FrameFinder: Explorative Multi-Perspective Framing Extraction from News Headlines

Markus Reiter-Haas
reiter-haas@tugraz.at
TU Graz, Austria

Beate Klösch
beate.kloesch@uni-graz.at
University of Graz, Austria

Markus Hadler
markus.hadler@uni-graz.at
University of Graz, Austria

Elisabeth Lex
elisabeth.lex@tugraz.at
TU Graz, Austria

ABSTRACT

Revealing the framing of news articles is an important yet neglected task in information seeking and retrieval. In the present work, we present FrameFinder, an open tool for extracting and analyzing frames in textual data. FrameFinder visually represents the frames of text from three perspectives, i.e., (i) frame labels, (ii) frame dimensions, and (iii) frame structure. By analyzing the well-established gun violence frame corpus, we demonstrate the merits of our proposed solution to support social science research and call for subsequent integration into information interactions.

CCS CONCEPTS

• **Information systems** → **Content analysis and feature selection**; World Wide Web; Language models; • **Computing methodologies** → *Information extraction.*

KEYWORDS

Computational Framing Extraction, Exploratory Content Analysis, Media Bias, Text Representations, Online News

ACM Reference Format:
Markus Reiter-Haas, Beate Klösch, Markus Hadler, and Elisabeth Lex. 2023. FrameFinder: Explorative Multi-Perspective Framing Extraction from News Headlines. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

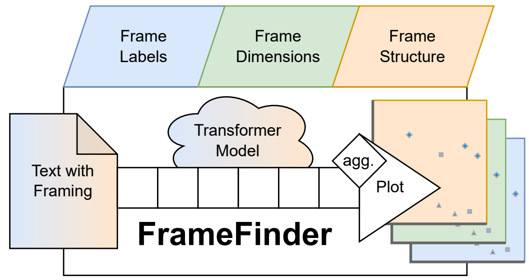


Figure 1: Schematic overview of the framing detection tool.

1 INTRODUCTION

Cognitive biases, such as framing effects, influence information seeking and retrieval behaviors [3]. In this vein, biased search results have been shown to affect user attitudes due to exposure [7]. Moreover, it has been well established in psychology that framing also affects the behavior and choices of people [32]. Detecting and understanding the framing of online news is thus important due to its influence on readers, but also very challenging [21]. While there are several approaches for computational framing analysis (see [2] for an overview), many rely on annotated data and train a classifier. However, framing is defined as *the selection and salience of aspects in a communicating text* [8] and thus requires a deeper understanding than just doing predictions. Moreover, even in such supervised settings, the amount of available data is typically rather sparse. The sparsity issue was therefore one of the main challenges in the recent shared framing detection task at SemEval 2023 [22], where only a few or even zero samples were available per language. Notably, the best performing teams all used pretrained Transformers to tackle the task at hand [25, 16, 35]. Due to these challenges, the landscape of framing detection tools is still shallow, especially regarding openly available ones (e.g., [5]).

In the present work, we expand upon existing computational framing research by providing a novel tool to discover and extract frames from texts with a focus on online news. *FrameFinder* extracts frames from three distinct perspectives using Transformer models [33]. As described in [24], frames can be analyzed (i) by their associated *frame labels*, (ii) their *frame dimensions*, and (iii) their *frame structure.* To showcase the benefits of the tool, we conducted an analysis on the gun violence frame corpus (GVFC) [17]. There we find that the discussion is mostly framed regarding security rather than health, despite the names of involved people being a major structural element. Besides, the **openly available library and online demonstration**[1] allows both social science researchers and novice users to analyze the framing of texts without requiring technical (e.g., programming) skills. For future research, we strive to incorporate framing analyses directly into the retrieval process of online news to accomplish more balanced media consumption of users, either by informing them about the framing bias or by adapting, e.g., reranking, the retrieved results.

