Monitoring Narratives About the Energy Transition in Germany

Social (In)Justice in the Energy Transition — From the Digital Debate to the Living World

Jonas Rieger[1], Lars Grönberg[1], Carmen Loschke[2] and Sibylle Braungardt[2]

[1]Department of Statistics, TU Dortmund University, 44221 Dortmund, Germany
[2]Oeko-Institut, Merzhauser Str. 173, 79100 Freiburg, Germany

Abstract

The energy transition as a global challenge is widely discussed in Germany and shaped by different narratives. We investigate how newspapers, comments and social media posts differ between sources, platforms, and entities to visualize and monitor respective indicators of the debate over time. For this purpose, we propose a framework of an unsupervised and automated topic model-based approach, the results of which are visualized in a dashboard in conjunction with results from a reliable and supervised classification model using parameter-efficient fine-tuning methods and language models. The analysis of the classification results is refined by integrating the information on annotator disagreement which is captured in a new human-coder labeled sample of our text corpus on the German energy transition. We place a particular emphasis on highlighting the issue of social (in)justice in the debate on energy transition.

Keywords

narrative, energy transition, topic model, parameter-efficient fine-tuning, annotation

1. Introduction

The energy transition is a critical global issue, addressing environmental, economic, and social challenges that affect and involve everyone. As a transformative process aimed at shifting from fossil fuels to renewable energy, it has become a central topic in (social) media and public debates. The respective coverage not only reflects public opinion but also plays a significant role in shaping it, influencing how society perceives and engages with the transition. Understanding and tracking these discussions is therefore essential to support informed, inclusive, and effective energy policy development.

However, the vast volume of journalistic texts, social media comments, and public opinions makes it nearly impossible for the general public to fully comprehend the scope of the ongoing debate. To address this challenge, the "Diskurs Energiewende" (energy transition debate) project seeks to develop tools that provide an automated, continuously updated summary of the debate, along with a comprehensive and reliable retrospective analysis. These tools aim to shed light on the dynamics of the energy transition debate and will be made publicly available to foster broader understanding and dialogue.

The development of these tools integrates expertise in narrative analysis and energy transition research, with a particular focus on Germany, where the energy transition is a highly contested and extensively debated topic. Following the heated discussions around the German heating energy law in 2023 [1], key issues such as the social justice implications of the transition have become central to the public debate, which is frequently emotionally charged and shaped by diverse actors, whose agendas and influence often steer the direction of public opinion.

A primary goal of the "Diskurs Energiewende" project is to investigate the drivers, key actors, and external shocks that shape the energy transition debate. By examining these factors, the project seeks to provide insights into how narratives evolve and impact public perceptions, ultimately supporting

In: R. Campos, A. Jorge, A. Jatowt, S. Bhatia, M. Litvak (eds.): Proceedings of the Text2Story'25 Workshop, Lucca (Italy), 10-April-2025

rieger@statistik.tu-dortmund.de (J. Rieger); lars.groenberg@tu-dortmund.de (L. Grönberg); c.loschke@oeko.de (C. Loschke); s.braungardt@oeko.de (S. Braungardt)
0000-0002-0007-4478 (J. Rieger); 0009-0002-0639-1121 (C. Loschke); 0000-0002-2966-7656 (S. Braungardt)

more equitable and inclusive policy development in the energy transition. As a complementary and expanding approach to Loschke et al. [2], which is limited to data from the platform X on the topic of German energy saving in the context of the Russian invasion of Ukraine, we cover the entire media debate on the German energy transition, including a focus on its socially (un)just implementation.

For the automated representation of the debate we will make use of continuous, i.e. updateable topic models, while for the snapshot analysis we will utilize parameter-efficient fine-tuning (PEFT) in combination with few-shot methods on (large) language models (LLMs). Our text corpus will primarily comprises national and regional newspapers, to which we selectively add online comments and social media content.

In the following, we present our methodological workflow, which is based on the findings of Schofield et al. (2015), that "suggest not automating these workflows but adding transparency and easy-to-use tools to support practitioners' context-specific judgments and interventions" [3]. **This paper aims to provide a concrete description of a best practice case study for the continuous and reliable extraction and presentation of narrative elements from real world application scenarios — and to open it up for further discussion and improvement.** To provide a first glimpse of the possibilities of our project, we present preliminary results of the our continuous topic models with integrated change detection as an example analysis. Further elements of the (later) presented framework are currently still under development. The database and the parameters used for the presented results are preliminary as well. However, their heuristic choice illustrates the potential of our methods for an even more sophisticated parameter choice. We provide the programming code, analyses and (permitted) data of the project in the GitHub repository https://github.com/LarsG321/Dissemination-Monitoring-narratives-about-the-energy-transition-in-Germany. We welcome any type of questions, suggestions or criticisms via the issue tracker of the respective repository.

2. Methods

Latent Dirichlet allocation (LDA) [4], estimated with a Gibbs sampler [5], is supposedly (still) the most frequently used topic model variant. In particular in the social sciences [6] it still enjoys great popularity years after the publication of the structural topic model (STM) [7], probably due to the simplicity of application, the modest model assumptions and the meaningful performance, which cannot be outperformed consistently even by neural topic models [8, 9].

After the publication of the seminal paper of the attention algorithm [10] and the resulting performance boost for the corresponding transformer-based language models, BERTopic [11] as a neural topic model (NTM) based on language models (LMs) has also become quite popular. That is likely because it is particularly convenient to use, in fact BERTopic is even more convenient than LDA as it estimates the number of topics data-driven by default. However, it has already been shown in a few analyses that BERTopic performs inferior to Gibbs LDA in terms of several (in particular those based on human judgments) performance measures [8, 9]. In combination with the ongoing scrutiny of existing automated evaluation methods for topic models [12, 13, 14], even if the contextualized topic model [15] works better for well-chosen parameters, overall LDA serves as a robust, stable, reliable universal tool — and many NTMs perform worse in comparison. In our project, since we want to update the resulting topic models continuously, we evaluate which of the two methods RollingLDA [16], including a dedicated monitoring tool to analyze the topic evolutions [17], or BERTrend [18], that is based on BERTopic, is better suited for our use case and select the more suitable model for application.

