Online popularity and parliamentary decisions: Comparing social media posts of German politicians and voting behavior

Rudy Alexandro Garrido Veliz, Adem Chanie Ali,
Tim Niklas Porth, Martin Semmann, Seid Muhie Yimam,
University of Hamburg
Correspondence: seid.muhie.yimam@uni-hamburg.de

Abstract

Social media has reshaped the democratic standards of Germany by becoming a key channel where citizens form political opinions. Unlike traditional, closely scrutinized platforms, it allows more diverse messaging. Politicians and parties have increasingly used it to gain influence and advantage. However, there is currently no path of verification to the position a German politician may try to portray to their constituents, and the true decision they take in the Bundestag. To alleviate this incongruence, this paper presents a framework that leverages natural language processing, allowing to induce the stance a post takes on a variety of topics that are relevant to the Bundestag voting. We build a new multiplatform dataset of 111.1k posts from 580 politician accounts on TikTok, Bluesky, and Twitter/X, and merge it with 261 Bundestag voting topics from the years 2020–2025. Through topic modeling and sentiment/stance detection, the framework unlocks a direct comparison between their social media activity and voting record. We test many methods for the pipeline, and a combination of TopicGPT and LLM-based stance detection yields the best result, with an accuracy of 64.1% to correctly predict the decision in the Bundestag. We also showcase how the framework can be used to explore the harmony, or lack of, between the social media activity and the votes of the party on the subject of environmental protection. With this tool, we find that there is a considerable distance for most of the leading parties in Germany, between what they try to portray in their online presence and what they choose to vote for in parliament.

1 Introduction

In the digital age, social media has become central to political communication by reshaping how politicians interact with citizens and how public discourse is formed (Lorenz-Spreen et al., 2023). Platforms such as X, TikTok, Bluesky, Facebook, and Instagram enable direct and immediate engagement, often blurring distinctions between information, opinion, and persuasion. Furthermore, their highly personalized and largely unfiltered formats allow politicians to tailor messages to specific audiences, which can affect both political attitudes and democratic stability (Lorenz-Spreen et al., 2023). This "double-edged sword" is reflected in lowered barriers to participation and new forms of mobilization (e.g. MeToo, Fridays for Future), but also in rising polarization, populism, and declining political trust (Lorenz-Spreen et al., 2023).

Concurrently, advances in natural language processing (NLP) provide automated tools for analyzing political text at scale. Topic modeling can extract and structure recurring themes in large corpora, using methods such as Latent Dirichlet Allocation (Blei et al., 2003) or neural approaches like BERTopic (Grootendorst, 2022). Complementing this, stance detection identifies the attitude expressed in a text, e.g., agreement, disagreement, or neutrality toward specific topics (Küçük and Can, 2020). Together, these methods enable systematic investigation beyond manual content analysis. This research examines whether the stances expressed by German Bundestag members on social media (TikTok, X, and Bluesky) correlate with their voting behavior in the Bundestag on corresponding topics.

Building a framework to compare stances across two distinct datasets, the work asks: (**RQ1**) to what extent do the positions stated by Bundestag members on social media correlate with their votes on the same topics? It further addresses (**RQ2**) which unsupervised topic modeling approaches produce the most coherent and suitable topics from parliamentary voting titles and social media posts, and (**RQ3**) how political positions can be automatically inferred from both social media texts and parliamentary voting records.

2 Related Works

Research on German political NLP has so far concentrated mainly on two areas: parliamentary speech analysis and social media analysis. On the parliamentary side, earlier work established corpora and NLP pipelines for digitized Bundestag records, including historical documents, enabling scalable topic and sentiment studies (Abrami et al., 2024). More recent research analyzed roughly 28k Bundestag speeches from 2019 to 2024 and reported that topic trajectories and sentiment distributions depend on both party identity and parliamentary role (Pätz et al., 2025). Overall, this body of work shows that political language use is closely tied to institutional context.

On the social media side, the 2021 Bundestag election has attracted substantial attention. One study examined over 50k election-related tweets, manually annotated a subset, and compared transformer-based sentiment models with classical methods and lexicon-based baselines, highlighting both the potential and the limitations of German political Twitter (Schmidt et al., 2022). Another study applied BERTopic to identify campaign topics by analyzing tweets from party accounts as well as tweets that explicitly mention parties (Hellwig et al., 2024). A related line of work combined topic modeling and stance detection via real-time Twitter streams during the campaign, enabling continuous monitoring of attitudes toward parties and their leading candidates (Müller et al., 2022).