2 FRAMEFINDER: FRAMING DETECTION

Framing has multiple definitions across various scientific disciplines [31]. In this work, we consider communicative frames following Entman [8] regarding the *selection and salience of aspects in a communicating text to promote a specific interpretation.* As a result,

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/XXXXXXX.XXXXXXX

[1]The demo is available at: **https://huggingface.co/spaces/Iseratho/frame-finder** and accompanied by a brief video introduction: **https://iseratho.github.io/external/frame-finder-video.html** The underlying code is also available as a standalone Python library for full customization of algorithms and configuration: **https://github.com/Iseratho/framefinder** that can be installed via *pip*: `pip install framefinder`

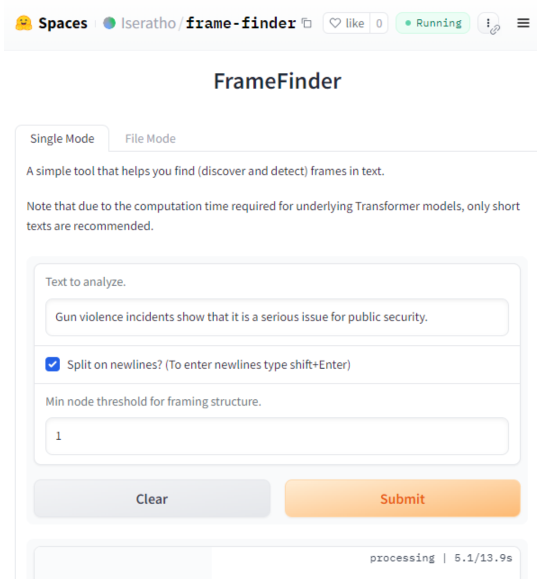


Figure 2: Truncated Screenshot of the Online Demo. For an overview of the generated plots, refer to experimental results in Figure 3.

framing deals with the presentation on both a micro and macro level [27]. Due to such nuances, framing is difficult to identify for algorithms [21]. Therefore, conceptualizations of framing are often only partially considered in automatic text processing [2].

FrameFinder is a tool to discover and extract frames from textual data using multiple distinct perspectives. As depicted in Figure 1, the tool takes texts as input that are (deliberately or undeliberately) framed in a certain way. The aim is to extract those frames in a human-comprehensible manner. To that end, we use the expressive power of Transformer models [33]. Internally, Transformers use embeddings, i.e., numerical vectors, to create rich representations of texts and parts thereof. The output representations, which can be probability vectors, alignment scores, or graph representations, are then aggregated and plotted. As previously identified in [24], we consider three distinct types of representations for framing analysis, i.e., frame labels, frame dimensions, and frame structure. For each type of representation, FrameFinder aggregates the result (when analyzing more than 1 sample) and visualizes them in a suitable format. Taken together, such a multi-perspective view of the data allows for a more nuanced framing analysis and the customizability of the library enables an explorative way to not only detect established but also discover novel frames.

Online Demonstration. For the online demonstration, we built the core part using HuggingFace Transformers [34] library and models. Together with Gradio [1], we deployed it as a HuggingFace space (see Figure 2). The demo runs on a CPU-only instance with 16 GB of RAM. The basic interface comprises two modes, a text-based and a file-based mode. The first allows entering example(s) in a text box, while the latter requires the upload of a text file. In both modes, the text is by default split on newlines into individual documents that are analyzed and aggregated. This option can be disabled to analyze the corpus as a single document (which is only recommended for short texts, as both the probability of frames being present and text structure tend to increase with text length). Additionally, there is a filtering option for the structural visualization based on node occurrence within graphs (i.e., the degree-weighted frequency across individual graphs). Finally, in the text-based mode, a few examples are provided that are cached (i.e., pre-computed) and thus evaluated instantly. For the deployed configuration (i.e., models and definitions of labels/dimensions) refer to the detailed description in Section 3 that was conducted with the same settings.

In the following, we describe the basic approaches of the three types of framing perspectives. Afterward, we discuss the relation of the tool to social science research.