The research fields of parameter-efficient fine-tuning (PEFT) and mixture of experts (MoE) are very volatile and at the same time strongly overlapping in both aims and architectures, so that a number of publications also explicitly consider combinations. While PEFT in its origin (and name) aims at efficiency and implements this by keeping a large part of the pre-trained parameters fixed, MoE (according to its name) primarily targets performance improvement through input-dependent (de-)activation or weighting of parameters. However, implicitly — or explicitly — the respective other objective is usually achieved as well. In recent years, many innovations, mostly with incremental improvements in certain domains, have been published. The supposedly most well-known method in this area is LoRA [19], which has subsequently undergone various further developments, such as PRILoRA [20], VeRA [21], DoRA [22], RoSA [23] and the quantized — or quantization-aware — versions QLoRA [24] and QA-LoRA [25]. Moreover, several studies were able to show that the combination of PEFT methods in MoE architectures proves effective. These include mixture of vectors (MoV) and mixture of LoRA (MoLoRA) [26] or parameter-efficient routed fine-tuning (PERFT) [27]. Furthermore, it was shown that in MoE PEFT architectures it is advantageous to activate only the PEFT modules that are most relevant for the task, while freezing the other experts for fine-tuning [28].

For our use case, we rely on the well-established PEFT variant LoRA, which has been proven to perform well across diverse tasks. In fact, we compare LoRA to its quantized version QLoRA and combine both with the language modeling head PETapter [29]. Thus, we unite the advantages of PEFT, i.e. efficiency, robustness, modularity, ease of use and sharing, with the additional performance gain by using a few-shot method that integrates the task description into the input utilizing the masked language modeling objective in a PET-alike manner [30]. For the pretrained base (L)LM, we are focusing on German or multilingual models since the corpus is in German language. Options include the models German BERT [31], LeoLM [32], Llama [33], Mixtral/Ministral [34], Gemma [35], Qwen [36], which we want to compare in an appropriate size in an evaluation study and use the best-performing model for the application.

3. Framework

For the design of our framework, we follow a human-in-the-loop best practice [37] with consideration of active-learning strategies for incremental improvement of the classification results. We conduct the evaluation process in a task-based manner [9, 12, 13, 38], so that we do not solely rely on existing NLP benchmarks for the selection of the (L)LM, as their reliability is limited by the inherent potential for data contamination [39].

In Figure 1, our project's methodological workflow is illustrated. As a database, we focus on national and regional German newspapers. We provide a list of example newspapers that we have access to and are considering to include in Appendix A. The resulting corpus is complemented with commentary content from the websites of, e.g., Bild, Spiegel, Süddeutsche Zeitung, Welt, taz, and Zeit, for which we need to evaluate the full public availability of the respective community comments. Further, we collect social media posts utilizing promising software packages for the platforms X [40] and Telegram [41].

Next, we define a filter that primarily aims for a recall, i.e. relevance with regard to the energy transition, close to 100% while maintaining the highest possible precision. In Section 4, we provide an example of a filter for the newspapers Bild, Süddeutsche Zeitung, and Welt. The filtered and preprocessed corpora are modeled using RollingLDA or BERTrend, respectively. Every week, the procedure is repeated with the newly added data. After postprocessing, the results of the modeling are made publicly available in our dashboard on an ongoing weekly basis. For this, we make use of automated topic labeling [42] and highlight hot/cold topics, popular topics per month [cf. 43], as well as a monitoring of the topics' evolution [17] with tailored graphics [44], all of which is implemented in an unsupervised manner. In addition, we present the most important entities (e.g., politicians, companies, countries) in the respective corpus, put them in context using network analysis methods, and report the overall sentiment over time as well as top articles per topic and month. Furthermore, as far as the data (availability) allows, we intend to not only highlight temporal changes but also spatial differences, e.g., among different federal states or regions of Germany [cf. 45]. Besides, we explore the use of LLMs for writing short summaries of topics for the latest analyzed week.

The interpretation of the results is used for the development of the codebook [cf. 46] for supervised classification, which we use for a restrospective snapshot analysis. In the definition of the codebook, we will (need to) focus on the operationalizable definition of what social justice means in order to avoid too much disagreement in the labels due to insufficiently precise definitions [cf. 47, 48, 49]). We aim for the resulting labeled data set to provide evidence for (at least) the following hypotheses:

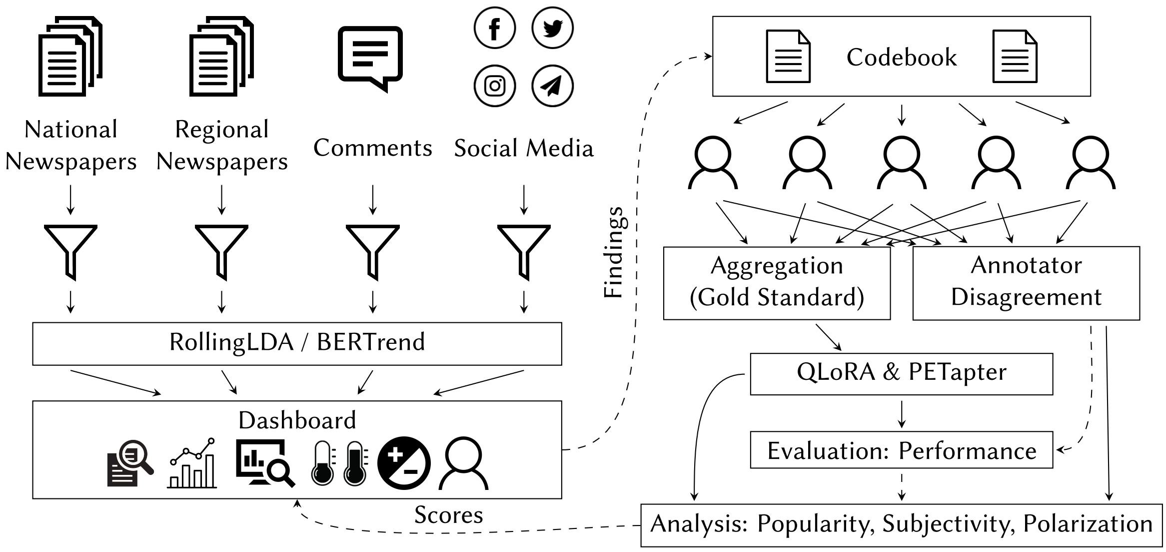

Figure 1: Our project framework for the identification of narrative elements of the German energy transition (individual symbols are from pngegg.com).

- The (media-perceptible) acceptance for the German energy transition dropped in the last two years (2023, 2024).
- The negative aspects regarding the energy transition are disseminated disproportionately often by certain media.
- The argument of social (in)justice is more often used as a counter-argument for the German energy transition as a whole than as an argument for a more socially just implementation of it.

The annotation is done by five independent coders in order to be able to provide not only a (reliable) gold standard label but also topic-based uncertainty estimates through (reasonable) annotator disagreement for our evaluations. In addition to the evaluation of the annotator disagreement [47, 48], we make use of majority vote gold standard labels for fine-tuning [49]. We implement an active learning approach [37] to keep track of and be able to incrementally improve in terms of the model's performance measures. Our architecture uses QLoRA and PETapter, the utility of which has already been proven in real-world social sciences applications [29, 46].

