

Public Communication on the Energy Transition - a Natural Language Processing Approach for Analyzing News Data

Mareike Petrosjan, Dinesh Korrapati, Christin Hoffmann, Felix Müsgens, BTU Cottbus-Senftenberg, Cottbus, Germany
 petrosja@b-tu.de, korradin@b-tu.de, hoffmanc@b-tu.de, felix.muesgens@b-tu.de

Abstract

Global mitigation targets as defined by the Paris Agreement can only be achieved through a fast and efficient energy transition. Pace and efficiency of the energy transition are strongly determined by its public acceptance. However, as the climate crisis itself, as well as one of its main driving solutions - the energy transition- is a complex, transformative, and politically diverse topic, communication about it is at risk of knowledge gaps, misinformation, and emotional polarization. Individual preferences are likely influenced and updated by any additional information received. Consequently, neutral and informative communication about the energy transition is becoming increasingly important, particularly in fostering public acceptance. We apply natural language processing (NLP) techniques to analyse public communication on the energy transition, focusing on sentiment analysis of transcripts from Germany's most-watched news program, *Tagesschau*. We address the challenge of efficiently detecting tone and emotion in this domain-specific context by offering a fine-tuned model that enables sentiment classification of texts related to the energy transition.

1 Introduction

Climate change constitutes one of the most pressing environmental, economic, and societal challenges of the present era [1]. Addressing mitigation targets, policymakers worldwide develop energy strategies [2], which encompass the large-scale adoption of renewable energy, improvements in energy efficiency, and the related adjustments in energy distribution, storage, and consumption patterns, collectively referred to as the energy transition [3]. Internationally, Germany is recognized as a pioneer in restructuring its energy system to incorporate an increasing share of renewable energy [4], [5]. To address the associated challenges, public acceptance plays a crucial role in the successful and rapid deployment of emerging energy technologies [6]. The transition demands substantial investment, labour market shifts, landscape interventions and broader societal transformation [7].

Given the complexity and transformative nature of the energy transition, substantial knowledge gaps exist, and individual preferences are continually shaped by new information received [8]. However, an efficient and effective energy transition depends on public acceptance that can be fostered by clear, neutral, and effective communication. [9]. On the contrary, limited understanding and emotional polarization may hinder well-informed decisions [10] that result in higher costs and more time needed for the realization of this transformation project.

Within the field of communication, the media plays an important role. Public opinion, political priorities, and media narratives influence each other [11]. While the media closely reacts to political debates [12], it can also elevate specific issues by selecting topics and framing them in particular ways, thereby shaping public attention and opinions

[13]. In the context of the energy transition, media coverage can influence whether emerging technologies are facilitated or constrained [14].

Previous studies have mainly focused on newspaper portrayals of the German energy transition, examining aspects as scenario modelling communication [15] or international differences in reporting [16]. Applying topic modeling and sentiment analysis on coverage of the Renewable Energy Act (EEG), [2] could show a shift from positive to cost-focused critical framings. However, broader analyses beyond newspapers in Germany remain limited.

We focus our analysis on the *Tagesschau* as a high-reach news source that is accessed via traditional television broadcasting and digital devices, recognizing its role as a key medium for information acquisition [17] and as a channel for science and environmental communication [18]. Our analysis is focusing on the energy transition media coverage of Germany's leading TV news program *Tagesschau* [19]. Some studies have analysed TV content related to aspects or specific dimensions of the energy transition. [20] analyse the climate change coverage in *Tagesschau* using neural topic modelling. For instance, [17] focuses on the TV news coverage of wind energy on three different American news channels.

In this study, we contribute to the literature by (1) addressing the research gap in the topic of public communication about the energy transition providing a sentiment analysis of the German TV format *Tagesschau* and (2) introducing a novel domain-specific model specialized on energy-related text data. Summarizing, three guiding questions structure our research:

- How is the energy transition communicated in the German news media debate over time?
- Which topics related to the energy transition dominate the news media discourse in Germany?

- Are there patterns of sentiment regarding specific topics of energy-related news?

