, this error type occurs.
4. **Inversion of Meaning:** takes place when the translation engine inverts the semantic meaning of a headline, resulting in a seemingly meaningful translation with a dissimilar semantic meaning.
5. **Miscellaneous Error:** a category reserved for errors that do not fit into any of the aforementioned categories.

Table 7 provides an example with appropriate justifications for each of these error types. In the majority of cases, the translation engine's lack of contextual awareness resulted in mistranslations. In some cases, the missing context could be inferred from the headline alone, whereas in others, reading the entire article or researching the entities mentioned in the headline appears to be the only way to obtain adequate context. Errors linked to a lack of vocabulary or other factors were less common.

Overall, the performance gap between Slovenian and other languages could be attributed to the language's poor translation quality relative to the other languages, as evidenced by the relatively numerous instances of improper translation. Given the complexity of the language and the small number of native speakers, the conclusion seems plausible.

6. Insights and implications

Predicting the political bias of news headlines has many positive implications. It can not only help readers identify politically biassed news but also allow journalists and the individuals involved in the news production process to assess their work objectively. Furthermore, such insights would also be interesting for researchers and social scientists. In this section, we further discuss the research implications and the ways in which our proposed framework can enhance practical applications.

6.1. Research implications

Our study proposes a new perspective by leveraging IC_Knwl via a Translate–Retrieve–Translate strategy to facilitate comprehension of the overall narrative of the multilingual headlines. Using IC_Knwl, it introduces a language-agnostic learning framework to enhance the prediction of political bias in multilingual news headlines. To the best of our knowledge, our proposed framework is one of the earliest attempts to leverage IC_Knwl in a multilingual context for bias prediction of news headlines. Since the existing work lacks annotated datasets for the task, it presents a dataset of multilingual news headlines. It simulates the real-world challenges of imbalanced data distribution by annotating headlines in five low-resource European languages with their respective political biases. Our experimental investigation demonstrates the advantages of using IC_Knwl, shedding light on the prospects of utilising it for downstream tasks. It also demonstrates the effectiveness of multiple state-of-the-art, multilingual, pre-trained language models.

6.2. Practical implications

Our study highlights the role of IC_Knwl in facilitating the comprehension of short news headline text. It demonstrates that the IC_Knwl when used in conjunction with the translate–retrieve–translate technique, can effectively aid in the comprehension of narratives in a multilingual context. When fused with multilingual PLMs, it enhances the political bias prediction of multilingual news headlines. Both implicit and explicit knowledge are expected in effective systems. The performance enhancement achieved by fusing the implicit knowledge obtained from the PLMs with explicit knowledge in the form of IC Knwl supports this view.

Given that a system is expected to deal with low-resource situations in the real world, our proposed framework is language-agnostic and thus adaptable to such scenarios. Another common problem in real-world scenarios is the scarcity of annotated data. Our proposed dataset, which focuses on low-resource languages with an imbalanced distribution, addresses this issue. Furthermore, our framework for data generation facilitates future expansion and the creation of custom datasets for related tasks.

7. Conclusions and future works

In this paper, we introduced a language-agnostic learning framework infused with IC_Knwl for enhancing the prediction of political bias in multilingual news headlines under an imbalanced sample distribution. We proposed to leverage IC_Knwl through a TRT strategy to help uncover contextual features for comprehension of the overall narrative of the multilingual headlines. Since not all the retrieved inferences are expected to be of equal relevance, we also employed an attention mechanism to emphasise relevant inferences. We used the neural-network model COMET trained on the ATOMIC$^{20}_{20}$ knowledge graphs to retrieve IC_Knwl and employed the Google Translate API for translation. Furthermore, we presented an annotated dataset of news headlines in five low-resource European languages.

We conducted an extensive evaluation of our framework with several multilingual PLMs. The evaluation results revealed the impressive performance of the multilingual PLMs, which can be attributed to their complex network architectures. The results also demonstrated that incorporating IC_Knwl and employing attention significantly enhanced their performance. Overall, the results indicate that the proposed framework for bias prediction is effective regardless of the models used. Even the models

Table 7
Case study of Slovenian headlines to understand the translation error types (translation: from Slovenian to English).

Entity Detection Error	
Slovenian Headline:	**Vodomec na 32. Liffu filmu Pohodi plin!**
Generated Translation:	Aquarius on the 32nd Liff movie Walk the Gas!
Correct Translation:	Kingfisher on the 32nd Liff awarded to movie Walk the Gas!
Comment:	The entity '***Vodomec***', which means '***Common Kingfisher***', is translated incorrectly as '***Aquarius***'. However, it refers to the name of an award in this context.
Comprehension Error	
Slovenian Headline:	**Počivalšek: Janša SMC ni ničesar prepustil**
Generated Translation:	Resting place: Janša SMC did not leave anything
Correct Translation:	Počivalšek: Janša left nothing for SMC
Comment:	The surname '***Počivalšek***' is mistranslated as '***Resting place***'. Furthermore, there exists no distinction between the surname '***Janša***' and the political party '***SMC***'.
Improper Sentence Formation	
Slovenian Headline:	**Nad zdravstvene delavce z grožnjami in žalitvami**
Generated Translation:	Above health professionals with threats and insults
Correct Translation:	Threats, insults towards health professionals
Comment:	Depending on the context, '***Nad***' could mean '***Above***' or '***Towards***'. The translation engine misinterprets '***Nad***' in this case, resulting in an improper sentence formation.
Inversion of Meaning	
Slovenian Headline:	**Na spletu podatki 533 milijonov Facebook uporabnikov, tudi 230.000 Slovencev**
Generated Translation:	There are 533 million Facebook users online, including 230,000 Slovenians
Correct Translation:	Data of 533 million Facebook users leaked online, including 230,000 Slovenians
Comment:	Although the translation is comprehensible, it refers to Facebook users instead of Facebook user data.
Miscellaneous Error	
Slovenian Headline:	**Grujović naj bi streljal v silobranu, priča trdi drugače**
Generated Translation:	Grujović allegedly shot in the silobran, the witness claims otherwise
Correct Translation:	Grujović allegedly shot in self-defence, the witness claims otherwise
Comment:	Since '***silobranu***' is misinterpreted as an entity, there is no attempt to translate '***v silobranu***', which means '***in self-defence***'.

evaluated for individual languages present plausible results. Furthermore, we conducted a thorough case study on the Slovenian headlines to investigate translation errors. The study uncovered numerous instances of improper translation, indicating that the performance gap between Slovenian and other languages may be attributable to the language's poor translation quality.

In the future, we plan to diversify our additional knowledge sources. In particular, we intend to investigate how knowledge sources such as Wiktionary and ConceptNet influence the task of bias prediction. Handling the explainability and interpretability of the model would be another challenge that we intend to address. Another possible direction is to extend this study beyond bias prediction to its quantification and correction. It would also be interesting to experiment with auxiliary tasks involving news headlines in a multitask learning paradigm.

CRediT authorship contribution statement

Swati Swati: Conceptualization, Data curation, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Adrian Mladenić Grobelnik:** Writing – review & editing, Investigation, Validation. **Dunja Mladenić:** Conceptualization, Supervision, Writing – review & editing. **Marko Grobelnik:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Dataset and scripts available at: https://github.com/Swati17293/KG-Multi-Bias.

Acknowledgements

This work was supported by the Slovenian Research Agency under the project J2-1736 Causalify and the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 812997 and enRichMyData European Union's Horizon Europe project under Grant Agreement 101070284 (http://dx.doi.org/10.3030/101070284).

References

[1] N. Helberger, On the democratic role of news recommenders, Digital J. 7 (8) (2019) 993–1012, http://dx.doi.org/10.1080/21670811.2019.1623700.

[2] B. McNair, Journalism and democracy, in: The Handbook of Journalism Studies, Routledge, 2009, pp. 257–269, http://dx.doi.org/10.4324/9780203877685-27.

[3] J.M. Miller, J.A. Krosnick, News media impact on the ingredients of presidential evaluations: Politically knowledgeable citizens are guided by a trusted source, Am. J. Polit. Sci. (2000) 301–315, http://dx.doi.org/10.2307/2669312.

[4] S. Park, S. Kang, S. Chung, J. Song, NewsCube: delivering multiple aspects of news to mitigate media bias, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, 2009, pp. 443–452, http://dx.doi.org/10.1145/1518701.1518772.

[5] S.R. Davis, C.J. Worsnop, E.M. Hand, Gender bias recognition in political news articles, Mach. Learn. Appl. 8 (2022) 100304, http://dx.doi.org/10.1016/j.mlwa.2022.100304.