Supposing a sufficiently high annotator agreement, we employ the fine-tuned model based on the codebook-labeled partial dataset to predict and analyze a variety of descriptive characteristics, e.g., polarization and subjectivity scores for topics, arguments, newspapers and media types, respectively, over time. These findings are in turn fed back into our dashboard to provide an comprehensive overview of the German news media debate on the energy transition and its social (in)justice.

4. First Insights

This first analysis is concerned with the left half of our framework in Figure 1 up to the development of the dashboard and presents first results for the RollingLDA methodology.

In order to investigate the German energy transition and its accompanying narratives, we retrieve articles from three major German newspapers Bild, Süddeutsche Zeitung (SZ), and Welt from 2014 to 2024. These outlets are selected to capture a broad ideological and stylistic range: SZ is generally considered left-leaning, Welt is viewed as conservative, and Bild is known for its more opinion-based and sensationalist approach. We apply a pre-filtering step to retrieve only relevant articles, i.e. containing terms explicitly related to the German energy transition. For this, we make use of the keywords

- Bild: (wende | gesetz | politik | habeck | hammer) & (waerme | energie | klima | heiz)
- Süddeutsche Zeitung and Welt: (wende | gesetz | politik) & (waerme | energie | klima | heizung)

In addition, for Bild, we exclude articles belonging to the Sports section according to meta data as well as all articles containing the pattern *leserbriefe* or *seite* in the title since these articles revealed to be noisy and seem to be falsely extracted by the data provider. These steps result in a dataset of 21,046 articles from SZ, 21,107 from Welt, and 2,263 from Bild.

4.1. Topic Modeling

To explore the thematic structure of the filtered articles, we applied RollingLDA [16] for various topic numbers in combination with a change detection procedure [17] to identify changes in topics and the most influential words for these respective changes. The topic models were trained independently on each newspaper corpus. In Appendix B, we provide details about the preprocessing and heuristic hyperparameter choices, i.e. a more sophisticated choice of parameters could lead to even clearer (and better interpretable) results.. As explained for our framework in Figure 1, we employed automated topic labeling leveraging LLMs [42] in combination with subsequent human verification and refinement (where necessary). Among the tested configurations, $K = 10$ appeared most promising in terms of interpretability and thematic clarity across the three corpora.

The basic topic model results, i.e. top ten explanatory words per topic and the monthly topic frequencies, can bee accessed in Appendix C. Within the SZ corpus, the topics covered both the German energy transition itself and international political affairs, with frequent mentions of Ukraine and Russia. Additional thematic clusters included *climate protection and protest* (Topic 4) and *climate change in relation to agriculture* (Topic 2), indicating that the filtering process effectively isolated content directly relevant to the German energy transition. Notably, topics related to sustainability and renewable energies were found consistently across multiple topic number configurations. A similarly broad thematic distribution emerged within the Welt corpus. The identified topics ranged from *Russia-EU relations in the context of the Ukraine conflict* (Topic 8) and *German energy policy* (e.g., topics 1, 4, 6, 7) to *real estate and infrastructure* (e.g., topics 5, 10). As in the SZ corpus, themes tied to sustainability, transportation, and the German energy transition were prevalent and showed consistent patterns across different topic number configurations. Furthermore, comparative analyses between the two newspapers revealed that certain topic clusters — particularly those pertaining to the Ukraine conflict, sustainable energy, and transportation — were present in both corpora, with occasional co-moving patterns in topic prevalence over time. In contrast, the Bild corpus exhibited a more pronounced focus on national politics and prominent public figures. The most frequently identified topics included *German domestic politics in the context of the Ukraine conflict* (Topic 1), *Thuringian state-level politics* (Topic 2) and German politicians and public figures (nearly all topics). This orientation aligns with Bild's reputation for spotlighting high-profile personalities and emphasizing sensationalist content. Despite these differences, Bild's articles also contained relevant discussions on the German energy transition, demonstrating the effectiveness of the pre-filtering approach in capturing pertinent material, even within a more personality-driven news environment.

Overall, the RollingLDA analysis successfully extracted thematically coherent topics that illuminate the debate on the German energy transition. Süddeutsche Zeitung and Welt both covered a broad range of energy- and sustainability-related issues alongside international political affairs, while Bild's coverage skewed more heavily toward national politics and individual political figures. These findings offer an important foundation for understanding how different newspaper orientations shape narratives surrounding the German energy transition, laying the groundwork for deeper investigations into public debate and policy implications in this domain.

4.2. Change Detection

Moving on to the change detection, several meaningful changes were identified. Figure 2 displays some results for the change detection. Blue curves indicate intra-topic similarities over time, while the red curves represent statistically motivated bootstrapped thresholds [17]. That is, vertical lines mark detected changes in time (top row). For illustration, the plots start in 2018 displaying five distinct

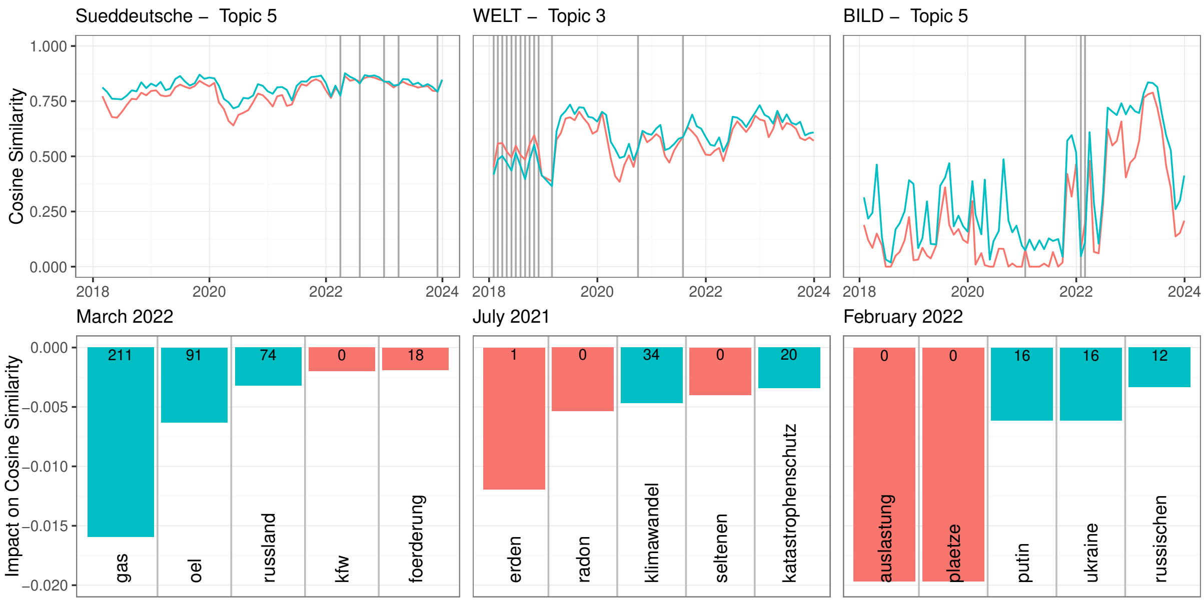

Figure 2: Intra-topic cosine similarity over time and word impacts for a selected change in it [cf. 17]. The first row shows identified changes (vertical lines) for three selected topics from the topic models of SZ, Welt, and Bild, respectively. The second row shows the most influential words for a specific change in the respective topic and visualizes whether the word count increased (blue) or decreased (red) compared to the previous reference period.