2.1 Frame Labels

Framing detection can be approached as a classification task, in which specific *frame labels* are predicted to be either present or absent. This typically requires an annotated corpus. However, such corpora are scarce, with notable examples including the media frame corpus [6], the gun violence frame corpus [17], and the SemEval Task 3 Subtask 2 corpus [22]. Moreover, the number of samples within these corpora are typically rather small[2]. Alternatively, when given label definitions, the label prediction can also be modeled as a zero-shot prediction task.

Recent efforts to predict frame labels include contributions to the SemEval tasks (e.g., [35, 16, 25]) and the OpenFraming tool [5]. The latter differentiates between *frame discovery* using topic models and *frame prediction*, which involves training a classification model. Due to this explorative nature, it is similar in spirit to FrameFinder but requires expert knowledge and labor to annotate the data through content analysis. In contrast, we strive to avoid manual annotations, by considering multiple perspectives instead.

For aggregation of the prediction, we consider the mean and standard error of the label probabilities per sample. We then visualize the aggregated scores using a bar chart, and typically consider a threshold of 0.5 (denoted by color) to be indicative of which frame labels to assign to the corpus as a whole.

2.2 Frame Dimensions

Some frames are defined antagonistically, such as concerning moral foundations [11]. Considering the antagonistic care/harm pair, a text can be framed either positively emphasizing care (i.e., as a virtue) or negatively with harm in mind (i.e., as a vice), but not both. Such dimensions can be analyzed by considering the alignment within the embedding spaces of words and documents. Example approaches of dimensional framing analysis are moral framing in news [20], political framing on social media [13], or both, i.e., moral framing of political messages on social media [26].

The framing of documents can be analyzed either on a per-word basis using e.g. Word2Vec [19] or on a per-document basis using e.g. Sentence-Transformers [23]. In both scenarios, the position of an embedding (of a word or document) concerning the anchor embeddings (from the antagonistic pair) is determined. Herein, the

[2] The media frame corpus version 2 contains three subcorpora with 6327 on average but is deprecated due to changes in LexisNexis interface. The gun violence frame corpus contains 2990 samples, while the shared task in SemEval contains 2,049 split among train/dev/test set and nine languages (with three languages only in test).

FrameAxis method [14] scores the *frame bias* and intensity by projecting embeddings onto the axis formed by the antagonistic pair. The frame bias is defined as the mean of the scores, while the intensity considers the variance. Hence, the former specifies the leaning towards a frame, while the latter determines the activity along an axis. In the present work, we use FrameAxis for aggregating alignment scores but apply it to documents rather than words. The dimensions are plotted using horizontal lines, with the position of the projected points specifying the bias and their size specifying the intensity after aggregation.

2.3 Frame Structure

Some frames within a text are even more nuanced and require the consideration of the semantic structure. In this regard, the relations between the parts of text (e.g., words or phrases) are vital to extract the framing. One potential method for structural analysis is semantic role labeling (SRL) [10] that assigns tags that identify the type of argument in relation to a predicate. Two common examples of semantic roles are the *agent* tag, which is typically the subject, and *patients*, which are usually objects. An example approach for framing analysis is detailed in [13], where the agents and patients are visualized as tree stumps.

Alternatively, abstract meaning representations (AMR) [4] explicitly capture the semantic relations as rooted, directed, acyclic graphs[3]. In addition to extracting the semantic roles, these semantic graphs transform words and phrases into simplified semantic concepts, which improves comparability and subsequent transformations. Therefore, and in line with [24], we use AMR in the present work. When aggregating multiple semantic graphs, we create a weighted metagraph by superimposition of individual graphs. Thus, more pronounced concepts and relations get more emphasis, while additionally allowing filtering operations to only retain the most common elements of the metagraph.