[6] T. Spinde, An interdisciplinary approach for the automated detection and visualization of media bias in news articles, in: 2021 International Conference on Data Mining Workshops, ICDMW, IEEE, 2021, pp. 1096–1103, http://dx.doi.org/10.1109/ICDMW53433.2021.00144.
[7] T. Spinde, L. Rudnitckaia, J. Mitrović, F. Hamborg, M. Granitzer, B. Gipp, K. Donnay, Automated identification of bias inducing words in news articles using linguistic and context-oriented features, Inf. Process. Manage. 58 (3) (2021) 102505, http://dx.doi.org/10.1016/j.ipm.2021.102505.
[8] K. Chen, M. Babaeianjelodar, Y. Shi, K. Janmohamed, R. Sarkar, I. Weber, T. Davidson, M. De Choudhury, S. Yadav, A. Khudabukhsh, et al., Partisan US news media representations of Syrian refugees, 2022, http://dx.doi.org/10.48550/arXiv.2206.09024, arXiv preprint arXiv:2206.09024.
[9] W. Chipidza, The effect of toxicity on COVID-19 news network formation in political subcommunities on Reddit: An affiliation network approach, Int. J. Inf. Manage. 61 (2021) 102397, http://dx.doi.org/10.1016/j.ijinfomgt.2021.102397.
[10] F. Hamborg, K. Donnay, B. Gipp, Automated identification of media bias in news articles: an interdisciplinary literature review, Int. J. Digit. Libraries 20 (4) (2019) 391–415, http://dx.doi.org/10.1007/s00799-018-0261-y.
[11] R.R.R. Gangula, S.R. Duggenpudi, R. Mamidi, Detecting political bias in news articles using headline attention, in: Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, Association for Computational Linguistics, 2019, pp. 77–84, http://dx.doi.org/10.18653/v1/W19-4809.
[12] P. Laban, L. Bandarkar, M.A. Hearst, News headline grouping as a challenging NLU task, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, 2021, pp. 3186–3198, http://dx.doi.org/10.18653/v1/2021.naacl-main.255.
[13] K. Holmqvist, J. Holsanova, M. Barthelson, D. Lundqvist, Reading or scanning? A study of newspaper and net paper reading, in: The Mind's Eye, Elsevier, 2003, pp. 657–670, http://dx.doi.org/10.1016/B978-044451020-4/50035-9.
[14] B.C. Andrew, Media-generated shortcuts: Do newspaper headlines present another roadblock for low-information rationality? Harvard Int. J. Press/Politics 12 (2) (2007) 24–43, http://dx.doi.org/10.1177/1081180X07299795.
[15] U.K. Ecker, S. Lewandowsky, E.P. Chang, R. Pillai, The effects of subtle misinformation in news headlines, J. Experim. Psychol. Appl. 20 (4) (2014) 323, http://dx.doi.org/10.1037/xap0000028.
[16] K. Molek-Kozakowska, Towards a pragma-linguistic framework for the study of sensationalism in news headlines, Discourse Commun. 7 (2) (2013) 173–197, http://dx.doi.org/10.1177/1750481312471668.
[17] E. Ifantidou, Newspaper headlines and relevance: Ad hoc concepts in ad hoc contexts, J. Pragmat. 41 (4) (2009) 699–720, http://dx.doi.org/10.1016/j.pragma.2008.10.016.
[18] M. McCluskey, A content analysis of 2004 presidential election headlines of the los angeles times and the washington times, Electronic Theses Dissertations 358 (2005) https://stars.library.ucf.edu/etd/358.
[19] S.M. Jovanović, et al., Headlines against democracy: Operational code analysis of the Serbian daily informer's headlines in relation to the anti-government protests' first phase (2018–2019), J. Med. Re.-Rev. Stud. Med. 14 (41) (2021) 23–41, https://www.ceeol.com/search/article-detail?id=1000890.
[20] D. Zeng, W. Gong, S. Li, Critical discourse analysis on the news headline about terrorism: A case study of the english reports on counter-terrorism from Turkey, Modern Linguist. 6 (3) (2018) 496–501, http://dx.doi.org/10.12677/ml.2018.63057.
[21] B. Andrew, Political journalism represented by headline news: Canadian public and commercial media compared, Can. J. Polit. Sci./Rev. Can. Sci. Polit. 46 (2) (2013) 455–478, http://dx.doi.org/10.1017/S0008423913000462.
[22] P. Navia, R. Osorio, El mercurio lies, and la tercera lies more. Political bias in newspaper headlines in Chile, 1994–2010, Bullet. Latin Am. Res. 34 (4) (2015) 467–485, http://dx.doi.org/10.1111/blar.12364.
[23] F. Hamborg, N. Meuschke, A. Aizawa, B. Gipp, Identification and analysis of media bias in news articles, in: Everything Changes, Everything Stays the Same? Understanding Information Spaces. Proceedings of the 15th International Symposium of Information Science, ISI 2017, 2017, pp. 224–236, http://dx.doi.org/10.18452/1446.
[24] F. Hamborg, N. Meuschke, B. Gipp, Bias-aware news analysis using matrix-based news aggregation, Int. J. Digit. Libraries 21 (2) (2020) 129–147, http://dx.doi.org/10.1007/s00799-018-0239-9.
[25] D. Aksenov, P. Bourgonje, K. Zaczynska, M. Ostendorff, J.M. Schneider, G. Rehm, Fine-grained classification of political bias in german news: A data set and initial experiments, in: Proceedings of the 5th Workshop on Online Abuse and Harms, WOAH 2021, 2021, pp. 121–131, http://dx.doi.org/10.18653/v1/2021.woah-1.13.
[26] S. Guo, K.Q. Zhu, Modeling multi-level context for informational bias detection by contrastive learning and sentential graph network, 2022, http://dx.doi.org/10.48550/arXiv.2201.10376, arXiv preprint arXiv:2201.10376.
[27] T.M. Doan, J.A. Gulla, A survey on political viewpoints identification, Online Soc. Netw. Media 30 (2022) 100208, http://dx.doi.org/10.1016/j.osnem.2022.100208.
[28] S. Park, S. Kang, S. Chung, J. Song, A computational framework for media bias mitigation, ACM Trans. Interact. Intell. Syst. (TiiS) 2 (2) (2012) 1–32, http://dx.doi.org/10.1145/2209310.2209311.
[29] R. Del Gratta, F. Frontini, A.F. Khan, J. Mariani, C. Soria, The lremap for under-resourced languages, in: Workshop Collaboration and Computing for under-Resourced Languages in the Linked Open Data Era, Satellite Workshop of LREC, vol. 14, 2014, https://www.academia.edu/download/39974165/The_LREMap_for_Under-Resourced_Languages20151113-28056-rmmjbf.pdf.
[30] E.G. Bruneau, M. Cikara, R. Saxe, Going beyond the headlines: Narratives mitigate intergroup empathy bias, in: Proceedings of the 34th Annual Meeting of the Cognitive Science Society, CogSci 2012, cognitivesciencesociety.org, 2012, https://mindmodeling.org/cogsci2012/papers/0475/index.html.
[31] R.T. Berner, Commentary: The narrative and the headline, Newspap. Res. J. 4 (3) (1983) 33–40, http://dx.doi.org/10.1177/073953298300400305.
[32] D. Li, X. Zhu, Y. Li, S. Wang, D. Li, J. Liao, J. Zheng, Enhancing emotion inference in conversations with commonsense knowledge, Knowl.-Based Syst. 232 (2021) 107449, http://dx.doi.org/10.1016/j.knosys.2021.107449.
[33] J. Li, Z. Lin, P. Fu, W. Wang, Past, present, and future: Conversational emotion recognition through structural modeling of psychological knowledge, in: Findings of the Association for Computational Linguistics: EMNLP 2021, Association for Computational Linguistics, 2021, pp. 1204–1214, http://dx.doi.org/10.18653/v1/2021.findings-emnlp.104.
[34] L. Du, X. Ding, K. Xiong, T. Liu, B. Qin, Enhancing pretrained language models with structured commonsense knowledge for textual inference, Knowl.-Based Syst. 254 (2022) 109488, http://dx.doi.org/10.1016/j.knosys.2022.109488.
[35] R. Li, Z. Jiang, L. Wang, X. Lu, M. Zhao, D. Chen, Enhancing transformer-based language models with commonsense representations for knowledge-driven machine comprehension, Knowl.-Based Syst. 220 (2021) 106936, http://dx.doi.org/10.1016/j.knosys.2021.106936.
[36] A. Lieto, G.L. Pozzato, S. Zoia, V. Patti, R. Damiano, A commonsense reasoning framework for explanatory emotion attribution, generation and re-classification, Knowl.-Based Syst. 227 (2021) 107166, http://dx.doi.org/10.1016/j.knosys.2021.107166.
[37] S. Swati, M. Grobelnik, Ic-Bait: An inferential commonsense-driven model for predicting political polarity in news headlines, 2022, http://dx.doi.org/10.2139/ssrn.4114271, Available at SSRN 4114271.
[38] J.D. Hwang, C. Bhagavatula, R. Le Bras, J. Da, K. Sakaguchi, A. Bosselut, Y. Choi, (Comet-) atomic 2020: On symbolic and neural commonsense knowledge graphs, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, 2021, pp. 6384–6392, https://ojs.aaai.org/index.php/AAAI/article/view/16792.
[39] Y. Fang, S. Wang, Y. Xu, R. Xu, S. Sun, C. Zhu, M. Zeng, Leveraging knowledge in multilingual commonsense reasoning, in: Findings of the Association for Computational Linguistics: ACL 2022, Association for Computational Linguistics, 2022, pp. 3237–3246, http://dx.doi.org/10.18653/v1/2022.findings-acl.255.
[40] A. Bonyadi, M. Samuel, Headlines in newspaper editorials: A contrastive study, Sage Open 3 (2) (2013) 2158244013494863, http://dx.doi.org/10.1177/2158244013494863.
[41] MBFC, Finnish news - media bias/fact check, 2022, https://mediabiasfactcheck.com/finnish-news/. (Accessed 04 July 2022).
[42] M. Li, H. Zhou, J. Hou, P. Wang, E. Gao, Is cross-linguistic advert flaw detection in Wikipedia feasible? A multilingual-BERT-based transfer learning approach, Knowl.-Based Syst. 252 (2022) 109330, http://dx.doi.org/10.1016/j.knosys.2022.109330.
[43] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, G. Zhang, Transfer learning using computational intelligence: A survey, Knowl.-Based Syst. 80 (2015) 14–23, http://dx.doi.org/10.1016/j.knosys.2015.01.010.
[44] E.W. Pamungkas, V. Basile, V. Patti, A joint learning approach with knowledge injection for zero-shot cross-lingual hate speech detection, Inf. Process. Manage. 58 (4) (2021) 102544, http://dx.doi.org/10.1016/j.ipm.2021.102544.
[45] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, W. Wang, Language-agnostic bert sentence embedding, 2020, http://dx.doi.org/10.48550/arXiv.2007.01852, arXiv preprint arXiv:2007.01852.
[46] N. Reimers, I. Gurevych, Sentence-BERT: Sentence embeddings using siamese BERT-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 3982–3992, http://dx.doi.org/10.18653/v1/D19-1410.

[47] Z. Yang, Y. Yang, D. Cer, J. Law, E. Darve, Universal sentence representation learning with conditional masked language model, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021, pp. 6216–6228, http://dx.doi.org/10.18653/v1/2021.emnlp-main.502.