changes for topic 5 "German Energy Transition" of SZ. We select the first of those five changes for a more comprehensive analysis, focusing on the most influential words associated with this change (bottom row). Accordingly, it was triggered by an increased usage of terms related to fossil fuels (e.g., *gas*, *oel*) and the word *russland*. This aligns with the start of the Russian invasion of Ukraine in February 2022, which significantly impacted the fossil fuel crisis in Germany. For Welt, the third topic emerged as the most relevant to the energy transition. Here, the most pertinent change occurred in July 2021, attributed to the catastrophic flooding (180 deaths) in the German Ahrtal, which received extensive media attention. The increased frequency of terms such as *klimawandel* (climate change) and *katastrophenschutz* (disaster prevention) supports this interpretation. For Bild, it is noteworthy that the intra-topic cosine similarity are rather low, indicating that topic 5 for Bild, due to the smaller number of articles, demonstrates less stability over time compared to the other two corpora. This volatility leads to a reduced cosine similarity, suggesting a less persistent thematic presence within this corpus. Nonetheless, the method identified the beginning of the Russia-Ukraine conflict as a relevant change, leveraged by an increased usage of the terms *Putin*, *Ukraine*, and *Russischen*.

To summarize the preliminary findings, it becomes evident that the Süddeutsche Zeitung and Welt appear to contain several relevant topics related to the German energy transition, whereas the analysis of Bild reveals challenges due to a relative lack of energy-related topics. In addition, we observed the cross theme of *social injustice* to be barely represented in the current corpus, indicating areas for further refinement and investigation for future analyses.

5. Challenges

Despite our concise approach, there are still several challenges that we will briefly address here. It is likely to be easily possible to classify thematic aspects, to depict sentiment over time, as well as to present key actors in the debate. However, it is still challenging to render entire narratives (over time) beyond individual elements [cf., e.g., 50, 51]. Narratives are complex to extract (even if, by nature, they rely on content simplification) and have several components, whereby there are already many different concepts in the (theoretical) definition and hence also in the (operationalized) extraction.

Regarding data, it should be noted that social media data availability (e.g., X as most prominent example) is increasingly burdening analyses utilizing such data. Furthermore, Instagram, TikTok, but also X contain ever-increasing video and image content, posing the challenge to appropriately combine these different data types for modeling (we limit ourselves to text data). This is already evident for the articles of the Bild newspaper, which, due to their heavy use of visual content, but also due to their use of neologisms (especially in connection with the 2023 Gebäudeenergiegesetz *GEG*, "Buildings Energy Act", which coined the term *Heizungshammer*, "heating hammer"), require special treatment when it comes to preprocessing and keyword searches (cf. Section 4). Further, it is unclear how to adequately take into account different, e.g., age-related, media consumption habits when evaluating perception. In case we get access to regional newspaper data, moreover, we will investigate the regional differences in reporting on the German energy transition, similar to Ozgun and Broekel [45], who were able to prove a strong link between regional characteristics and the amount of news on innovation.

5.1. Next Steps

In our project "Diskurs Energiewende", additionally to the presented framework, we aim to contribute to the development of NLP methods. For RollingLDA, we plan to implement an extension of emerging and fading topics, for which we will evaluate the positive and negative aspects of how BERTrend handles them. This might be especially helpful for the RollingLDA results presented above, since the method fixates the topics in the first iteration. This leads to the effect that non-present topics in later iterations are mapped to topics of the first iteration. Thus, if the narrative of social injustice gets important in the year 2020, for example, and was not present in the first iteration in year 2018 it does not receive its own topic. Further, we intend to develop advancements in the overlapping area of narrative detection using PEFT and the use of advanced learning methods making special use of annotator disagreement. These would allow us to analyze the dissemination and timeline of narratives through different media sources. In turn, it will be challenging to develop a fair performance score to compare models fine-tuned with classical gold standard labels and those incorporating annotator disagreement [cf. 47, 48, 49]).

While the current preliminary analysis is capable of detecting topics with respect to the energy transition, certain limitations with regard to the narrative detection of social injustice apply. To overcome this problem, we will extend the RollingLDA analyses and compare them with other topic model results. We will extend the keyword search to obtain a larger corpus, i.e. a higher recall. In addition, a focused sub-corpus analysis of social injustice related articles within the corpus of the energy transition articles will be conducted. Further, as incremental improvements we will incorporate lemmatization and improve our preprocessing function to exclude non-alphanumeric letters (e.g. "€"), which further normalizes the text corpus and might improve the narrative extraction. Based on the improved corpus, we will evaluate the usefulness of RollingLDA, BERTrend [18] and BERTopic [11] to identify narratives in our data and to implement it in our final publicly available dashboard.

Acknowledgments

This work was funded by the German Federal Ministry for Economic Affairs and Climate Action (Bundesministerium für Wirtschaft und Klimaschutz, BMWK) as part of the joint project *Social (in)justice in the energy transition - from the digital debate to the living world* (project no. 03EI5267B).

References

[1] S. Braungardt, F. Keimeyer, C. Loschke, Is the "heating hammer" hitting energy efficiency policy? Learnings from the debate around the German Buildings Energy Act, in: Proceedings of the eceee summer study, 2024. URL: https://www.oeko.de/fileadmin/oekodoc/3-028-24_Braungardt.pdf.

[2] C. Loschke, S. Braungardt, J. Rieger, What motivates and demotivates energy savings in times of crisis? – An argument mining analysis using X/Twitter data, Energy Efficiency 18 (2025). doi:10.1007/s12053-024-10283-0.