2.4 Relation to Social Science Research

In the social sciences, such as sociology or communication studies, the analysis of frames also plays an important role, particularly in qualitative social research such as content analysis. Here, texts are typically coded manually, either deductively, i.e., on the basis of predetermined theoretical aspects [18], or inductively derived from the data material, as in the case of grounded theory [30]. FrameFinder works similarly to deductive content analysis by assigning predefined frames, i.e., frame labels (2.1) or moral dimensions (2.2), as codes to text passages. The detection of frame structures (2.3) is comparable to the basic principles of axial coding in the grounded theory approach, where identified codes and concepts, i.e., frames, are interpretatively contrasted and linked to each other. A tool like FrameFinder can help to get a first impression of the frames used in the text corpora and to decide on the further way of analysis. The frames found can then be integrated into MAXQDA [28] or other qualitative coding software for more in-depth analysis. However, social researchers need to consider the pre-defined labels and dimensions that underlie this tool in order to interpret and extend their manual analyses accordingly. The adoption of frame detection tools such as FrameFinder in social science research will depend on the choice of underlying framing concepts and their adaptability to various contexts and research goals.

[3]For details of the node and edge types refer to the guidelines: `https://github.com/amrisi/amr-guidelines/blob/master/amr.md`

GVFC Themes	# Events	# Issues
Total headlines	1269	1339
Economic consequences	3	92
Gun control/regulation	16	306
Gun/2nd Amendment rights	7	59
Mental health	51	29
Politics	32	401
Public opinion	18	244
Race/ethnicity	84	50
School or public space safety	28	156
Society/culture	4	44
Total labels	243	1381

Table 1: Statistics of the annotated GVFC.

3 DEMONSTRATION WITH THE GVFC

To demonstrate the merits of the framing extraction tool, we analyze the gun violence frame corpus (GVFC) [17]. The corpus consists of 2990 news headlines about gun violence in the United States. Figure 3 shows the results extracted with FrameFinder[4].

Models and Configuration. In the code, the models and their configuration can be adapted before computation. For the analysis of the GVFC, we use the same configuration (i.e., definitions of labels and dimensions), as well as models that are deployed in the online demonstration for consistency's sake. We choose three popular models, together with the well-established labels of the media frame corpus as labels and moral foundation theory as dimensions.

For the frame label extraction, we use a zero-shot classification model based on BART [15], i.e., *facebook/bart-large-mnli*. For zero-shot labels, we used the 14 specific media frames (and 1 unspecific other category) defined by their keyword list in [6].

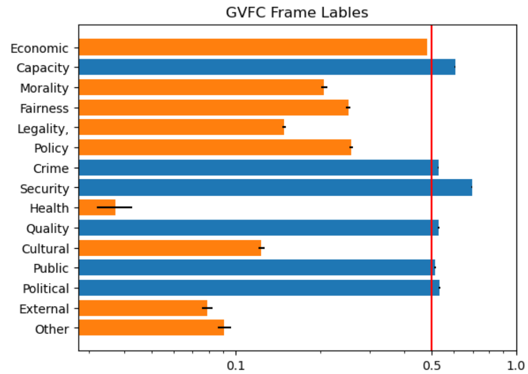
For the frame dimensions, we use an encoder model based on MPNet [29], i.e., *sentence-transformers/all-mpnet-base-v2*. For the poles of the dimensions, we use the instructions from the moral foundation dictionary [9] (i.e., version two) of the five axes: harm/care, cheating/fairness, betrayal/loyalty, subversion/authority, and degradation/sanctity.

For the frame structure, we use another BART-based based model trained on abstract meaning representations (AMR) [4], i.e., *model_parse_xfm_bart_base-v0_1_0*[5]. We set the threshold for nodes to 300 and only plot the largest weakly connected component together with another zoomed-in version using a threshold of 1000.