[48] A. Kruspe, M. Häberle, I. Kuhn, X.X. Zhu, Cross-language sentiment analysis of European Twitter messages during the COVID-19 pandemic, in: Proceedings of the 1st Workshop on NLP for COVID-19 At ACL 2020, Association for Computational Linguistics, 2020, https://aclanthology.org/2020.nlpcovid19-acl.14.

[49] Y. Pei, S. Chen, Z. Ke, W. Silamu, Q. Guo, AB-LaBSE: Uyghur sentiment analysis via the pre-training model with BiLSTM, Appl. Sci. 12 (3) (2022) 1182, http://dx.doi.org/10.3390/app12031182.

[50] R.N. Patel, E. Burgin, H. Assem, S. Dutta, Efficient multi-lingual sentence classification framework with sentence meta encoders, in: 2021 IEEE International Conference on Big Data (Big Data), IEEE, 2021, pp. 1889–1899, http://dx.doi.org/10.1109/BigData52589.2021.9671714.

[51] Z. Talat, A. Neveol, S. Biderman, M. Clinciu, M. Dey, S. Longpre, S. Luccioni, M. Masoud, M. Mitchell, D. Radev, et al., You reap what you sow: On the challenges of bias evaluation under multilingual settings, in: Proceedings of BigScience Episode# 5–Workshop on Challenges & Perspectives in Creating Large Language Models, Association for Computational Linguistics, 2022, pp. 26–41, http://dx.doi.org/10.18653/v1/2022.bigscience-1.3.

[52] A. Roy, K. Basak, A. Ekbal, P. Bhattacharyya, A deep ensemble framework for fake news detection and multi-class classification of short political statements, in: Proceedings of the 16th International Conference on Natural Language Processing, 2019, pp. 9–17, https://aclanthology.org/2019.icon-1.2/.

[53] T. Saikh, A. Anand, A. Ekbal, P. Bhattacharyya, A novel approach towards fake news detection: deep learning augmented with textual entailment features, in: International Conference on Applications of Natural Language To Information Systems, Springer, 2019, pp. 345–358, http://dx.doi.org/10.1007/978-3-030-23281-8_30.

[54] T. Saikh, A. De, A. Ekbal, P. Bhattacharyya, A deep learning approach for automatic detection of fake news, in: Proceedings of the 16th International Conference on Natural Language Processing, 2019, pp. 230–238, https://aclanthology.org/2019.icon-1.27.

[55] K.K. King, B. Wang, Diffusion of real versus misinformation during a crisis event: a big data-driven approach, Int. J. Inf. Manage. (2021) 102390, http://dx.doi.org/10.1016/j.ijinfomgt.2021.102390.

[56] L. Rotim, M. Tutek, J. Šnajder, Takelab at semeval-2017 task 5: Linear aggregation of word embeddings for fine-grained sentiment analysis of financial news, in: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), 2017, pp. 866–871, http://dx.doi.org/10.18653/v1/S17-2148.

[57] D. Korenčić, S. Ristov, J. Šnajder, Document-based topic coherence measures for news media text, Expert systems with Applications 114 (2018) 357–373, http://dx.doi.org/10.1016/j.eswa.2018.07.063.

[58] M.B. Pandur, J. Dobša, S. Beliga, A. Meštrović, Topic modelling and sentiment analysis of COVID-19 related news on Croatian internet portal, Inform. Soc. 2020 (2020) 5–9.

[59] E. Müller-Budack, J. Theiner, S. Diering, M. Idahl, S. Hakimov, R. Ewerth, Multimodal news analytics using measures of cross-modal entity and context consistency, Int. J. Multimed. Inform. Retrieval 10 (2) (2021) 111–125, http://dx.doi.org/10.1007/s13735-021-00207-4.

[60] G. Tahmasebzadeh, S. Hakimov, E. Müller-Budack, R. Ewerth, A feature analysis for multimodal news retrieval, in: Proceedings of the 1st International Workshop on Cross-Lingual Event-Centric Open Analytics Co-Located with the 17th Extended Semantic Web Conference (ESWC 2020), Aachen: RWTH, 2020, http://dx.doi.org/10.34657/5197.

[61] D. D'Alessio, Media Bias in Presidential Election Coverage, 1948-2008: Evaluation Via Formal Measurement, Lexington Books, 2012.

[62] N. Palić, J. Vladika, D. Čubelić, I. Lovrenčić, M. Buljan, J. Šnajder, TakeLab at SemEval-2019 task 4: Hyperpartisan news detection, in: Proceedings of the 13th International Workshop on Semantic Evaluation, 2019, pp. 995–998, http://dx.doi.org/10.18653/v1/S19-2172.

[63] R.L. Stevenson, M.T. Greene, A reconsideration of bias in the news, J. Quart. 57 (1) (1980) 115–121, http://dx.doi.org/10.1177/107769908005700117.

[64] W.-F. Chen, H. Wachsmuth, K. Al Khatib, B. Stein, Learning to flip the bias of news headlines, in: Proceedings of the 11th International Conference on Natural Language Generation, Association for Computational Linguistics, 2018, pp. 79–88, http://dx.doi.org/10.18653/v1/W18-6509.

[65] T. Groseclose, J. Milyo, A measure of media bias, Q. J. Econ. 120 (4) (2005) 1191–1237, http://dx.doi.org/10.1162/003355305775097542.

[66] M. Iyyer, P. Enns, J. Boyd-Graber, P. Resnik, Political ideology detection using recursive neural networks, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2014, pp. 1113–1122, https://aclanthology.org/P14-1105.pdf.

[67] N.R. Naredla, F.F. Adedoyin, Detection of hyperpartisan news articles using natural language processing technique, I. J. Inform. Manag. Data Insights 2 (1) (2022) 100064, http://dx.doi.org/10.1016/j.jjimei.2022.100064.

[68] I. Tourni, L. Guo, T.H. Daryanto, F. Zhafransyah, E.E. Halim, M. Jalal, B. Chen, S. Lai, H. Hu, M. Betke, et al., Detecting frames in news headlines and lead images in US gun violence coverage, in: Findings of the Association for Computational Linguistics: EMNLP 2021, 2021, pp. 4037–4050, http://dx.doi.org/10.18653/v1/2021.findings-emnlp.339.

[69] J.-D. Krieger, T. Spinde, T. Ruas, J. Kulshrestha, B. Gipp, A domain-adaptive pre-training approach for language bias detection in news, in: Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries, 2022, pp. 1–7, http://dx.doi.org/10.1145/3529372.3530932.

[70] V. Magotra, E. Hirani, V. Mehta, S. Dholay, News bias detection using transformers, in: Communication and Intelligent Systems, Springer, 2022, pp. 319–326, http://dx.doi.org/10.1007/978-981-19-2130-8_26.

[71] A. Hoyer, Spanish news framing of the Syrian refugee crisis, WWU Honors Program Senior Projects 26 (2016) https://cedar.wwu.edu/wwu_honors/26/.

[72] L. Fan, M. White, E. Sharma, R. Su, P.K. Choubey, R. Huang, L. Wang, In plain sight: Media bias through the lens of factual reporting, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 6343–6349, http://dx.doi.org/10.18653/v1/D19-1664.

[73] R. Baly, G. Da San Martino, J. Glass, P. Nakov, We can detect your bias: Predicting the political ideology of news articles, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP, Association for Computational Linguistics, 2020, pp. 4982–4991, http://dx.doi.org/10.18653/v1/2020.emnlp-main.404.

[74] F. Petroni, T. Rocktäschel, S. Riedel, P. Lewis, A. Bakhtin, Y. Wu, A. Miller, Language models as knowledge bases? in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, 2019, pp. 2463–2473, http://dx.doi.org/10.18653/v1/D19-1250.

[75] V. Shwartz, P. West, R. Le Bras, C. Bhagavatula, Y. Choi, Unsupervised commonsense question answering with self-talk, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP, Association for Computational Linguistics, 2020, pp. 4615–4629, http://dx.doi.org/10.18653/v1/2020.emnlp-main.373.

[76] N. Do, E. Pavlick, Are rotten apples edible? Challenging commonsense inference ability with exceptions, in: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Association for Computational Linguistics, 2021, pp. 2061–2073, https://aclanthology.org/2021.findings-acl.181.pdf.

[77] N. Kassner, H. Schütze, Negated and misprimed probes for pretrained language models: Birds can talk, but cannot fly, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, 2020, pp. 7811–7818, https://aclanthology.org/2020.acl-main.698.pdf?ref=https://githubhelp.com.

[78] Q. Tu, Y. Li, J. Cui, B. Wang, J.-R. Wen, R. Yan, MISC: A mixed strategy-aware model integrating COMET for emotional support conversation, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022, pp. 308–319, http://dx.doi.org/10.18653/v1/2022.acl-long.25.

[79] T. Young, E. Cambria, I. Chaturvedi, H. Zhou, S. Biswas, M. Huang, Augmenting end-to-end dialogue systems with commonsense knowledge, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, 2018, https://ojs.aaai.org/index.php/AAAI/article/view/11923.

[80] H. Zhou, T. Young, M. Huang, H. Zhao, J. Xu, X. Zhu, Commonsense knowledge aware conversation generation with graph attention., in: IJCAI, 2018, pp. 4623–4629, https://www.ijcai.org/Proceedings/2018/0643.pdf.

[81] T. Mihaylov, A. Frank, Knowledgeable reader: Enhancing cloze-style reading comprehension with external commonsense knowledge, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018, pp. 821–832, http://dx.doi.org/10.18653/v1/P18-1076.

[82] Y.K. Lal, H. Liu, N. Tandon, N. Chambers, R. Mooney, N. Balasubramanian, Analyzing the contribution of commonsense knowledge sources for why-question answering, in: ACL 2022 Workshop on Commonsense Representation and Reasoning, 2022, https://openreview.net/pdf?id=H4xz8zteub9.

[83] J. Chen, J. Chen, Z. Yu, Incorporating structured commonsense knowledge in story completion, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019, pp. 6244–6251, https://ojs.aaai.org/index.php/AAAI/article/view/5183.