[3] A. Schofield, S. Wu, T. Bayard de Volo, T. Kuze, A. Gomez, S. Sultana, "My very subjective human interpretation": Domain expert perspectives on navigating the text analysis loop for topic models, Proceedings of the ACM on Human-Computer Interaction 9 (2025). doi:10.1145/3701201.
[4] D. M. Blei, A. Y. Ng, M. I. Jordan, Latent Dirichlet allocation, Journal of Machine Learning Research 3 (2003) 993–1022. doi:10.1162/jmlr.2003.3.4-5.993.
[5] T. L. Griffiths, M. Steyvers, Finding scientific topics, Proceedings of the National Academy of Sciences 101 (2004) 5228–5235. doi:10.1073/pnas.0307752101.
[6] D. Ramage, E. Rosen, J. Chuang, C. D. Manning, D. A. McFarland, Topic modeling for the social sciences, in: Workshop on Applications for Topic Models, NIPS, 2009. URL: http://vis.stanford.edu/papers/topic-modeling-social-sciences.
[7] M. E. Roberts, B. M. Stewart, D. Tingley, E. M. Airoldi, The structural topic model and applied social science, in: NIPS-Workshop on Topic Models: Computation, Application, and Evaluation, 2013. URL: https://www.wcfia.harvard.edu/files/wcfia/files/stmnips2013.pdf.
[8] A. M. Hoyle, R. Sarkar, P. Goel, P. Resnik, Are neural topic models broken?, in: Findings of the Association for Computational Linguistics: EMNLP 2022, Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 2022, pp. 5321–5344. doi:10.18653/v1/2022.findings-emnlp.390.
[9] Z. Li, A. Mao, D. Stephens, P. Goel, E. Walpole, A. Dima, J. Fung, J. Boyd-Graber, Improving the TENOR of labeling: Re-evaluating topic models for content analysis, in: Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, St. Julian's, Malta, 2024, pp. 840–859. URL: https://aclanthology.org/2024.eacl-long.51/.
[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, I. Polosukhin, Attention is all you need, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, volume 30, Curran Associates, Inc., 2017. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
[11] M. Grootendorst, BERTopic: Neural topic modeling with a class-based tf-idf procedure, 2022. arXiv:2203.05794.
[12] C. Doogan, W. Buntine, Topic model or topic twaddle? Re-evaluating semantic interpretability measures, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Online, 2021, pp. 3824–3848. doi:10.18653/v1/2021.naacl-main.300.
[13] A. Hoyle, P. Goel, A. Hian-Cheong, D. Peskov, J. Boyd-Graber, P. Resnik, Is automated topic model evaluation broken? The incoherence of coherence, in: Advances in Neural Information Processing Systems, volume 34, Curran Associates, Inc., 2021, pp. 2018–2033. URL: https://proceedings.neurips.cc/paper_files/paper/2021/file/0f83556a305d789b1d71815e8ea4f4b0-Paper.pdf.
[14] D. Stammbach, V. Zouhar, A. Hoyle, M. Sachan, E. Ash, Revisiting automated topic model evaluation with large language models, in: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Singapore, 2023, pp. 9348–9357. doi:10.18653/v1/2023.emnlp-main.581.
[15] F. Bianchi, S. Terragni, D. Hovy, Pre-training is a hot topic: Contextualized document embeddings improve topic coherence, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Association for Computational Linguistics, Online, 2021, pp. 759–766. doi:10.18653/v1/2021.acl-short.96.
[16] J. Rieger, C. Jentsch, J. Rahnenführer, RollingLDA: An update algorithm of Latent Dirichlet Allocation to construct consistent time series from textual data, in: Findings of the Association for Computational Linguistics: EMNLP 2021, Association for Computational Linguistics, Punta Cana, Dominican Republic, 2021, pp. 2337–2347. doi:10.18653/v1/2021.findings-emnlp.201.
[17] J. Rieger, K.-R. Lange, J. Flossdorf, C. Jentsch, Dynamic change detection in topics based on rolling LDAs, in: Proceedings of the Text2Story'22 Workshop, CEUR-WS 3117, 2022, pp. 5–13. URL:

https://ceur-ws.org/Vol-3117/paper1.pdf.
[18] A. Boutaleb, J. Picault, G. Grosjean, BERTrend: Neural topic modeling for emerging trends detection, in: Proceedings of the Workshop on the Future of Event Detection (FuturED), Association for Computational Linguistics, Miami, Florida, USA, 2024, pp. 1–17. doi:10.18653/v1/2024.futured-1.1.
[19] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen, LoRA: Low-rank adaptation of large language models, in: International Conference on Learning Representations, 2022. URL: https://openreview.net/forum?id=nZeVKeeFYf9.
[20] N. Benedek, L. Wolf, PRILoRA: Pruned and rank-increasing low-rank adaptation, in: Y. Graham, M. Purver (Eds.), Findings of the Association for Computational Linguistics: EACL 2024, Association for Computational Linguistics, St. Julian's, Malta, 2024, pp. 252–263. URL: https://aclanthology.org/2024.findings-eacl.18/.
[21] D. J. Kopiczko, T. Blankevoort, Y. M. Asano, VeRA: Vector-based random matrix adaptation, in: The Twelfth International Conference on Learning Representations, 2024. URL: https://openreview.net/forum?id=NjNfLdxr3A.
[22] S.-Y. Liu, C.-Y. Wang, H. Yin, P. Molchanov, Y.-C. F. Wang, K.-T. Cheng, M.-H. Chen, DoRA: Weight-decomposed low-rank adaptation, in: R. Salakhutdinov, Z. Kolter, K. Heller, A. Weller, N. Oliver, J. Scarlett, F. Berkenkamp (Eds.), Proceedings of the 41st International Conference on Machine Learning, volume 235, PMLR, 2024, pp. 32100–32121. URL: https://proceedings.mlr.press/v235/liu24bn.html.
[23] M. Nikdan, S. Tabesh, E. Crnčević, D. Alistarh, RoSA: Accurate parameter-efficient fine-tuning via robust adaptation, in: Proceedings of the 41st International Conference on Machine Learning, ICML'24, JMLR.org, 2024. URL: https://dl.acm.org/doi/10.5555/3692070.3693618.
[24] T. Dettmers, A. Pagnoni, A. Holtzman, L. Zettlemoyer, QLoRA: Efficient finetuning of quantized LLMs, in: Advances in Neural Information Processing Systems, volume 36, Curran Associates, Inc., 2023, pp. 10088–10115. URL: https://proceedings.neurips.cc/paper_files/paper/2023/file/1feb87871436031bdc0f2beaa62a049b-Paper-Conference.pdf.
[25] Y. Xu, L. Xie, X. Gu, X. Chen, H. Chang, H. Zhang, Z. Chen, X. Zhang, Q. Tian, QA-LoRA: Quantization-aware low-rank adaptation of large language models, in: The Twelfth International Conference on Learning Representations, 2024. URL: https://openreview.net/forum?id=WvFoJccpo8.
[26] T. Zadouri, A. Üstün, A. Ahmadian, B. Ermis, A. Locatelli, S. Hooker, Pushing mixture of experts to the limit: Extremely parameter efficient MoE for instruction tuning, in: The Twelfth International Conference on Learning Representations, 2024. URL: https://openreview.net/forum?id=EvDeiLv7qc.
[27] Anonymous, PERFT: Parameter-efficient routed fine-tuning for mixture-of-expert model, 2025. URL: https://openreview.net/forum?id=PPjpGTPG5K.
[28] Z. Wang, D. Chen, D. Dai, R. Xu, Z. Li, Y. Wu, Let the expert stick to his last: Expert-specialized fine-tuning for sparse architectural large language models, in: Y. Al-Onaizan, M. Bansal, Y.-N. Chen (Eds.), Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Miami, Florida, USA, 2024, pp. 784–801. doi:10.18653/v1/2024.emnlp-main.46.
[29] J. Rieger, M. Ruckdeschel, G. Wiedemann, PETapter: Leveraging PET-style classification heads for modular few-shot parameter-efficient fine-tuning, 2024. arXiv:2412.04975.
[30] T. Schick, H. Schütze, Exploiting cloze-questions for few-shot text classification and natural language inference, in: P. Merlo, J. Tiedemann, R. Tsarfaty (Eds.), Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Association for Computational Linguistics, Online, 2021, pp. 255–269. doi:10.18653/v1/2021.eacl-main.20.
[31] B. Chan, S. Schweter, T. Möller, German's next language model, in: D. Scott, N. Bel, C. Zong (Eds.), Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Barcelona, Spain (Online), 2020, pp. 6788–6796. doi:10.