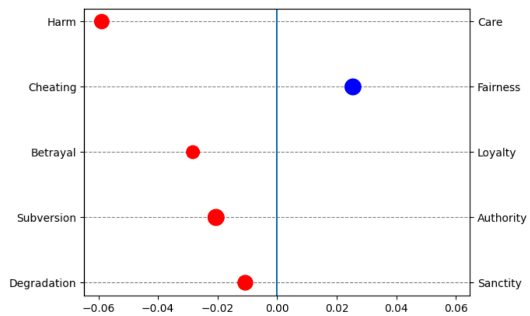
Framing Analysis. From Figure 3a, we observe that gun violence headlines are mostly framed from a security viewpoint. Other important frames are about resources (i.e., capacity), crime, quality of life, public opinion, as well as political frames. In comparison, it is not seen as a health issue. Interestingly, there appears to be an

[4]Note that while the results were computed using the same underlying code, for efficiency, we extracted the frames using a GPU rather than using the free CPU-only online demo interface.

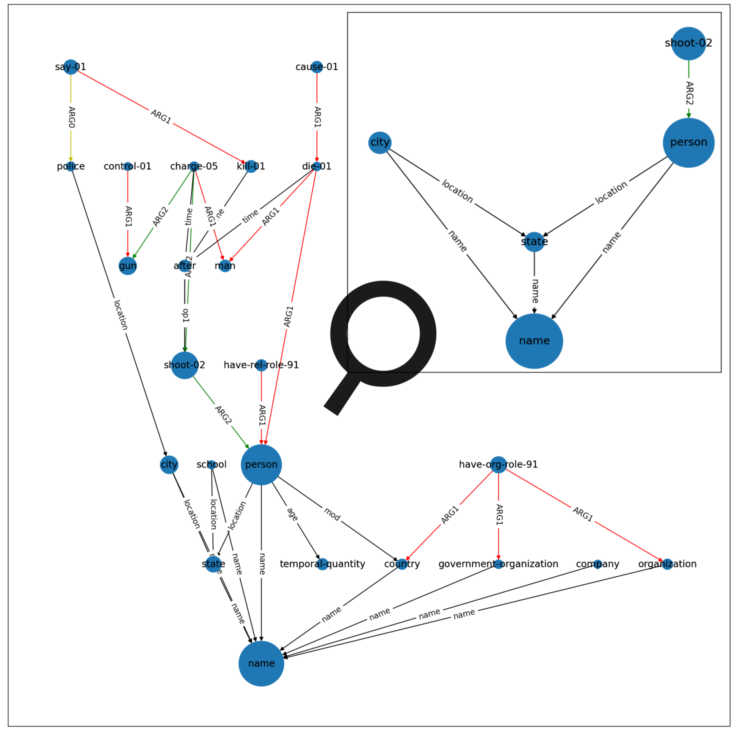
[5]The model can be downloaded from the AMRlib model GitHub repo: `https://github.com/bjascob/amrlib-models`



(a) Frame Labels



(b) Frame Dimensions



(c) Frame Structure (with zoomed-in substructure at the top right).

Figure 3: Framing visualizations of the GVFC.

absence of certain frames regarding morality and fairness, which can be investigated with the framing dimensions.

From a moral standpoint (see Figure 3b), it revolves most about the harm caused. Overall, the moral framing is rather negative with betrayal, subversion, and degradation residing on the vice side. In contrast, the fairness frame is the only positive (i.e., virtue) frame invoked in the headlines. While the bias of the frames differs noticeably (i.e., regarding their positions), the differences in intensity (i.e., point size) are much less pronounced. In this regard, the subversion/authority axis appears to be more emphasized compared to the betrayal/loyalty axis. While we clearly observe differences, with the overall negativity and fairness being less biased compared to harm, it shows that these moral values are of lesser concern when framing the news headlines.

Considering the structural view of the arguments (i.e., Figure 3c) shows that, while complex in nature, the headlines have a common theme. Specifically, as shown in the zoomed-in version, the headlines typically refer to the name of the victim of the shooting rather than the shooter (which is specified by the ARG2 role). Noteworthy is that guns and police have a subordinate role in the headlines.