[84] Q.-F. Wang, E. Cambria, C.-L. Liu, A. Hussain, Common sense knowledge for handwritten chinese text recognition, Cogn. Comput. 5 (2) (2013) 234–242, http://dx.doi.org/10.1007/s12559-012-9183-y.

[85] P. Zhong, D. Wang, P. Li, C. Zhang, H. Wang, C. Miao, Care: Commonsense-aware emotional response generation with latent concepts, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, 2021, pp. 14577–14585, https://ojs.aaai.org/index.php/AAAI/article/view/17713.

[86] L. Zhu, G. Pergola, L. Gui, D. Zhou, Y. He, Topic-driven and knowledge-aware transformer for dialogue emotion detection, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021, pp. 1571–1582, http://dx.doi.org/10.18653/v1/2021.acl-long.125.

[87] R. Speer, J. Chin, C. Havasi, ConceptNet 5.5: An open multilingual graph of general knowledge, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31, 2017, https://dl.acm.org/doi/abs/10.5555/3298023.3298212.

[88] E. Cambria, S. Poria, D. Hazarika, K. Kwok, SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, 2018, https://ojs.aaai.org/index.php/AAAI/article/view/11559.

[89] N. Mostafazadeh, A. Kalyanpur, L. Moon, D. Buchanan, L. Berkowitz, O. Biran, J. Chu-Carroll, GLUCOSE: GeneraLized and contextualized story explanations, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP, Association for Computational Linguistics, 2020, pp. 4569–4586, http://dx.doi.org/10.18653/v1/2020.emnlp-main.370.

[90] H. Rashkin, M. Sap, E. Allaway, N.A. Smith, Y. Choi, Event2Mind: Commonsense inference on events, intents, and reactions, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, 2018, pp. 463–473, http://dx.doi.org/10.18653/v1/P18-1043.

[91] J. Romero, S. Razniewski, Inside quasimodo: Exploring construction and usage of commonsense knowledge, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020, pp. 3445–3448, http://dx.doi.org/10.1145/3340531.3417416.

[92] N. Tandon, G. De Melo, G. Weikum, Webchild 2.0: Fine-grained commonsense knowledge distillation, in: Proceedings of ACL 2017, System Demonstrations, Association for Computational Linguistics, 2017, pp. 115–120, http://dx.doi.org/10.18653/v1/P17-4020.

[93] S. Gabriel, C. Bhagavatula, V. Shwartz, R. Le Bras, M. Forbes, Y. Choi, Paragraph-level commonsense transformers with recurrent memory, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, 2021, pp. 12857–12865, https://ojs.aaai.org/index.php/AAAI/article/view/17521.

[94] J.D.M.-W.C. Kenton, L.K. Toutanova, 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, vol. 1, Association for Computational Linguistics, 2019, pp. 4171–4186, http://dx.doi.org/10.18653/v1/N19-1423.

[95] K. Clark, M.-T. Luong, Q.V. Le, C.D. Manning, Electra: Pre-training text encoders as discriminators rather than generators, 2020, http://dx.doi.org/10.48550/arXiv.2003.10555, arXiv preprint arXiv:2003.10555.

[96] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R.R. Salakhutdinov, Q.V. Le, Xlnet: Generalized autoregressive pretraining for language understanding, Adv. Neural Inform. Process. Syst. 32 (2019) http://papers.neurips.cc/paper/8812-xlnet-generalized-autoregressive-pretraining-for-language-understanding.pdf.

[97] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, R. Soricut, Albert: A lite bert for self-supervised learning of language representations, 2019, http://dx.doi.org/10.48550/arXiv.1909.11942, arXiv preprint arXiv:1909.11942.

[98] M.E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep contextualized word representations, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,(Long Papers), vol. 1, Association for Computational Linguistics, 2018, pp. 2227–2237, http://dx.doi.org/10.18653/v1/N18-1202.

[99] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inform. Process. Syst. 30 (2017) 6000–6010, https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[100] T. Lin, Y. Wang, X. Liu, X. Qiu, A survey of transformers, 2021, http://dx.doi.org/10.48550/arXiv.2106.04554, arXiv preprint arXiv:2106.04554.

[101] B. Gain, D. Bandyopadhyay, T. Saikh, A. Ekbal, Iitp@ COLIEE 2019: legal information retrieval using BM25 and BERT, 2021, http://dx.doi.org/10.48550/arXiv.2104.08653, arXiv preprint arXiv:2104.08653.

[102] K. Mishra, M. Firdaus, A. Ekbal, Please be polite: Towards building a politeness adaptive dialogue system for goal-oriented conversations, Neurocomputing 494 (2022) 242–254, http://dx.doi.org/10.1016/j.neucom.2022.04.029.

[103] S. Yadav, M. Sarrouti, D. Gupta, NLM at MEDIQA 2021: Transfer learning-based approaches for consumer question and multi-answer summarization, in: Proceedings of the 20th Workshop on Biomedical Language Processing, 2021, pp. 291–301, http://dx.doi.org/10.18653/v1/2021.bionlp-1.34.

[104] S. Yadav, D. Gupta, A.B. Abacha, D. Demner-Fushman, Question-aware transformer models for consumer health question summarization, J. Biomed. Inform. 128 (2022) 104040, http://dx.doi.org/10.1016/j.jbi.2022.104040.

[105] S. Pingali, S. Yadav, P. Dutta, S. Saha, Multimodal graph-based transformer framework for biomedical relation extraction, in: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, 2021, pp. 3741–3747, https://aclanthology.org/2021.findings-acl.328.pdf.

[106] J. Shin, Y. Lee, K. Jung, Effective sentence scoring method using bert for speech recognition, in: Asian Conference on Machine Learning, PMLR, 2019, pp. 1081–1093, https://proceedings.mlr.press/v101/shin19a.html.

[107] G.V. Singh, M. Firdaus, A. Ekbal, P. Bhattacharyya, Unity in diversity: Multilabel emoji identification in tweets, IEEE Trans. Comput. Soc. Syst. (2022) http://dx.doi.org/10.1109/TCSS.2022.3162865.

[108] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, É. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, 2020, pp. 8440–8451, http://dx.doi.org/10.18653/v1/2020.acl-main.747.

[109] Y. Liu, J. Gu, N. Goyal, X. Li, S. Edunov, M. Ghazvininejad, M. Lewis, L. Zettlemoyer, Multilingual denoising pre-training for neural machine translation, Trans. Assoc. Comput. Linguist. 8 (2020) 726–742, http://dx.doi.org/10.1162/tacl_a_00343, URL https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00343/96484/Multilingual-Denoising-Pre-training-for-Neural.

[110] A. Kumar, V.H.C. Albuquerque, Sentiment analysis using XLM-r transformer and zero-shot transfer learning on resource-poor Indian language, Trans. Asian Low-Resour. Language Inform. Process. 20 (5) (2021) 1–13, http://dx.doi.org/10.1145/3461764.

[111] E. Novak, L. Bizjak, D. Mladenić, M. Grobelnik, Why is a document relevant? Understanding the relevance scores in cross-lingual document retrieval, Knowl.-Based Syst. 244 (2022) 108545, http://dx.doi.org/10.1016/j.knosys.2022.108545.

[112] B. Muller, Y. Elazar, B. Sagot, D. Seddah, First align, then predict: Understanding the cross-lingual ability of multilingual BERT, in: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Association for Computational Linguistics, 2021, pp. 2214–2231, http://dx.doi.org/10.18653/v1/2021.eacl-main.189.

[113] D. Cer, Y. Yang, S.-y. Kong, N. Hua, N. Limtiaco, R.S. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar, et al., Universal sentence encoder, 2018, http://dx.doi.org/10.48550/arXiv.1803.11175, arXiv preprint arXiv:1803.11175.

[114] N. Reimers, I. Gurevych, Making monolingual sentence embeddings multilingual using knowledge distillation, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP, Association for Computational Linguistics, 2020, pp. 4512–4525, http://dx.doi.org/10.18653/v1/2020.emnlp-main.365.

[115] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, W. Wang, Language-agnostic BERT sentence embedding, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, 2022, pp. 878–891, http://dx.doi.org/10.18653/v1/2022.acl-long.62.

[116] V. Chaudhary, Y. Tang, F. Guzmán, H. Schwenk, P. Koehn, Low-resource corpus filtering using multilingual sentence embeddings, in: Proceedings of the Fourth Conference on Machine Translation (Volume 3: Shared Task Papers, Day 2), Association for Computational Linguistics, 2019, pp. 261–266, http://dx.doi.org/10.18653/v1/W19-5435.

[117] A.-S. Mohammad, M.M. Hammad, A. Sa'ad, A.-T. Saja, E. Cambria, Gated recurrent unit with multilingual universal sentence encoder for arabic aspect-based sentiment analysis, Knowl.-Based Syst. (2021) 107540, http://dx.doi.org/10.1016/j.knosys.2021.107540.

[118] A. Conneau, D. Kiela, H. Schwenk, L. Barrault, A. Bordes, Supervised learning of universal sentence representations from natural language inference data, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2017, pp. 670–680, http://dx.doi.org/10.18653/v1/D17-1070.

[119] MBFC, Media bias/fact check - search and learn the bias of news media, 2022, https://mediabiasfactcheck.com. (Accessed 06 June 2022).

[120] R. Baly, G. Karadzhov, D. Alexandrov, J. Glass, P. Nakov, Predicting factuality of reporting and bias of news media sources, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2018, pp. 3528–3539, http://dx.doi.org/10.18653/v1/D18-1389.