18653/v1/2020.coling-main.598.
[32] B. Plüster, LeoLM: Igniting German-language LLM research, 2023. URL: https://laion.ai/blog/leo-lm/.
[33] Llama Team, The Llama 3 herd of models, 2024. arXiv:2407.21783, (for full author list see arXiv).
[34] Mixtral Team, Mixtral of experts, 2024. arXiv:2401.04088, (for full author list see arXiv).
[35] Gemma Team, Gemma 2: Improving open language models at a practical size, 2024. arXiv:2408.00118, (for full author list see arXiv).
[36] Qwen Team, Qwen2 technical report, 2024. arXiv:2407.10671, (for full author list see arXiv).
[37] J. F. Fung, Z. Li, D. K. Stephens, A. Mao, P. Goel, E. Walpole, A. Dima, J. L. Boyd-Graber, Human-in-the-loop Technical Document Annotation, Technical Report, National Institute of Standards and Technology, Gaithersburg, MD, 2024. doi:10.6028/NIST.TN.2287, NIST TN 2287.
[38] Q. V. Liao, Z. Xiao, Rethinking model evaluation as narrowing the socio-technical gap, 2023. arXiv:2306.03100.
[39] Y. Dong, X. Jiang, H. Liu, Z. Jin, B. Gu, M. Yang, G. Li, Generalization or memorization: Data contamination and trustworthy evaluation for large language models, in: L.-W. Ku, A. Martins, V. Srikumar (Eds.), Findings of the Association for Computational Linguistics: ACL 2024, Association for Computational Linguistics, Bangkok, Thailand, 2024, pp. 12039–12050. doi:10.18653/v1/2024.findings-acl.716.
[40] S. Peeters, Zeeschuimer, 2023. URL: https://github.com/digitalmethodsinitiative/zeeschuimer. doi:10.5281/zenodo.7525702.
[41] P. Kessling, F. V. Münch, tegracli: A convenience wrapper around Telethon and the Telegram API for research purposes, 2023. URL: https://github.com/Leibniz-HBI/tegracli. doi:10.5281/zenodo.8043362.
[42] J. Rieger, F. Peters, A. Fischer, T. Lauer, A. Bittermann, topiclabels: Automated Topic Labeling with Language Models, 2024. URL: https://github.com/PetersFritz/topiclabels. doi:10.32614/CRAN.package.topiclabels, R package version 0.2.0.
[43] A. Bittermann, J. Rieger, Finding scientific topics in continuously growing text corpora, in: Proceedings of the Third Workshop on Scholarly Document Processing, Association for Computational Linguistics, Gyeongju, Republic of Korea, 2022, pp. 7–18. URL: https://aclanthology.org/2022.sdp-1.2/.
[44] C. Krause, J. Rieger, J. Flossdorf, C. Jentsch, F. Beck, Visually analyzing topic change points in temporal text collections, in: Vision, Modeling, and Visualization, The Eurographics Association, 2023. doi:10.2312/vmv.20231231.
[45] B. Ozgun, T. Broekel, The geography of innovation and technology news - an empirical study of the German news media, Technological Forecasting and Social Change 167 (2021) 120692. doi:https://doi.org/10.1016/j.techfore.2021.120692.
[46] J. Rieger, K. Yanchenko, M. Ruckdeschel, G. von Nordheim, K. Kleinen-von Königslöw, G. Wiedemann, Few-shot learning for automated content analysis: Efficient coding of arguments and claims in the debate on arms deliveries to Ukraine, Studies in Communication and Media 13 (2024) 72–100. doi:10.5771/2192-4007-2024-1-72.
[47] X. Wang, B. Plank, ACTOR: Active learning with annotator-specific classification heads to embrace human label variation, in: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Singapore, 2023, pp. 2046–2052. doi:10.18653/v1/2023.emnlp-main.126.
[48] N. Deng, X. Zhang, S. Liu, W. Wu, L. Wang, R. Mihalcea, You are what you annotate: Towards better models through annotator representations, in: Findings of the Association for Computational Linguistics: EMNLP 2023, Association for Computational Linguistics, Singapore, 2023, pp. 12475–12498. doi:10.18653/v1/2023.findings-emnlp.832.
[49] E. Fleisig, R. Abebe, D. Klein, When the majority is wrong: Modeling annotator disagreement for subjective tasks, in: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Singapore, 2023, pp. 6715–6726. doi:10.18653/v1/2023.emnlp-main.415.
[50] A. Piper, R. J. So, D. Bamman, Narrative theory for computational narrative understanding, in: M.-F.

Moens, X. Huang, L. Specia, S. W.-t. Yih (Eds.), Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 2021, pp. 298–311. doi:10.18653/v1/2021.emnlp-main.26.
[51] E. Ash, G. Gauthier, P. Widmer, Relatio: Text semantics capture political and economic narratives, Political Analysis 32 (2024) 115–132. doi:10.1017/pan.2023.8.