For analysing large text corpora, Natural Language Processing (NLP) techniques have proven to be an effective and economical method compared to manual content analysis [21]. In order to better understand the emotional content of texts, sentiment analysis is a classical tool of use [22], which we will apply to our dataset.

2 Method

To understand the public communication about the energy transition, we analyse transcripts of the German newscast *Tagesschau*, which were prepared for natural language processing. We developed the domain-specific sentiment model E3M through a human–AI annotation workflow and a stacking ensemble approach. An overview of the NLP-Workflow guiding through the steps of pre-processing, filtering, sentiment analysis, and topic modelling is described in the following sections.

2.1 Tagesschau Data

The datasets used in this study consist of complete transcripts of the German news program *Tagesschau*. *Tagesschau* is a flagship newscast of Germany’s public broadcasters, financed through mandatory household contributions, and is aired daily at 8 p.m. with a duration of 15 minutes and can be accessed online at any time. With an average audience of approximately 9.5 million viewers per day (as of 2024), it is the most-watched news program in Germany, with a market share of TV news of about 40.5 %. Among younger audiences, the program likewise maintains the highest viewership share [19]. A distinctive feature of *Tagesschau*, compared to international news formats, is its widespread perception as highly objective [23], [24].

2.2 Pre-processing

The dataset used in this study covers the period from October 30, 2007, to October 29, 2022, and continues with a second segment spanning July 12, 2023, to March 20, 2025. Program data were extracted from the publicly accessible *Tagesschau* archive using automated methods, with the first segment being provided through a dataset shared by [25]. For detailed analysis, the data were divided into one-minute segments. In a subsequent step, the audio tracks were extracted from the video files and automatically transcribed. For the second dataset, archived material containing program subtitles was accessed again.

The resulting raw dataset consists of unpunctuated transcribed speech. The absence of grammatical structures and the presence of transcription artefacts posed particular challenges for natural language processing. To address this, the “DeepMultilingualPunctuation” model [26] was applied during preprocessing to restore punctuation and grammatical structure, as it has demonstrated strong performance on German texts as well [27]. For subsequent topic delineation and sentiment analysis, the transcripts

were segmented sentence by sentence and aligned with their respective timestamps, thereby increasing accuracy compared to sentiment values aggregated across full one-minute segments.

2.3 News Story Identification

To extract content relevant to the German energy transition, we applied a simple keyword-based approach comprising more than 200 terms. The keyword set covered domains such as energy sources, policy frameworks, and storage technologies, and was derived from publications of German energy agencies, specialized glossaries, journal articles, and newspapers. Following the keyword filtering, the challenge remained of ensuring that only thematically coherent and continuous reports were retained. To address this, we conducted manual checks for topic relevance and removed sentences that were not related to the topic. A translation into English using Python (Python Software Foundation, version 3.12.11) [28] was added to ensure that non-German speakers could contribute effectively to the process and to apply the models in the following procedure. Each news story, consisting of several sentences, was assigned a unique identification number per year to distinguish it from others during the subsequent steps.

2.4 Sentiment Analysis

The sentiment analysis in this study follows a stacking ensemble approach, wherein multiple base models are integrated through a supervised meta-learner that optimizes their combined predictive performance [29]. Prior research has demonstrated that hybrid deep-learning strategies can outperform single-model architectures by capturing complementary linguistic and contextual nuances across domains [30], [29]. Given the presence of both energy transition-related terminology and the distinct, predominantly objective tone of *Tagesschau* reporting, we expect that combining existing models would yield improved performance for our dataset.

2.4.1 Model Architecture

The developed sentiment framework follows a two-layer stacking structure. At the base layer, three transformer models are processing the translated dataset sentence by sentence: XLM-RoBERTa, a multilingual model, enables robust cross-linguistic sentiment recognition [31]. SoLeimanian Financial-RoBERTa, trained on financial and economic news, captures nuanced polarity shifts within market- and energy-related narratives [32]. CardiffNLP Twitter-RoBERTa is tailored to detect informal, opinion-rich expressions characteristic of social media discourse [33]. Each model produces a SoftMax probability distribution, which expresses the model’s confidence over the sentiment categories (positive, negative and neutral), where the sum of probabilities equals one. Instead of relying on “hard” class labels (e.g., only the most likely class), the ensemble uses these probability vectors, preserving uncertainty and nuanced confidence information from each model. The outputs from the three models (each with three

class probabilities) are concatenated into a nine-dimensional feature vector. Probability-level fusion of this kind has been shown to enhance accuracy and reduce classification noise in hybrid sentiment architectures [30].