[121] P. Resnick, A. Ovadya, G. Gilchrist, Iffy quotient: A platform health metric for misinformation, Center Soc. Med. Responsibil. 17 (2018) http://umsi.info/iffy-quotient-whitepaper.
[122] MBFC, Left vs. Right bias: How we rate the bias of media sources - media bias/fact check, 2022, https://mediabiasfactcheck.com/left-vs-right-bias-how-we-rate-the-bias-of-media-sources/. (Accessed 06 June 2022).
[123] MBFC, Methodology - media bias/fact check, 2022, https://mediabiasfactcheck.com/methodology/ (Accessed 06 June 2022).
[124] G. Leban, B. Fortuna, J. Brank, M. Grobelnik, Event registry: learning about world events from news, in: Proceedings of the 23rd International Conference on World Wide Web, 2014, pp. 107–110, http://dx.doi.org/10.1145/2567948.257702.
[125] S. Swati, D. Mladenić, T. Erjavec, EveOut: an event-centric news dataset to analyze an outlet's event selection patterns, Informatica 45 (7) (2021) http://dx.doi.org/10.31449/inf.v45i7.3410.
[126] S. Swati, D. Mladenić, Are you following the right news-outlet? A machine learning based approach to outlet prediction, in: In Proceedings of the Slovenian KDD Conference on Data Mining and Data Warehouses (SiKDD), 2020, https://ailab.ijs.si/Dunja/SiKDD2020/Papers/08-swati_outlet_prediction.pdf.
[127] S. Swati, D. Mladenić, Understanding the impact of geographical bias on news sentiment: A case study on London and rio olympics, in: In Proceedings of the Slovenian KDD Conference on Data Mining and Data Warehouses (SiKDD), 2021, https://ailab.ijs.si/dunja/SiKDD2021/Papers/Swati+Mladenic.pdf.
[128] A. Bosselut, H. Rashkin, M. Sap, C. Malaviya, A. Celikyilmaz, Y. Choi, COMET: Commonsense transformers for automatic knowledge graph construction, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, 2019, pp. 4762–4779, http://dx.doi.org/10.18653/v1/P19-1470.
[129] S. Sabour, C. Zheng, M. Huang, Cem: Commonsense-aware empathetic response generation, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, no. 10, 2022, pp. 11229–11237, http://dx.doi.org/10.1609/aaai.v36i10.21373.
[130] Y. Liu, J. Gao, J. Du, L. Zhou, R. Xu, Empathetic response generation with state management, 2022, http://dx.doi.org/10.48550/arXiv.2205.03676, arXiv preprint arXiv:2205.03676.
[131] N. Majumder, P. Hong, S. Peng, J. Lu, D. Ghosal, A. Gelbukh, R. Mihalcea, S. Poria, MIME: Mimicking emotions for empathetic response generation, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP, Association for Computational Linguistics, 2020, pp. 8968–8979, http://dx.doi.org/10.18653/v1/2020.emnlp-main.721.
[132] K. Nantomah, On some properties of the sigmoid function, Asia Math. (2019) https://hal.archives-ouvertes.fr/hal-02635089/.
[133] N. Balakrishnan, Handbook of the Logistic Distribution, CRC Press, 2013.
[134] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: International Conference on Learning Representations, ICLR, 2015, pp. 1–13, https://hdl.handle.net/11245/1.505367.
[135] U.N. Sayar Ghosh Roy, T. Raha, Z. Abid, V. Varma, Leveraging multilingual transformers for hate speech detection, in: CEUR Workshop Proceedings: FIRE 2020 - Forum for Information Retrieval Evaluation, 2020, http://ceur-ws.org/Vol-2826/T2-4.pdf.
[136] Y. Yang, D. Cer, A. Ahmad, M. Guo, J. Law, N. Constant, G.H. Abrego, S. Yuan, C. Tar, Y.-H. Sung, et al., Multilingual universal sentence encoder for semantic retrieval, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 2020, pp. 87–94, http://dx.doi.org/10.18653/v1/2020.acl-demos.12.
[137] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2018, http://dx.doi.org/10.48550/arXiv.1810.04805, arXiv preprint arXiv:1810.04805.
[138] K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, D. Brown, Text classification algorithms: A survey, Information 10 (4) (2019) 150, http://dx.doi.org/10.3390/info10040150.
[139] R. Real, J.M. Vargas, The probabilistic basis of Jaccard's index of similarity, Systematic Biol. 45 (3) (1996) 380–385, http://dx.doi.org/10.2307/2413572.
[140] B. Nagle, A proposal for dealing with grade inflation: The relative performance index, J. Educ. Bus. 74 (1) (1998) 40–43, http://dx.doi.org/10.1080/08832329809601659.

Chapter 5

Conclusions and Future Works

In this thesis, we begin by exploring various dimensions of news reporting bias, such as event selection, news sentiment, and political orientation, among others. By critically assessing existing work, we identified main challenges in this field and laid the foundation for identifying future research directions. In addition, we introduced an adaptable data-generation method and provided its implementation as scripts to facilitate the creation of a custom dataset for related tasks such as analysing complex event-outlet relationships. We then investigated the role of inferential commonsense knowledge in comprehending news headlines, addressing the challenges of capturing syntactic and semantic information in their short text. We have introduced a neural network framework using this knowledge for bias prediction in headlines. Extending this approach to multilingual contexts, we then introduced a language-agnostic framework that leverages inferential knowledge using a translate-retrieve-translate strategy for bias prediction in multilingual headlines. We have also presented datasets for evaluating the two aforementioned frameworks. The overall findings demonstrate the effectiveness of these frameworks, presenting plausible results even for individual languages.

5.1 Contributions to Science

The thesis has three key scientific contributions, which are discussed as follows:

- **SC 1. A Novel Data Generation Method:**

  We have designed an adaptable data generation method to generate customised datasets for related tasks and provided its implementation as scripts. Using the scripts, we have introduced a publicly available, novel dataset with a diverse set of features that focuses on events and news outlets, making it ideal for studying and analysing complex event-outlet relationships. We have then identified a potential use case to demonstrate how the dataset can be utilised for analysing the event coverage patterns of media outlets and estimating the correlation between them using conditional probabilistic models. We have also conducted a comprehensive statistical analysis with respect to multiple event features to compare and contrast the coverage patterns of the selected news outlets. Our analysis uncovered interesting findings, such as the outlets exhibiting a preference for specific categories of events and their coverage pattern being influenced by their geographical origin.

- **SC 2. Commonsense-Infused Bias Prediction Framework:**

  We have proposed leveraging inferential commonsense knowledge to enhance the comprehension of news headlines, where capturing syntactic and semantic information within its short text is challenging. We have shown that it is possible to employ this knowledge in a manner analogous to how people use common sense to perform their routine activities. By exploiting this knowledge with an emphasis on important inferences, we have introduced a novel neural network framework to address the challenge of predicting political bias in short headline texts. We also presented two datasets of news headlines annotated with political bias to evaluate our framework. We conducted experiments with several state-of-the-art language models on both of them to showcase the reliability and effectiveness of our framework. These models demonstrated significant improvements within our framework, with accuracy ranging from 2.0% to 10.0%, macro-averaged F1 scores from 2.2% to 22.2%, Jaccard scores up to 15.1%, and micro-averaged F1 scores up to 18.6%. The results clearly demonstrate that the language models, when used within our framework, exhibited superior performance on both datasets.

Our methodology highlights the potential of leveraging commonsense inferences to explain events that are not explicitly stated in context-deficient headlines. By enriching headlines with such knowledge, we showed that a model can focus not only on important entities and events but also on explaining unstated events, leading to enhanced predictions. As far as we know, our work is the first attempt at harnessing commonsense inferences to enhance bias prediction in short news headlines.

- **SC 3. Multilingual Language-Agnostic Bias Prediction Framework:**

  We have presented an extension of the proposed framework in SC 2 to deal with low-resource multilingual headlines under an imbalanced sample distribution in a language-agnostic setting. To achieve this, we have proposed leveraging multilingual commonsense knowledge through a translate-retrieve-translate strategy. Utilising this knowledge with proper attention mechanisms, we have introduced a learning framework for enhancing the prediction of political bias in multilingual headlines. We have proposed the framework to be language-agnostic, ensuring its adaptability to low-resource situations in real-world scenarios. In particular, we suggested integrating implicit knowledge derived from pre-trained language models with explicit inferential knowledge, as both implicit and explicit knowledge are crucial for the development of efficient systems. We also presented a dataset of headlines in five European languages with limited resources, along with their corresponding political bias annotations. We designed it to simulate the real-world challenges of imbalanced data distribution. We extensively evaluated our framework using several state-of-the-art multilingual, pre-trained language models. Overall, the best-performing model, trained solely with headlines, demonstrated a 0.90 accuracy, a 0.83 F1 score, and a 0.83 Jaccard score. The application of additional knowledge resulted in a significant improvement, increasing accuracy and F1 to 0.92 and the Jaccard score to 0.86. The results demonstrate that our framework is effective across different models, even yielding plausible results for individual languages.

  We highlighted the significance of multilingual commonsense inferences in facilitating comprehension of the overall narrative of short multilingual news headlines. To the best of our knowledge, our proposed strategy is one of the first attempts to use such knowledge in a multilingual setting to predict bias in news headlines.

5.2 Future Works

In the future, several key areas can be explored. For instance, an assessment of learning processes and the representativeness of state-of-the-art language models can be conducted, acknowledging and addressing the potential computational challenges they might pose. Exploring methods to address bias is another potential direction for investigation. Investigating the potential impact of inferential commonsense knowledge, external knowledge bases, and multimodal contexts can provide valuable insights for diverse use cases. Ensuring model transparency through enhanced explainability and interpretability can stand out as another possibility. Commitment to dataset expansion, including the incorporation of data from social media and multimedia sources, offers the potential to create a more comprehensive learning context. These directions are discussed in the following paragraphs.

Pre-trained large language models. Although the large state-of-the-art pre-trained language models used in the study outperformed humans on many tasks, we believe it is important to evaluate what they are learning and how representative they are. Also, these models require significant computing resources, which presents a challenge that is yet to be addressed. Additionally, we acknowledge the significance of exploring ChatGPT's learning dynamics and understanding its contribution to a deeper understanding of language models and their ability to capture real-world nuances.

Bias. Biases in language models, knowledge bases, or dataset annotations have the potential to misguide the prediction framework. Mitigating such biases involves exploring various workarounds, but it remains an active research area due to the near impossibility of completely eliminating these biases. In the future, exploring approaches like adversarial filtering and expanding the proposed datasets can prove to be instrumental in minimising biases.