A. Potential national and regional newspapers under study

The following newspapers are considered to be taken into account in our analyses:

- national newspapers: Bild, Der Tagesspiegel, Die Welt, Süddeutsche Zeitung, taz — die Tageszeitung
- magazines/weekly newspapers: Der Spiegel, Stern
- regional newspapers: Reutlinger Nachrichten, Schwarzwälder Bote, Stuttgarter Nachrichten, Stuttgarter Zeitung, Südwest Presse
- thematic outlets: EID Energie Informationsdienst, Energie & Management/Powernews.org, Handelsblatt, KI — Kälte Luft Klimatechnik,

B. Preprocessing and hyperparameters

The programming code is publicly available via our GitHub project repository https://github.com/LarsG321/Dissemination-Monitoring-narratives-about-the-energy-transition-in-Germany. Due to license restrictions, we are unable to share the raw data. The following preprocessing steps are conducted for all corpora:

- Filter for texts between 2014-01-01 and 2024-03-31
- Remove duplicated texts
- Remove umlauts, HTML and XML nodes
- Remove special characters (apparently we missed removing the € sign)
- Replaced line breaks with spaces
- Lowercase the text
- Remove German stopwords (leveraging a common and rather conservative list)
- Remove punctuation
- Remove numbers

LDA hyperparameters used for all corpora:

- $K = 5, 6, 7, 8, 9, 10, 15, 20$; `alpha = eta = 1/K`

RollingLDA hyperparameters used for all corpora (if not specified, we use the default parameters from the function `RollingLDA` in the software package `rollinglda` [16]):

- `init = "2014-02-01"` (first month of data)
- `chunks = "month"` (monthly updates)
- `memory = "month"` (one month of memory)
- `memory.fallback = 0, 1, 2` (number of documents used as memory if there are no documents in the “original” memory time period)

Change detection [17] hyperparameters used for all corpora:

- $z = \{1, 2, 3, 4\}$ (maximum number of months in the reference period)
- $q = \{0.80, 0.81, 0.82, 0.83, 0.84, 0.85, 0.86, 0.87, 0.88, 0.89, 0.90\}$ (quantile to determine threshold)

Final parameters for RollingLDA and change detection:

- Bild: `memory.fallback=0`, $K = 10, z = 1, q = 0.85$
- SZ: `memory.fallback=0`, $K = 10, z = 1, q = 0.85$
- Welt: `memory.fallback=2`, $K = 10, z = 1, q = 0.80$

C. Topic model results

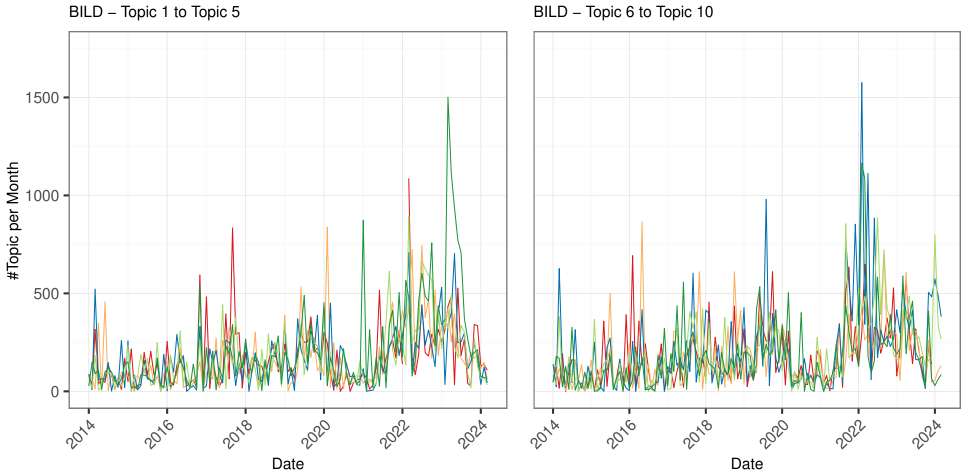

Figure 3: Monthly topic frequencies for Bild corpus

Table 1
Top 10 words per topic for Bild corpus

#	Topic 1: German Politics and Ukraine Conflict	Topic 2: Thuringian State Politics	Topic 3: German Politics and Conflicts	Topic 4: German Politics and Personalities	Topic 5: German Green Economy Policy	Topic 6: German Publishing Companies	Topic 7: German Politics and Health	Topic 8: German Politics and International Relations	Topic 9: Euro, Jobs, and Health	Topic 10: German Politics and Economy
1	afd	euro	luet-zerath	scholz	habeck	stv	muskeln	baerbock	euro	€
2	trump	kem-merich	graichen	kohl	gruene	ltg	merz	scholz	tagestrend	helmut
3	roy	manta	krim	laschet	robert	gmbh	merkel	ampel	jobgeld	tagestrend
4	putin	tina	hitler	habeck	trump	€	cdu	putin	gesund-heit	kohl
5	ukraine	ramelow	stein-meier	robert	wirtschafts-minister	springer	kanzlerin	eingezahlt	liebe	jobgeld
6	rechts	haut	flick	heynckes	waerme-pumpe	digital	neubauer	ukraine	tipp	charles
7	land	schweiger	goetz	olaf	gasheiz-ungen	axel	luisa	tipp	ps	neuinfek-tionen
8	lindner	inflation	syrien	jupp	akw	druckerei	klamroth	olaf	geld	gorbat-schow
9	putins	adresse	frodeno	eier	heiz-hammer	newsprint	erdogan	russland	zaman	euro
10	siegfried	turner	staats-sekretaer	wirtschafts-minister	gesetz	may	landkreis	kuebl-boeck	kosten	rasen