At the meta layer, a Logistic Regression classifier functions as the meta-learner, learning optimal weights and interactions among the base-model outputs to produce the final sentiment prediction. Owing to its linear structure and interpretability, logistic regression provides an effective and widely adopted approach for integrating probabilistic features within stacking ensemble frameworks [29].

2.4.2 Training and Testing Data

In the subsequent step, we prepared the training-testing dataset for the sentiment analysis. [34] demonstrated the high potential of ChatGPT for annotation tasks, showing that the model can outperform human annotators in text classification. Similarly, [35] highlighted the use of Natural Language Explanations (NLEs) to support annotation tasks in cases where classifications are not immediately clear. While NLEs can reinforce human judgment, they may also bias annotators if the model's decision is incorrect, which requires careful handling when employing ChatGPT as an annotation support tool.

In our human-AI annotation workflow for the training-testing data, researchers first conducted independent sentiment annotations. ChatGPT then generated parallel annotations (positive/negative/neutral) together with one-sentence explanations. The same prompt was repeatedly used to ensure the consistent reliability of the statements. Researchers were able to review their initial assessments, particularly in cases of uncertainty and revise them with reference to the AI-generated outputs.

The training-testing set was developed to evaluate the accuracy of the newly created stacking ensemble model. Several identified news stories of 2022 were classified and used here as a 70/30 training-testing file. That means, 70% of the sentences were used for training and 30% for testing the model.

All models were implemented using the Hugging Face Transformers library [36] and executed on GPU hardware under fixed random seeds, ensuring reproducibility [37]. The resulting Energy Economic ensemble model, E3M, is designed explicitly for energy-related data and achieved an overall accuracy of 77.78 %, exceeding the average accuracy of the individual base models. In the final step, we are applying this sentiment analysis sentence by sentence to the complete corpus of energy-related content.

2.4.3 Story-Level Sentiment Aggregation

To obtain a sentiment representation at the story level, the individual sentence-level predictions by the E3M model are aggregated into an overall score for each story. Each sentence was first assigned a numerical sentiment value (positive = +1, neutral = 0, negative = -1), enabling quantitative aggregation. For each story s comprising n sentences, the overall story-level sentiment score \bar{S}_s was computed as the arithmetic mean of its sentence-level sentiment values, where v_i represents the sentiment value of sentence

i (-1, 0, or +1), and n is the total number of sentences in story s :

$$\bar{S}_s = \frac{1}{n} \sum_{i=1}^n v_i$$

The resulting continuous score \bar{S}_s was then discretized into categorical labels using fixed thresholds to ensure interpretability:

$$\text{Story Sentiment} = \begin{cases} \text{Positive,} & \text{if } \bar{S}_s > 0.2 \\ \text{Neutral,} & \text{if } -0.2 \leq \bar{S}_s \leq 0.2 \\ \text{Negative,} & \text{if } \bar{S}_s < -0.2 \end{cases}$$

This mean-based aggregation captures the overall polarity of each story while minimizing the impact of sentence-level variations.

2.5 Topic Modelling

To examine both the content and temporal patterns within the text corpus, we apply techniques of automated content analysis, specifically semi-automatic topic modelling, as the next step. Topic modelling is a computational approach designed to uncover main themes within large text corpora [38]. In a recent study, [39] used LLMs to perform topic modeling on social-media discussions related to data-center energy consumption. In this study, keyword-based extraction was integrated with language-model-assisted topic generation to enhance thematic coherence and contextual precision [40].