Inferential commonsense knowledge. There exist several inferential knowledge bases with varying capabilities. Their impact on prediction performance can be investigated. However, incorporating commonsense renders minimal or insignificant enhancement in the absence of valuable information. Similarly, there are cases where it alone proves to be adequate for accurate prediction. Determining when to incorporate or omit commonsense in an accurate and quantitative manner remains a challenge that can be addressed. Furthermore, despite their potential to significantly enhance the framework's performance, the exploration of advanced techniques for knowledge selection and subsequent integration into the framework is yet to be investigated and can thus be explored in the future.

External knowledge bases. Additional insights acquired from external knowledge bases have yet to be explored. In future investigations, the integration of additional information to enrich our framework can be explored. Integrating knowledge from external sources can contribute valuable dimensions to our framework, enriching its capacity for nuanced understanding and predictions.

Multimodality. It would also be interesting to explore the task of bias prediction in a multimodal setting. The integration of common-sense knowledge from diverse modalities has the potential to add new dimensions to the challenge.

Use-cases. While our evaluation was not specifically designed to analyse the effectiveness of our method across various domains or use cases, the versatility of our approach allows for testing in diverse scenarios following comprehensive assessments. Consequently, there is potential to extend our study beyond bias prediction. For instance, conducting experiments involving auxiliary tasks such as bias quantification and correction within a multitask learning framework could offer valuable insights and add depth to our investigation. It could also enhance prediction of political bias by leveraging shared representations between tasks, improving generalisation across languages and domains, and promoting model interpretability through simultaneous learning of related tasks.

Explainability. Another aspect that can be addressed is the challenge of ensuring the model's explainability and interpretability. This includes developing methods for understanding and conveying how the model arrives at its predictions.

Dataset expansion. In the future, our dataset can be extended by incorporating diverse sources, such as social media and multimedia content, to ensure a more comprehensive representation of real-world scenarios. This expansion will not only enrich the dataset but also allow the model to learn from a broader set of contexts, enhancing its adaptability and predictive capabilities.

References

Aksenov, D., Bourgonje, P., Zaczynska, K., Ostendorff, M., Schneider, J. M., & Rehm, G. (2021). Fine-grained classification of political bias in german news: A data set and initial experiments. *Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)*, 121–131. https://doi.org/10.18653/v1/2021.woah-1.13

Andrew, B. C. (2007). Media-generated shortcuts: Do newspaper headlines present another roadblock for low-information rationality? *Harvard International Journal of Press/Politics*, *12*(2), 24–43. https://doi.org/10.1177/1081180X07299795

Bago, B., Rand, D., & Pennycook, G. (2019). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines.

Baker, B. H., Graham, T., & Kaminsky, S. (1994). *How to identify, expose & correct liberal media bias*. Media Research Center Alexandria, VA.

Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Van Der Goot, E., Halkia, M., Pouliquen, B., & Belyaeva, J. (2013). Sentiment analysis in the news. *arXiv preprint arXiv:1309.6202*. https://doi.org/10.48550/arXiv.1309.6202

Baly, R., Da San Martino, G., Glass, J., & Nakov, P. (2020). We can detect your bias: Predicting the political ideology of news articles. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 4982–4991. https://doi.org/10.18653/v1/2020.emnlp-main.404

Berner, R. T. (1983). Commentary: The narrative and the headline. *Newspaper Research Journal*, *4*(3), 33–40. https://doi.org/10.1177/073953298300400305

Bernhardt, D., Krasa, S., & Polborn, M. (2008). Political polarization and the electoral effects of media bias. *Journal of Public Economics*, *92*(5-6), 1092–1104.

Bestvater, S. E., & Monroe, B. L. (2023). Sentiment is not stance: Target-aware opinion classification for political text analysis. *Political Analysis*, *31*(2), 235–256. https://doi.org/10.1017/pan.2022.10

Bourgeois, D., Rappaz, J., & Aberer, K. (2018). Selection bias in news coverage: Learning it, fighting it. *Companion Proceedings of the The Web Conference 2018*, 535–543.

Bruneau, E. G., Cikara, M., & Saxe, R. (2012). Going beyond the headlines: Narratives mitigate intergroup empathy bias. [https://mindmodeling.org/cogsci2012/papers/0475/index.html]. *Proceedings of the 34th Annual Meeting of the Cognitive Science Society, CogSci 2012*.

Bucher, H.-J., & Schumacher, P. (2006). The relevance of attention for selecting news content. an eye-tracking study on attention patterns in the reception of print and online media. *Communications*, *31*(3), 347–368.

Bursztyn, L., Rao, A., Roth, C. P., & Yanagizawa-Drott, D. H. (2020). *Misinformation during a pandemic* (tech. rep.). National Bureau of Economic Research. https://doi.org/10.3386/w27417

Chakrabarty, T., Ghosh, D., Muresan, S., & Peng, N. (2020). R^3: Reverse, retrieve, and rank for sarcasm generation with commonsense knowledge [10.18653/v1/2020.acl-main.711]. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 7976–7986.

Chen, C., Li, Y., Wei, C., Cui, J., Wang, B., & Yan, R. (2024). Empathetic response generation with relation-aware commonsense knowledge. *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, 87–95. https://doi.org/10.1145/3616855.3635836

Chen, W.-F., Al Khatib, K., Stein, B., & Wachsmuth, H. (2020). Detecting media bias in news articles using gaussian bias distributions. *Findings of the Association for Computational Linguistics: EMNLP 2020*, 4290–4300. https://doi.org/10.18653/v1/2020.findings-emnlp.383

Chen, W.-F., Wachsmuth, H., Al Khatib, K., & Stein, B. (2018). Learning to flip the bias of news headlines. *Proceedings of the 11th International Conference on Natural Language Generation*, 79–88. https://doi.org/10.18653/v1/W18-6509

Cheng, C., Barceló, J., Hartnett, A. S., Kubinec, R., & Messerschmidt, L. (2020). Covid-19 government response event dataset (coronanet v. 1.0). *Nature human behaviour*, *4*(7), 756–768. https://doi.org/10.1038/s41562-020-0909-7

Chiang, C.-F., & Knight, B. (2011). Media bias and influence: Evidence from newspaper endorsements. *The Review of economic studies*, *78*(3), 795–820. https://doi.org/10.1093/restud/rdq037

Cho, D. B., Lee, H. Y., Jung, W. S., & Kang, S. S. (2021). Automatic classification and vocabulary analysis of political bias in news articles by using subword tokenization. *KIPS Transactions on Software and Data Engineering*, *10*(1), 1–8.

Chowdhury, S. B. R., & Chaturvedi, S. (2021). Does commonsense help in detecting sarcasm? *Proceedings of the Second Workshop on Insights from Negative Results in NLP*, 9–15. https://doi.org/10.18653/v1/2021.insights-1.2

D'Alessio, D. (2012). *Media bias in presidential election coverage, 1948-2008: Evaluation via formal measurement*. Lexington Books.

Dallmann, A., Lemmerich, F., Zoller, D., & Hotho, A. (2015). Media bias in german online newspapers. *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, 133–137. https://doi.org/10.1145/2700171.2791057

Davis, S. R., Worsnop, C. J., & Hand, E. M. (2022). Gender bias recognition in political news articles. *Machine Learning with Applications*, *8*, 100304. https://doi.org/10.1016/j.mlwa.2022.100304

De Vreese, C. H. (2005). News framing: Theory and typology. *Information Design Journal & Document Design*, *13*(1).

Doan, T. M., & Gulla, J. A. (2022). A survey on political viewpoints identification. *Online Social Networks and Media*, *30*, 100208. https://doi.org/10.1016/j.osnem.2022.100208

Du, L., Ding, X., Xiong, K., Liu, T., & Qin, B. (2022). Enhancing pretrained language models with structured commonsense knowledge for textual inference. *Knowledge-Based Systems*, *254*, 109488. https://doi.org/10.1016/j.knosys.2022.109488

Dunham, W. R. (2013). Framing the right suspects: Measuring media bias. *Journal of Media Economics*, *26*(3), 122–147. https://doi.org/10.1080/08997764.2013.811587

Ecker, U. K., Lewandowsky, S., Chang, E. P., & Pillai, R. (2014). The effects of subtle misinformation in news headlines. *Journal of experimental psychology: applied*, *20*(4), 323. https://doi.org/10.1037/xap0000028

Einsohn, A., & Schwartz, M. (2019). *The copyeditor's handbook: A guide for book publishing and corporate communications*. University of California Press. https://doi.org/10.1525/9780520972278

El Emam, K., Mosquera, L., & Hoptroff, R. (2020). *Practical synthetic data generation: Balancing privacy and the broad availability of data* [https://cdn.ttgtmedia.com/rms/pdf/Practical_Synthetic_Data_Generation.pdf]. O'Reilly Media.

Estrin, J. (2014). The real story about the wrong photos in# bringbackour-girls. *The New York Times*.

Fan, L., White, M., Sharma, E., Su, R., Choubey, P. K., Huang, R., & Wang, L. (2019). In plain sight: Media bias through the lens of factual reporting. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 6343–6349. https://doi.org/10.18653/v1/D19-1664

Fang, Y., Wang, S., Xu, Y., Xu, R., Sun, S., Zhu, C., & Zeng, M. (2022). Leveraging knowledge in multilingual commonsense reasoning. *Findings of the Association for Computational Linguistics: ACL 2022*, 3237–3246. https://doi.org/10.18653/v1/2022.findings-acl.255

Gangula, R. R. R., Duggenpudi, S. R., & Mamidi, R. (2019). Detecting political bias in news articles using headline attention. *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, 77–84. https://doi.org/10.18653/v1/W19-4809

Gentzkow, M., & Shapiro, J. M. (2006). Media bias and reputation. *Journal of political Economy*, *114*(2), 280–316.

Grefenstette, G., Qu, Y., Shanahan, J. G., & Evans, D. A. (2004). Coupling niche browsers and affect analysis for an opinion mining application. *Proceedings of Recherche d'Information Assistée par Ordinateur (RIAO)*.

Groseclose, T., & Milyo, J. (2005). A measure of media bias. *The Quarterly Journal of Economics*, *120*(4), 1191–1237.