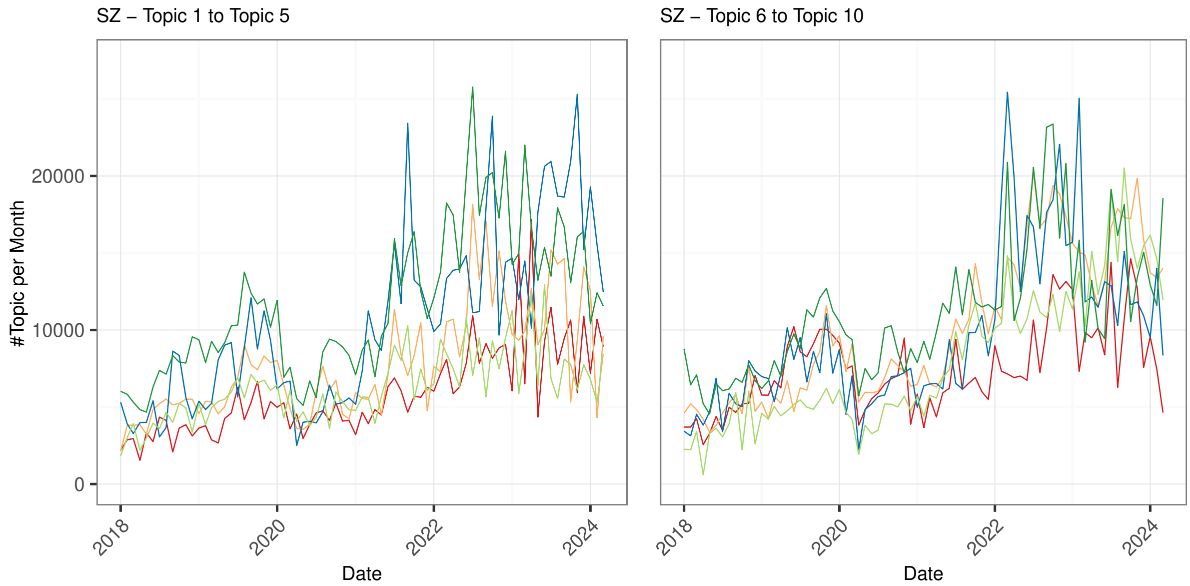

Figure 4: Monthly topic frequencies for SZ corpus

Table 2
Top 10 words per topic for SZ corpus

#	Topic 1: Bayern Munich Soccer	Topic 2: Climate Change and Agriculture	Topic 3: German Politics	Topic 4: Climate Activism	Topic 5: German Energy Transition	Topic 6: Politics and Social Issues	Topic 7: German Economy and Finance	Topic 8: International Politics	Topic 9: Art and Culture	Topic 10: Family and Life
1	trainer	wasser	spd	stadt	strom	trump	unter-nehmen	russland	museum	kinder
2	fc	tiere	gruenen	schueler	euro	biden	euro	eu	kunst	frau
3	spieler	grad	partei	meter	wasser-stoff	polizei	inflation	ukraine	buch	leben
4	wm	pflanzen	soeder	wasser	gas	aktivis-ten	milliar-den	putin	men-schen	mutter
5	fussball	natur	cdu	for	kosten	genera-tion	firmen	china	the	eltern
6	mannschaft	land-wirtschaft	fdp	thunberg	autos	israel	preise	trump	kuenstler	mann
7	saison	erde	scholz	greta	deutsch-land	hamas	wirtschaft	usa	welt	leute
8	bayern	baeume	csu	future	bundes-regierung	demokratie	geld	praesi-dent	leben	familie
9	spiel	arten	koalition	ki	pro	men-schen	ezb	europa	geschichte	frauen
10	sport	klima-wandel	afd	fridays	energien	gesellschaft	banken	russ-ischen	gesellschaft	kind

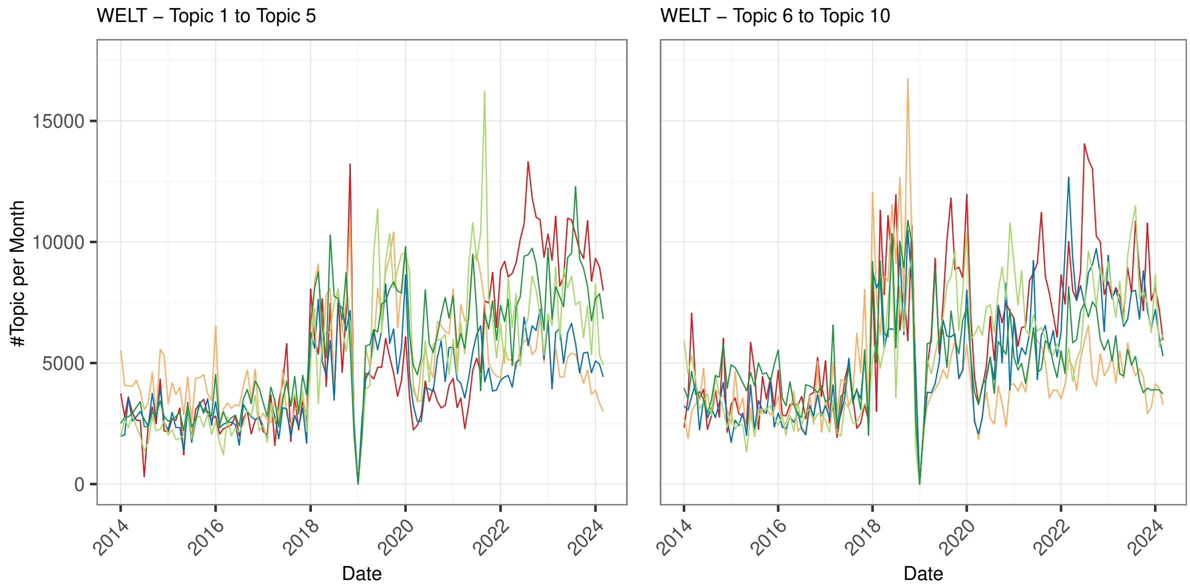

Figure 5: Monthly topic frequencies for Welt corpus

Table 3
Top 10 words per topic for Welt corpus

#	Topic 1: German Politics and Sports	Topic 2: Life and Family	Topic 3: Climate Change and Environment	Topic 4: German Political Parties	Topic 5: Eurozone Inflation and Interest Rates	Topic 6: Russian Energy and Politics	Topic 7: German Politics	Topic 8: Russian-Ukraine Conflict	Topic 9: Social Issues and Politics	Topic 10: EU Economy and Trade
1	seite	leben	wasser	gruenen	inflation	russland	spd	russland	men-schen	eu
2	lesen	mann	forscher	partei	unter-nehmen	trump	partei	ukraine	gesellschaft	milliar-den
3	fc	buch	erde	spd	euro	euro	cdu	biden	leben	euro
4	dax	frau	co	cdu	milliar-den	putin	csu	putin	gibt	unter-nehmen
5	innen-politik	geschichte	grad	afd	jahr	eu	gruenen	trump	kinder	dollar
6	sport	eltern	wissen-schaftler	fdp	wirtschaft	strom	merkel	russis-chen	politik	europae-ischen
7	polizei	mutter	meter	union	ezb	wasser-stoff	afd	china	freiheit	europa
8	angaben	men-schen	klima-wandel	laschet	zinsen	bundes-regierung	seehofer	usa	jungen	china
9	trainer	stadt	holz	scholz	deutsch-land	ukraine	fdp	russische	geht	europae-ische
10	teilte	kinder	pflanzen	koalition	dollar	co	koalition	stream	pan-demie	bruessel