Each news story in the data set was annotated with three keywords describing the main theme. Here, KeyBERT, a method using BERT embeddings - numerical representations that capture the contextual meaning and usage of words, was applied [41]. Next, the large language model (LLM) Mistral-Instruct model [42], a version of Mistral designed to follow natural language instructions, was prompted to produce a single short label (up to four words) summarizing each story. The prompt included the extracted keywords and the full story text. To ensure consistent and reproducible results, the generation process employed a fixed random seed, a numerical value that guarantees identical outcomes across runs, and deterministic decoding with a temperature of 0. Deterministic decoding means that the model always selects the most probable word at each step, producing the same output for identical inputs and thereby eliminating randomness. The temperature hyperparameter controls the randomness of language model outputs: lower values, such as 0, make the model select the most probable tokens deterministically, resulting in stable and repeatable text generation, while higher values introduce greater variability [37].

In the next stage, we manually consolidated the model-generated topics to produce a more coherent thematic structure. We first established a set of broad thematic categories by conducting an exploratory, year-by-year analysis of the material. Subsequently, we identified consistent subtopics

for each narrative to harmonize terminology and avoid divergent labels for conceptually similar themes. News stories that address multiple themes are classified by the topic that exerts the most decisive influence. Consequently, thematic overlaps may occur, for example, a predefined main topic can be subsumed as a subtopic under another main topic, when its presence within the news story is less than that of the other one. Different main topics can include similar subtopics.

3 Preliminary Results

For presenting preliminary results, we are focusing on the time frame from July 2023 until March 2025, with particular attention to the modelled topics of energy policy and energy costs.

Table 1 shows an overview of these two main topics, including their subtopics, story and sentence count. Within the category of Energy Policy, the number of reports was the highest for energy pricing and political debates, including each eight news stories, followed by legislative measures such as the Gaspreisbremse and Gebäudeenergiegesetz, and industry prices. Several themes, including CO₂ pricing, consumer prices, energy saving, cooperation, Energiepreisbremse, energy law, and international affairs, were each mentioned only once, indicating a comparatively limited presence in the overall reporting.

The number of reports covering the main topic of energy costs is the highest for industry prices, with a story count of 16. Consumer prices and energy pricing are mentioned in two stories, while CO₂ pricing, Energiepreisbremse, Gaspreisbremse and Hospital Financial Crisis are mentioned each one time. Table 1 is showing the sentence count of each subtopic adding specific weight on the story count.

Main Topic	Subtopics	Story Count	Sentence Count
Energy Policy	Energy Pricing	8	132
	Political Debate Energy	8	111
	Gebäudeenergiegesetz	2	60
	Industry Prices	2	37
	Gaspreisbremse	2	29
	Energiepreisbremse	1	28
	Energy Saving	1	12
	Resources	1	11
	Cooperation	1	8
	CO ₂ Pricing	1	7
	Consumer Prices	1	5
	Energy Law	1	3
International	1	2	
Energy Costs	Industry Prices	16	204
	Consumer Prices	2	23
	Energy Pricing	2	13
	Gaspreisbremse	1	26
	Energiepreisbremse	1	13
	CO ₂ Pricing	1	10
	Hospital Financial Crisis	1	7

Table 1 Overview of Main Topics, Subtopics, Story and Sentence Count (Data from July 2023 - March 2025)

Regarding sentiment analysis, Figure 1 presents a combined visualization of monthly sentiment and story frequency for energy policy and energy costs from 2023 to 2025. The x-axis displays the timeline in monthly intervals. The left y-axis represents the mean sentiment scores, ranging from negative to positive values, and is depicted using line and point markers for both thematic categories. The right y-axis indicates the number of news stories published per month, illustrated as grey and hatched bars for energy policy and energy costs, respectively. In the topic of energy policy, months with the highest number of reports can be found in August, September, and November 2023, as well as in February 2025. The mean sentiment oscillates around neutral values with intermittent positive and negative peaks. In contrast, sentiment related to energy costs is consistently more negative, with several months showing pronounced downturns. Months with the highest number of reports here are August and September 2023.