Guo, S., & Zhu, K. Q. (2022). Modeling multi-level context for informational bias detection by contrastive learning and sentential graph network. *arXiv preprint arXiv:2201.10376*. https://doi.org/10.48550/arXiv.2201.10376

Hamborg, F., Donnay, K., & Gipp, B. (2019). Automated identification of media bias in news articles: An interdisciplinary literature review. *International Journal on Digital Libraries*, *20*(4), 391–415. https://doi.org/10.1007/s00799-018-0261-y

Hamborg, F., Meuschke, N., & Gipp, B. (2020). Bias-aware news analysis using matrix-based news aggregation. *International Journal on Digital Libraries*, *21*(2), 129–147. https://doi.org/10.1007/s00799-018-0239-9

Haselmayer, M., Wagner, M., & Meyer, T. M. (2017). Partisan bias in message selection: Media gatekeeping of party press releases. *Political communication*, *34*(3), 367–384. https://doi.org/10.1080/10584609.2016.1265619

Holmqvist, K., Holsanova, J., Barthelson, M., & Lundqvist, D. (2003). Reading or scanning? a study of newspaper and net paper reading. In *The mind's eye* (pp. 657–670). Elsevier. https://doi.org/10.1016/B978-044451020-4/50035-9

Huang, L., Le Bras, R., Bhagavatula, C., & Choi, Y. (2019). Cosmos qa: Machine reading comprehension with contextual commonsense reasoning. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2391–2401. https://doi.org/10.18653/v1/D19-1243

Hwang, J. D., Bhagavatula, C., Le Bras, R., Da, J., Sakaguchi, K., Bosselut, A., & Choi, Y. (2021). (comet-) atomic 2020: On symbolic and neural commonsense knowledge graphs [https://ojs.aaai.org/index.php/AAAI/article/view/16792]. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*, 6384–6392.

Kahneman, D., & Tversky, A. (2013). Choices, values, and frames. In *Handbook of the fundamentals of financial decision making: Part i* (pp. 269–278). World Scientific.

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. *Multimedia tools and applications*, *82*(3), 3713–3744.

Knobloch-Westerwick, S., Mothes, C., & Polavin, N. (2020). Confirmation bias, ingroup bias, and negativity bias in selective exposure to political information. *Communication research*, *47*(1), 104–124. https://doi.org/10.1177/00936502177195

Knott, K., & Poole, E. (2016). *Media portrayals of religion and the secular sacred: Representation and change*. Routledge. https://doi.org/10.4324/9781315594514

Krieger, J.-D., Spinde, T., Ruas, T., Kulshrestha, J., & Gipp, B. (2022). A domain-adaptive pre-training approach for language bias detection in news. *Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries*, 1–7. https://doi.org/10.1145/3529372.3530932

Laban, P., Bandarkar, L., & Hearst, M. A. (2021). News headline grouping as a challenging nlu task. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 3186–3198. https://doi.org/10.18653/v1/2021.naacl-main.255

Lazaridou, K., Löser, A., Mestre, M., & Naumann, F. (2020). Discovering biased news articles leveraging multiple human annotations [=https://aclanthology.org/2020.lrec-1.159]. *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 1268–1277.

Leban, G., Fortuna, B., Brank, J., & Grobelnik, M. (2014). Event registry: Learning about world events from news. *Proceedings of the 23rd International Conference on World Wide Web*, 107–110. https://doi.org/10.1145/2567948.257702

Lee, J., Cha, J., & Park, E. (2023). Data-driven approaches into political orientation and news outlet discrimination: The case of news articles in south korea. *Telematics and Informatics*, *85*, 102066. https://doi.org/10.1016/j.tele.2023.102066

Lee, S., Kim, J., Kim, D., Kim, K. J., & Park, E. (2023). Computational approaches to developing the implicit media bias dataset: Assessing political orientations of nonpolitical news articles. *Applied Mathematics and Computation*, *458*, 128219. https://doi.org/10.1016/j.amc.2023.128219

Leetaru, K., & Schrodt, P. A. (2013). Gdelt: Global data on events, location, and tone, 1979–2012. *ISA annual convention*, *2*(4), 1–49. https://doi.org/10.1.1.686.6605

Li, D., Zhu, X., Li, Y., Wang, S., Li, D., Liao, J., & Zheng, J. (2021). Enhancing emotion inference in conversations with commonsense knowledge [10.1016/j.knosys.2021.107449]. *Knowledge-Based Systems*, *232*, 107449.

Li, J., Lin, Z., Fu, P., & Wang, W. (2021). Past, present, and future: Conversational emotion recognition through structural modeling of psychological knowledge. *Findings of the Association for Computational Linguistics: EMNLP 2021*, 1204–1214. https://doi.org/10.18653/v1/2021.findings-emnlp.104

Li, Q., Peng, H., Li, J., Xia, C., Yang, R., Sun, L., Yu, P. S., & He, L. (2022). A survey on text classification: From traditional to deep learning. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *13*(2), 1–41.

Lieto, A., Pozzato, G. L., Zoia, S., Patti, V., & Damiano, R. (2021). A commonsense reasoning framework for explanatory emotion attribution, generation and re-classification. *Knowledge-Based Systems*, *227*, 107166. https://doi.org/10.1016/j.knosys.2021.107166

Liu, D., Gong, Y., Yan, Y., Fu, J., Shao, B., Jiang, D., Lv, J., & Duan, N. (2020). Diverse, controllable, and keyphrase-aware: A corpus and method for news multi-headline generation. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 6241–6250. https://doi.org/10.18653/v1/2020.emnlp-main.505

Liu, R., Jia, C., & Vosoughi, S. (2021). A transformer-based framework for neutralizing and reversing the political polarity of news articles. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–26. https://doi.org/10.1145/3449139

Liu, R., Wang, L., Jia, C., & Vosoughi, S. (2021). Political depolarization of news articles using attribute-aware word embeddings. *Proceedings of the International AAAI Conference on Web and Social Media*, *15*, 385–396. https://doi.org/10.1609/icwsm.v15i1.18069

Lundman, R. J. (2003). The newsworthiness and selection bias in news about murder: Comparative and relative effects of novelty and race and gender typifications on newspaper coverage of homicide. *Sociological forum*, *18*, 357–386. https://doi.org/10.1023/A:1025713518156

Mahmood, S. (2009). Religious reason and secular affect: An incommensurable divide? *Critical inquiry*, *35*(4), 836–862. https://doi.org/10.1086/599592

Majumder, N., Hong, P., Peng, S., Lu, J., Ghosal, D., Gelbukh, A., Mihalcea, R., & Poria, S. (2020). Mime: Mimicking emotions for empathetic response generation. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 8968–8979. https://doi.org/10.18653/v1/2020.emnlp-main.721

Martin, A. W. (2005). Addressing the selection bias in media coverage of strikes: A comparison of mainstream and specialty print media. In *Research in social movements, conflicts and change* (pp. 141–178). Emerald Group Publishing Limited. https://doi.org/10.1016/S0163-786X(05)26005-6

Mastrorocco, N., & Minale, L. (2018). News media and crime perceptions: Evidence from a natural experiment. *Journal of Public Economics*, *165*, 230–255. https://doi.org/10.1016/j.jpubeco.2018.07.002

McCarthy, K. J., & Dolfsma, W. (2014). Neutral media? evidence of media bias and its economic impact. *Review of Social Economy*, *72*(1), 42–54. https://doi.org/10.1080/00346764.2013.806110

Mello, C., Cheema, G. S., & Thakkar, G. (2023). Combining sentiment analysis classifiers to explore multilingual news articles covering london 2012 and rio 2016 olympics. *International Journal of Digital Humanities*, *5*(2), 131–157. https://doi.org/10.1007/s42803-022-00052-9

Meyrowitz, J. (1986). *No sense of place: The impact of electronic media on social behavior*. Oxford University Press.

Min, B., Ross, H., Sulem, E., Veyseh, A. P. B., Nguyen, T. H., Sainz, O., Agirre, E., Heintz, I., & Roth, D. (2023). Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Computing Surveys*, *56*(2), 1–40.

Mowery, J. (2021). Contrastive clustering: Toward unsupervised bias reduction for emotion and sentiment classification. *arXiv preprint arXiv:2111.07448*. https://doi.org/10.48550/arXiv.2111.07448

Nah, S., Luo, J., Kim, S., Chen, M., Mitson, R., & Joo, J. (2023). Rethinking artificial intelligence: Algorithmic bias and ethical issues| algorithmic bias or algorithmic reconstruction? a comparative analysis between ai news and human news. *International Journal of Communication*, *18*, 30.

Naredla, N. R., & Adedoyin, F. F. (2022). Detection of hyperpartisan news articles using natural language processing technique. *International Journal of Information Management Data Insights*, *2*(1), 100064. https://doi.org/10.1016/j.jjimei.2022.100064

Nisar, H., & Bleich, E. (2020). Group status, geographic location, and the tone of media coverage: Jews and muslims in new york times and guardian headlines, 1985–2014. *Comparative Migration Studies*, *8*(1), 3. https://doi.org/10.1186/s40878-019-0153-3

Oberländer, L. A. M., Kim, E., & Klinger, R. (2020). Goodnewseveryone: A corpus of news headlines annotated with emotions, semantic roles, and reader perception [https://www.aclweb.org/anthology/2020.lrec-1.194/]. *Proceedings of the 12th Language Resources and Evaluation Conference*, 1554–1566.