To summarize, gun violence headlines frame the topic as a security issue that causes harm, with specific persons, such as the victims (mentioned by their names), being a focal point.

Comparison to GVFC Annotations. Here, we compare our results with the ground truth labels of the GVFC. In GVFC, headlines can either be assigned to singular events/incidents or issues of gun violence as an ongoing problem. Additionally, each headline gets assigned zero to two labels that determine the theme of the news story. We provide an aggregated overview in Table 1 (refer to [17] for further details). Both types (i.e., events and issues) appear roughly equally, but issues are far more often associated with labels.

This highlights a limitation of non-exploratory framing analysis, which involves first creating a codebook and then applying it to a corpus. Our use of FrameFinder reveals that the corpus often emphasizes the victims in event headlines. Similar to the annotations, we observe that politics and public opinion are common themes, while mental health gets neglected. In sum, while the annotations and findings from the exploratory framing analysis using FrameFinder largely align, the latter offers additional insights, e.g., emphasis on victims, which was not explicitly annotated in the corpus.

4 CONCLUSION

Framing analysis is intrinsically explorative and spans multiple disciplines. To advance research in this complex field, we present FrameFinder: an *explorative multi-perspective framing extraction* tool. Our user-friendly online demo offers insights into three distinct types of framing present in a text.

Currently, FrameFinder is designed to serve as a support tool for social science researchers. However, we recommend extending its application to information retrieval systems in future work. With media biases being a societal concern [12], we advocate for the development of more refined automatic models for media analysis.

REFERENCES

[1] Abubakar Abid, Ali Abdalla, Ali Abid, Dawood Khan, Abdulrahman Alfozan, and James Zou. 2019. Gradio: hassle-free sharing and testing of ml models in the wild. *arXiv preprint arXiv:1906.02569*.

[2] Mohammad Ali and Naeemul Hassan. 2022. A survey of computational framing analysis approaches. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 9335–9348.

[3] Leif Azzopardi. 2021. Cognitive biases in search: a review and reflection of cognitive biases in information retrieval. In *Proceedings of the 2021 conference on human information interaction and retrieval*, 27–37.

[4] Laura Banarescu et al. 2013. Abstract meaning representation for sembanking. In *Proceedings of the 7th linguistic annotation workshop and interoperability with discourse*, 178–186.

[5] Vibhu Bhatia et al. 2021. Openframing: open-sourced tool for computational framing analysis of multilingual data. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 242–250.

[6] Dallas Card, Amber Boydstun, Justin H Gross, Philip Resnik, and Noah A Smith. 2015. The media frames corpus: annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 438–444.

[7] Tim Draws, Nava Tintarev, Ujwal Gadiraju, Alessandro Bozzon, and Benjamin Timmermans. 2021. This is not what we ordered: exploring why biased search result rankings affect user attitudes on debated topics. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 295–305.

[8] Robert M Entman. 1993. Framing: toward clarification of a fractured paradigm. *Journal of communication*, 43, 4, 51–58.

[9] Jeremy Frimer, Jonathan Haidt, Jesse Graham, M Dehghani, and Reihane Boghrati. 2017. Moral foundations dictionaries for linguistic analyses, 2.0. *Unpublished Manuscript. Retrieved from:* `www.jeremyfrimer.com/uploads/2/1/2/7/21278832/summary.pdf`.

[10] Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational linguistics*, 28, 3, 245–288.

[11] Jonathan Haidt and Craig Joseph. 2004. Intuitive ethics: how innately prepared intuitions generate culturally variable virtues. *Daedalus*, 133, 4, 55–66.

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

[13] Elise Jing and Yong-Yeol Ahn. 2021. Characterizing partisan political narrative frameworks about covid-19 on twitter. *EPJ data science*, 10, 1, 53.