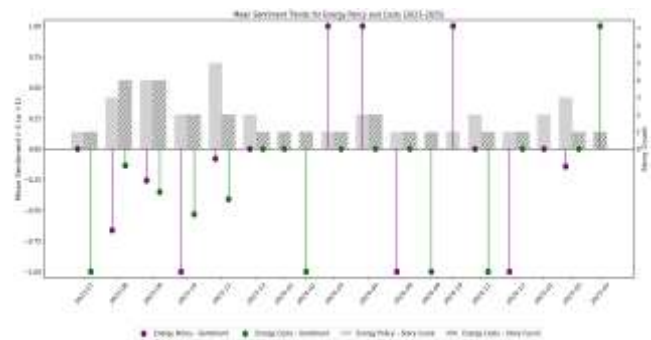


Figure 1 Sentiment Trends and Story Count on Energy Policy and Energy Costs

4 Discussion and Outlook

Our research approach offers particular value for scholars and data scientists applying quantitative methods in content analysis. In traditional manual approaches, assessing the emotional valence of topics through close reading is highly labour-intensive and time-consuming.

This study is subject to several limitations, which should be acknowledged. First, transcription errors may occur, but the large size of the dataset helps to mitigate the influence of individual inaccuracies. Second, sentiment analysis faces challenges in capturing subtle forms of expression such as implicit criticism, irony, or sarcasm; however, these cases are relatively infrequent in television news and thus unlikely to affect the broader patterns observed. Third, the reliance on television news data may introduce a slight bias toward neutrality, yet this also reflects the generally balanced reporting style of mainstream broadcasters. Finally, the analysis is restricted to a single national news channel, which ensures consistency and comparability; future research can extend the approach to regional channels for broader insights.

Acknowledgements

This research is supported by the German Federal Government, the Federal Ministry of Research, Technology and Space and the State of Brandenburg within the framework of the joint project EIZ: Energy Innovation Center (project numbers 85056897 and 03SF0693A) with funds from the Structural Development Act (Strukturstärkungsgesetz) for coal-mining regions. The paper benefited from helpful comments by participants of the Workshop on the Psychology of Climate Risk Communication Centre for Decision Research at the University of Leeds in September 2025. We thank our student assistants for their valuable support in conducting the manual content filtering.