Park, S., Kang, S., Chung, S., & Song, J. (2009). Newscube: Delivering multiple aspects of news to mitigate media bias. *Proceedings of the SIGCHI conference on human factors in computing systems*, 443–452. https://doi.org/10.1145/1518701.1518772

Park, S., Kang, S., Chung, S., & Song, J. (2012). A computational framework for media bias mitigation. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *2*(2), 1–32. https://doi.org/10.1145/2209310.2209311

Patil, R., Boit, S., Gudivada, V., & Nandigam, J. (2023). A survey of text representation and embedding techniques in nlp. *IEEE Access*. https://doi.org/10.1109/ACCESS.2023.3266377

Pingree, R. J., Quenette, A. M., Tchernev, J. M., & Dickinson, T. (2013). Effects of media criticism on gatekeeping trust and implications for agenda setting. *Journal of Communication*, *63*(2), 351–372. https://doi.org/10.1111/jcom.12016

Pino, C., Kavasidis, I., & Spampinato, C. (2016). Assessment and visualization of geographically distributed event-related sentiments by mining social networks and news. *2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, 354–358. https://doi.org/10.1109/CCNC.2016.7444806

Rakhecha, K., Rauniyar, S., Agrawal, M., & Bhatt, A. (2023). A survey on bias detection in online news using deep learning. *2023 2nd international conference on applied artificial intelligence and computing (ICAAIC)*, 396–403. https://doi.org/10.1109/ICAAIC56838.2023.10140917

Ranathunga, S., Lee, E.-S. A., Prifti Skenduli, M., Shekhar, R., Alam, M., & Kaur, R. (2023). Neural machine translation for low-resource languages: A survey. *ACM Computing Surveys*, *55*(11), 1–37.

Ravi, S., Chinchure, A., Sigal, L., Liao, R., & Shwartz, V. (2023). Vlc-bert: Visual question answering with contextualized commonsense knowledge. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1155–1165.

Raza, S., Garg, M., Reji, D. J., Bashir, S. R., & Ding, C. (2024). Nbias: A natural language processing framework for bias identification in text. *Expert Systems with Applications*, *237*, 121542. https://doi.org/10.1016/j.eswa.2023.121542

Ribeiro, F., Henrique, L., Benevenuto, F., Chakraborty, A., Kulshrestha, J., Babaei, M., & Gummadi, K. (2018). Media bias monitor: Quantifying biases of social media news outlets at large-scale. *Proceedings of the International AAAI Conference on Web and Social Media*, *12*(1). https://doi.org/10.1609/icwsm.v12i1.15025

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). " why should i trust you?" explaining the predictions of any classifier [10.1145/2939672.2939778]. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 1135–1144.

Rodrigo-Ginés, F.-J., Carrillo-de-Albornoz, J., & Plaza, L. (2023). A systematic review on media bias detection: What is media bias, how it is expressed, and how to detect it. *Expert Systems with Applications*, 121641. https://doi.org/10.1016/j.eswa.2023.121641

Sabour, S., Zheng, C., & Huang, M. (2022). Cem: Commonsense-aware empathetic response generation. *Proceedings of the AAAI Conference on Artificial Intelligence*, *36*, 11229–11237. https://doi.org/10.1609/aaai.v36i10.21373

Sap, M., Le Bras, R., Allaway, E., Bhagavatula, C., Lourie, N., Rashkin, H., Roof, B., Smith, N. A., & Choi, Y. (2019). Atomic: An atlas of machine commonsense for if-then reasoning. *Proceedings of the AAAI conference on artificial intelligence*, *33*(01), 3027–3035. https://doi.org/10.1609/aaai.v33i01.33013027

Spinde, T. (2021). An interdisciplinary approach for the automated detection and visualization of media bias in news articles. *2021 International Conference on Data Mining Workshops (ICDMW)*, 1096–1103. https://doi.org/10.1109/ICDMW53433.2021.00144

Spinde, T., Krieger, D., Plank, M., & Gipp, B. (2021). Towards a reliable ground-truth for biased language detection. *2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 324–325. https://doi.org/10.1109/JCDL52503.2021.00053

Spinde, T., Rudnitckaia, L., Mitrović, J., Hamborg, F., Granitzer, M., Gipp, B., & Donnay, K. (2021). Automated identification of bias inducing words in news articles using linguistic and context-oriented features. *Information Processing & Management*, *58*(3), 102505. https://doi.org/10.1016/j.ipm.2021.102505

Stefanov, P., Darwish, K., Atanasov, A., & Nakov, P. (2020). Predicting the topical stance and political leaning of media using tweets. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 527–537. https://doi.org/10.18653/v1/2020.acl-main.50

Stovall, J. G. (1985). The third-party challenge of 1980: News coverage of the presidential candidates. *Journalism Quarterly*, *62*(2), 266–271.

Swati, S., Grobelnik, A. M., Mladenić, D., & Grobelnik, M. (2023). A commonsense-infused language-agnostic learning framework for enhancing prediction of political bias in multilingual news headlines. *Knowledge-Based Systems*, *277*, 110838. https://doi.org/10.1016/j.knosys.2023.110838

Swati, S., & Mladenić, D. (2020). Are you following the right news outlet? a machine learning-based approach to outlet prediction. *Proceedings of the 23rd International Multiconference Information Society – IS 2020*, *100*(7), 33–36.

Swati, S., & Mladenić, D. (2021). Understanding the impact of geographical bias on news sentiment: A case study on london and rio olympics. *Proceedings of the 24th International Multiconference Information Society – IS 2021*, *100*(5), 21–24.

Swati, S., Mladenić, D., & Erjavec, T. (2021). Eveout: An event-centric news dataset to analyze an outlet's event selection patterns. *Informatica*, *45*(7). https://doi.org/10.31449/inf.v45i7.3410

Swati, S., Mladenić, D., & Grobelnik, M. (2023). An inferential commonsense-driven framework for predicting political bias in news headlines. *IEEE Access*. https://doi.org/10.1109/ACCESS.2023.3298877

Tan, Y. Y., Chow, C.-O., Kanesan, J., Chuah, J. H., & Lim, Y. (2023). Sentiment analysis and sarcasm detection using deep multi-task learning. *Wireless personal communications*, *129*(3), 2213–2237. https://doi.org/10.1007/s11277-023-10235-4

Vaca, C. K., Mantrach, A., Jaimes, A., & Saerens, M. (2014). A time-based collective factorization for topic discovery and monitoring in news. *Proceedings of the 23rd international conference on World wide web*, 527–538. https://doi.org/10.1145/2566486.2568041

Wessel, M., Horych, T., Ruas, T., Aizawa, A., Gipp, B., & Spinde, T. (2023). Introducing mbib-the first media bias identification benchmark task and dataset collection. *Proceedings of the 46th International ACM SIGIR Conference on Research and De-*

velopment in Information Retrieval, 2765–2774. https://doi.org/10.1145/3539618.3591882

Widyassari, A. P., Rustad, S., Shidik, G. F., Noersasongko, E., Syukur, A., Affandy, A., et al. (2022). Review of automatic text summarization techniques & methods. *Journal of King Saud University-Computer and Information Sciences*, *34*(4), 1029–1046.

Zhu, Y., Yuan, H., Wang, S., Liu, J., Liu, W., Deng, C., Dou, Z., & Wen, J.-R. (2023). Large language models for information retrieval: A survey. *arXiv preprint arXiv:2308.07107*.

Bibliography

Publications Related to the Thesis

Journal Articles

Swati, S., Grobelnik, A. M., Mladenić, D., & Grobelnik, M. (2023). A commonsense-infused language-agnostic learning framework for enhancing prediction of political bias in multilingual news headlines. *Knowledge-Based Systems*, *277*, 110838. https://doi.org/10.1016/j.knosys.2023.110838

Swati, S., Mladenić, D., & Erjavec, T. (2021). Eveout: An event-centric news dataset to analyze an outlet's event selection patterns. *Informatica*, *45*(7). https://doi.org/10.31449/inf.v45i7.3410

Swati, S., Mladenić, D., & Grobelnik, M. (2023). An inferential commonsense-driven framework for predicting political bias in news headlines. *IEEE Access*. https://doi.org/10.1109/ACCESS.2023.3298877

Conference Paper

Swati, S., Erjavec, T., & Mladenić, D. (2020). Eveout: Reproducible event dataset for studying and analyzing the complex event-outlet relationship. *Proceedings of the 23rd International Multiconference Information Society – IS 2020*, *100*(4), 17–20.

Swati, S., & Mladenić, D. (2020). Are you following the right news outlet? a machine learning-based approach to outlet prediction. *Proceedings of the 23rd International Multiconference Information Society – IS 2020*, *100*(7), 33–36.

Swati, S., & Mladenić, D. (2021). Understanding the impact of geographical bias on news sentiment: A case study on london and rio olympics. *Proceedings of the 24th International Multiconference Information Society – IS 2021*, *100*(5), 21–24.

Biography

Swati is currently pursuing her PhD at the Jožef Stefan International Postgraduate School in Ljubljana, Slovenia. Her current research focuses on the analysis and development of deep learning-based solutions for cross-lingual news reporting bias. It includes domains such as news-value determination and computational journalism. The scope of her work spans a wide range of dimensions, such as topic, language, geography, source, sentiment, time, and attention. The primary objective of her research is to examine how different methodologies address constraints and specific linguistic characteristics, particularly in languages that lack sufficient resources. She is actively involved in advancing research and development in machine learning, deep learning, data and text mining, and semantic technology while exploring their diverse practical applications. From explainable AI and inferential commonsense AI to continual learning, she possesses diverse expertise. This includes a thorough exploration of the potential and challenges associated with multimodal AI.

She also acquired valuable experience as an early-stage researcher for the EU-funded Horizon 2020 project CLEOPATRA, a Marie Skłodowska-Curie Innovative Training Network (MSCA-ITN) at the Jožef Stefan Institute in Ljubljana, Slovenia. During her involvement in the project, she worked as a visiting researcher at the Advanced Services Department at Unidade FCCN, Computação Científica Nacional in Lisbon, Portugal, with a focus on archives. She also worked in the Department of Media Studies at the University of Amsterdam in the Netherlands, where her research focused on digital methods in the field of new media studies.

Prior to this, she held an M.Tech. degree in mathematics and computing from the Indian Institute of Technology, Patna, India. The goal of her master's thesis was to enhance medical visual question answering by developing a hierarchical deep neural multimodal network with a question-type segregation module. The goal was to generate accurate answers to queries related to medical images, specifically addressing the generic VQA-Med problem. It was centred around various domains, including visual question answering, computer vision, natural language processing, machine learning, and deep learning.