[14] Haewoon Kwak, Jisun An, Elise Jing, and Yong-Yeol Ahn. 2021. Frameaxis: characterizing microframe bias and intensity with word embedding. *PeerJ Computer Science*, 7, e644.

[15] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.

[16] Qisheng Liao, Meiting Lai, and Preslav Nakov. 2023. Marseclipse at semeval-2023 task 3: multi-lingual and multi-label framing detection with contrastive learning. *arXiv preprint arXiv:2304.14339*.

[17] Siyi Liu, Lei Guo, Kate Mays, Margrit Betke, and Derry Tanti Wijaya. 2019. Detecting frames in news headlines and its application to analyzing news framing trends surrounding us gun violence. In *Proceedings of the 23rd conference on computational natural language learning (CoNLL)*, 504–514.

[18] Philipp Mayring. 2015. *Qualitative Inhaltsanalyse. Grundlagen und Techniken.* (12th ed.). Beltz Verlagsgruppe, Weinheim, Germany. ISBN: 9783407293930.

[19] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

[20] Negar Mokhberian, Andrés Abeliuk, Patrick Cummings, and Kristina Lerman. 2020. Moral framing and ideological bias of news. In *Social Informatics: 12th International Conference, SocInfo 2020, Pisa, Italy, October 6–9, 2020, Proceedings 12*. Springer, 206–219.

[21] Fred Morstatter, Liang Wu, Uraz Yavanoglu, Stephen R Corman, and Huan Liu. 2018. Identifying framing bias in online news. *ACM Transactions on Social Computing*, 1, 2, 1–18.

[22] Jakub Piskorski, Nicolas Stefanovitch, Giovanni Da San Martino, and Preslav Nakov. 2023. Semeval-2023 task 3: detecting the category, the framing, and the persuasion techniques in online news in a multi-lingual setup. In *Proceedings of the the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, 2343–2361.

[23] Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

[24] Markus Reiter-Haas. 2023. Exploration of framing biases in polarized online content consumption. In *Companion Proceedings of the ACM Web Conference 2023*, 560–564.

[25] Markus Reiter-Haas, Alexander Ertl, Kevin Innerebner, and Elisabeth Lex. 2023. Mcpt at semeval-2023 task 3: multilingual label-aware contrastive pretraining of transformers for few-and zero-shot framing detection. *arXiv preprint arXiv:2303.09901*.

[26] Markus Reiter-Haas, Simone Kopeinik, and Elisabeth Lex. 2021. Studying moral-based differences in the framing of political tweets. In *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 15, 1085–1089.

[27] Dietram A Scheufele and David Tewksbury. 2007. Framing, agenda setting, and priming: the evolution of three media effects models. *Journal of communication*, 57, 1, 9–20.

[28] VERBI Software. 2021. Maxqda 2022 [computer software]. berlin, germany: verbi software.

[29] Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. 2020. Mpnet: masked and permuted pre-training for language understanding. *Advances in Neural Information Processing Systems*, 33, 16857–16867.

[30] Jörg Strübing. 2014. *Grounded Theory. Zur sozialtheoretischen und epistemologischen Fundierung eines pramatistischen Forschungsstils.* (3rd ed.). Springer VS, Wiesbaden, Germany. ISBN: 9783531198965.

[31] Karen Sullivan. 2023. Three levels of framing. *Wiley Interdisciplinary Reviews: Cognitive Science*, e1651.

[32] Amos Tversky and Daniel Kahneman. 1981. The framing of decisions and the psychology of choice. *science*, 211, 4481, 453–458.

[33] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

[34] Thomas Wolf et al. 2019. Huggingface's transformers: state-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*.

[35] Ben Wu, Olesya Razuvayevskaya, Freddy Heppell, João A Leite, Carolina Scarton, Kalina Bontcheva, and Xingyi Song. 2023. Sheffieldveraai at semeval-2023 task 3: mono and multilingual approaches for news genre, topic and persuasion technique classification. In *Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023)*, 1995–2008.