5 Literaturverzeichnis

- [1] R. C. Rial, „Biofuels versus climate change: Exploring potentials and challenges in the energy transition,“ *Renewable and Sustainable Energy Reviews*, Bd. 196, 2024.
- [2] J. Dehler-Holland, K. Schumacher und W. Fichtner, „Topic Modeling Uncovers Shifts in Media Framing of the German Renewable Energy Act,“ *Patterns*, Bd. 2, Nr. 1, 2021.
- [3] IPCC, „IPCC special report on renewable energy sources and climate change mitigation,“ *Renewable Energy*, Bd. 20, Nr. 11, 2011.
- [4] R. Kunzig, „Energiewende—Vorbild Deutschland. In: *National Geographic*,“ 9 November 2017. [Online]. Available: <https://nationalgeographic.de/umwelt/2017/11/energiewende-vorbild-deutschland/>. [Zugriff am 22 September 2025].
- [5] S. Strunz, „The German energy transition as a regime shift,“ *Ecological Economics*, Bd. 100, pp. 150-158, 2014.
- [6] P. Emmerich, A.-G. Hülemeier, D. Jendryczko, M. J. Baumann, M. Weil und D. Baur, „Public acceptance of emerging energy technologies in context of the German energy transition,“ *Energy Policy*, Bd. 142, Nr. 111516, 2020.
- [7] M. Pastukhova und K. Westphal, „Governing the global energy transformation,“ *The geopolitics of the global energy transition*, pp. 341-364, 2020.
- [8] J. J. Häußermann, M. Maier und M. Schraudner, „Social acceptance of green hydrogen in Germany: building trust through responsible innovation,“ *Energy, Sustainability and Society*, Bd. 13, Nr. 22, 2023.
- [9] D. Arlt, C. Schumann und J. Wolling, „What does the public know about technological solutions for achieving carbon neutrality? Citizens' knowledge of energy transition and the role of media,“ *Frontiers in communication*, Bd. 8, Nr. 1005603, 2023.
- [10] F. De Silva-Schmidt, *Wissenserwerb zum Thema Klimapolitik aus Mediennutzung und interpersoneller Kommunikation*, Bd. 31, Ilmenau: Universitätsverlag, 2021, pp. 617-633.
- [11] M. E. McCombs, D. L. Shaw und D. H. Weaver, „New directions in agenda-setting theory and research,“ *Mass Communication and Society*, Bd. 17, pp. 781-802, 2014.
- [12] W. L. Bennett, „Toward a Theory of Press-State,“ *Journal of communication*, Bd. 40, Nr. 2, pp. 103-127, 1990.
- [13] T. Harcup und D. O’neill, „What is news? News values revisited (again),“ *Journalism studies*, Bd. 18, Nr. 12, pp. 1470-1488, 2017.
- [14] J. Lyytimäki, N. A. P. A. Nygrén und S. Rantala, „Energy transition looming behind the headlines? Newspaper coverage of biogas production in Finland,“ *Energy, Sustainability and Society*, Bd. 8, Nr. 1, p. 15, 2018.
- [15] D. Scheer, „Communicating energy system modeling to the wider public: An analysis of German media coverage,“ *Renewable and Sustainable Energy Reviews*, Bd. 80, pp. 1389-1398, 2017.
- [16] M. Antal und K. Karhunmaa, „The German energy transition in the British, Finnish and Hungarian news media,“ *Nature Energy*, Bd. 3, Nr. 11, pp. 994-1001, 2018.
- [17] S. Gearhart, O. Adegbola und M. Guerra, „Harvesting the Wind: Analyzing Television News Coverage of Wind Energy,“ *Environmental Communication*, Bd. 13, Nr. 7, pp. 943-957, 2019.
- [18] L. Kahlor und P. A. Stout, *Communicating science*, New York: Routledge, 2010.
- [19] „NDR,“ *Norddeutscher Rundfunk*, 2 Januar 2025. [Online]. Available: https://www.ndr.de/der_ndr/presse/mitteilungen/Tagesschau-um-20-Uhr-erreicht-Spitzenwerte-und-bleibt-Deutschlands-meist-gesehene-Nachrichtensendung-,pressemeldungndr24888.html. [Zugriff am 19 September 2025].
- [20] T. Schirmag, J. H. Wedemeyer und L. Wenz, „Neural topic modeling reveals German television’s climate change coverage,“ *Communications Earth & Environment*, Bd. 6, Nr. 1, p. 441, 2025.
- [21] S. Doropoulos, E. Karapalidou, P. Charitidis, S. Karakeva und S. Vologiannidis, „Beyond Manual Media Coding: Evaluating Large-Scale Text Analysis with NLP,“ *Applied Sciences*, Bd. 15, Nr. 14, 2025.
- [22] B. Pang und L. Lee, „Opinion mining and sentiment analysis,“ *Foundations and Trends® in information retrieval*, Bd. 2, Nr. 1, pp. 1-135, 2008.
- [23] „Häufige Fragen zur tagesschau: Gibt es eine Definition, wie eine ARD-aktuell-Sendung sein muss?,“ *ARD*, 2023. [Online]. Available: <https://www.tagesschau.de/ueber-uns>. [Zugriff am 22 September 2025].

- [24] U. Paetzold, „Wie objektiv können Nachrichten sein,“ *Gewerkschaftliche Monatshefte*, Bd. 24, Nr. 8, pp. 488-495, 1973.
- [25] R. Tschötschel, N. Schumann, R. Roloff und M. Brüggemann, „Der Klimawandel im öffentlich-rechtlichen Fernsehen. Inhaltsanalyse der "Tageschau" und des Gesamtprogramms von Das Erste, ZDF und WDR 2007 bis 2022,“ *Media Perspektiven*, Bd. 51, Nr. 12, pp. 574-851, 2022.
- [26] O. Guhr, A.-K. Schumann, F. Bahrmann und H.-J. Böhme, „Github,“ 26 September 2021. [Online]. Available: <https://github.com/oliverguhr/deepmultilingualpunctuation>. [Zugriff am 2025].
- [27] O. Guhr, A.-K. Schumann, F. Bahrmann und H.-J. Böhme, „FullStop: Multilingual Deep Models for Punctuation Prediction,“ 2021.
- [28] Python Software Foundation, „Python: A dynamic, open source programming language,“ [Online]. Available: <https://www.python.org/>. [Zugriff am 2025].
- [29] M. S. Sivri, „Combining sentiment analysis models using stacking ensemble learning techniques on BIST30 stocks,“ *Hendese Teknik Bilimler ve Mühendislik Dergisi*, Bd. 1, Nr. 2, pp. 91-97, 2024.
- [30] K. L. Tan, C. Lee, K. Lim und K. Anbanathen, „Sentiment analysis with ensemble hybrid deep learning model,“ *IEEE Access*, Bd. 10, pp. 103694-103704, 2022.
- [31] A. Conneau, A. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzman und V. Stoyanov, „Unsupervised cross-lingual representation learning at scale,“ *Proceedings of the 58th annual meeting of the association for computational linguistics*, pp. 8440-8451, 2020.
- [32] M. Soleimani, „Board Environmental, Social, and Governance (ESG) Expertise and the Usefulness of ESG Reports,“ 2024.
- [33] F. Barbieri, J. Camacho-Collados, L. Anke und L. Neves, „TweetEval: Unified benchmark and comparative evaluation for tweet classification,“ *Findings of the association for computational linguistics: EMNLP 2020*, pp. 1644-1650, 2020.
- [34] F. Gilardi, M. Alizadeh und K. Maël, „ChatGPT outperforms crowd workers for text-annotation tasks,“ *Proceedings of the National Academy of Sciences*, Bd. 120, Nr. 30, 2023.
- [35] F. Huang, H. Kwak und J. An, „Is chatgpt better than human annotators? potential and limitations of chatgpt in explaining implicit hate speech,“ *Companion proceedings of the ACM web conference*, pp. 294-297, 2023.
- [36] T. Wolf, L. Debut, V. Sanh, J. Chaumond und C. Delangue, „Huggingface's transformers: State-of-the-art natural language processing,“ *arXiv preprint arXiv:1910.03771*, 2019.
- [37] A. Holtzman, J. Buys, L. Du, M. Forbes und Y. Choi, „The curious case of neural text degeneration,“ *arXiv preprint arXiv:1904.09751*, 2019.
- [38] D. M. Blei, Y. Andrew und I. Michael, „Latent directed allocation,“ *Journal of machine Learning research*, Bd. 3, pp. 993-1022, 2003.
- [39] H. Wang, W. Hua, J. Peng und M. Hu, „Public sentiment analysis of data center energy consumption using social media data and large language models,“ *Energy and Buildings*, p. 115802, 2025.
- [40] Y. Zhu, H. Yuan, S. Wang, J. Liu und C. Deng, „Large language models for information retrieval: A survey,“ *ACM Transactions on Information Systems*, 2023.
- [41] M. Grootendorst, „KeyBERT: Minimal keyword extraction with BERT,“ 2020.
- [42] A. Jiang, A. De Wynter, N. Lambert und C. Raffel, „The curious case of neural text degeneration,“ *Proceedings of the International Conference on Learning Representations (ICLR)*, 2020.
- [43] N. Masaling und D. Suhartono, „Utilizing RoBERTa and XLM-RoBERTa pre-trained model for structured sentiment analysis,“ *INTERNATIONAL JOURNAL*, Bd. 13, Nr. 3, p. 410, 2024.
- [44] H. Kaplan, R.-P. Mundani, H. Rölke und A. Weichselbraun, „CrudeBERT: applying economic theory towards fine-tuning transformer-based sentiment analysis models to the crude oil market,“ *arXiv preprint arXiv:2305.06140*, 2023.
- [45] P. Zhang, J. Rojcek und M. Leippold, „EconBERT: A Large Language Model for Economics,“ Available at SSRN 5263616, 2025.
- [46] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, É. Grave, M. Ott, L. Zettlemoyer und V. Stoyan, „Facebook AI, Hugging Face,“ 26 September 2019. [Online]. Available: <https://huggingface.co/FacebookAI/xlm-roberta-base>. [Zugriff am 2025].