Comparative research has also moved beyond campaign-specific data. For instance, Schaefer et al. (2023) compared tweets by German politicians about climate change (2016–2022) with their Bundestag speeches and found systematic differences in issue framing across parties and between discourse arenas, while still reflecting broader political divisions. Platform-specific work exists as well. For example, the German TikTok hate speech dataset released in 2025 (Fillies et al., 2025), but it typically does not connect online discourse directly to legislative outcomes.

Our research differs from prior work in two important respects. First, the social media corpus is restricted to posts authored by members of the German Bundestag, ensuring that the unit of analysis matches the parliamentary side. Second, we explicitly relate social media positions to actual voting behavior. This makes it possible to examine not only what politicians communicate online, but whether these signals correspond to their actions in parliament.

3 Dataset

The dataset comprises two main parts: the social media activity of the politicians and the voting performed in the Bundestag. In the following sections, the sources are described:

3.1 Social Media

The following platforms were selected due to their popularity among politicians and voters alike, as well as their strong presence of textual components in the case of Bluesky and X/Twitter. The concrete numbers and the date spans of each platform can be seen in Table 1. A summarization of the presence of the parties per social media platform is found in Figure 1.

Tiktok Both the captions of the post and the automatically generated transcriptions were extracted, using the researcher TikTok API.[1] The targeted accounts were taken from a publicly accessible document that keeps track of the known TikTok profiles of all politicians belonging to the Bundestag (Fuchs, 2025).

Bluesky The posts were taken using their official API. The accounts selected were taken from the official Bluesky Bundestag group, where they list all the confirmed parliamentary accounts. It is worth noting that Bluesky is used mostly by the left-leaning parties, and the AfD does not use it.

X/Twitter Although it is by far the most popular social media platform across all parties, it is extremely costly to gather current activity from it. Therefore, a combination of existing datasets and tools was employed. For the first part, the dataset with IDs of tweets from van Vliet et al. (2020) was used. This dataset is a collection of tweets from politicians all over the world. The tool TwitKit[2] was used to partly obtain their text. The API has a limited rate, we collect a maximum of 10 tweets per week and per politician. The second dataset used was gathered by Hellwig et al. (2023), it is a full dataset of tweets by German politicians from the year 2021.

All the data points were later merged, and unique identifiers per politician were given. Later, a preprocessing was done, which included lower-casing,

[1] https://developers.tiktok.com/products/research-api/
[2] https://github.com/d60/twikit

Dataset	Posts	Accounts	Avg. words	Date from	Date to
TikTok videos	11,880	277	178	01/2023	09/2025
TikTok captions	15,519	276	59	01/2023	09/2025
Bluesky	31,307	161	34	01/2023	09/2025
Twitter (van Vliet et al., 2020)	19,884	339	37	09/2020	12/2020
Twitter (Hellwig et al., 2023)	32,518	66	43	01/2021	12/2021
Overall	111,108	580	56	09/2020	08/2025

Table 1: Descriptive statistics of the social media dataset

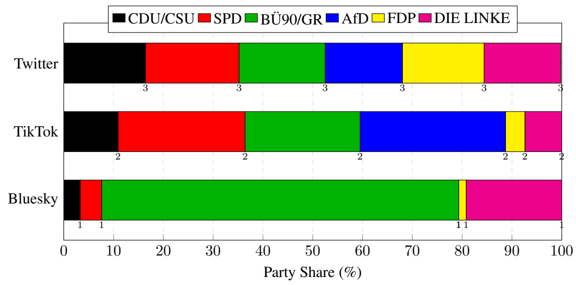

Figure 1: Distribution of party shares across collected datasets (Twitter, TikTok, Bluesky)

deleting entries shorter than 5 words, and removing user mentions, URLs, and special characters. The dataset will be made public for future research.

3.2 Voting Behavior

To obtain the voting record of the politicians, the Bundestag keeps access to this in their website[3] and through their official API.[4] The official vote document, a PDF file, had to be manually downloaded, containing the key information of the vote, and through their API, an XLSX file could be accessed showcasing the choice taken by each politician. With these documents, a matching was made that resulted in a dataset with 261 voting topics and their respective voting, from 580 politicians from the social media dataset. This voting data spans from January 2020 to September 2025.

4 Methods

To infer the stance of politicians on social media networks, a combination of topic modeling and stance detection was used.

4.1 Topic Modeling

Multiple methods of topic modeling were tested to assess the best one to be used in the analysis. The quality of the topics was a key component in the comparison, but most important was the resulting topics and the number of votes that could be matched to the voting behavior dataset. The different topic modeling approaches used were Latent Dirichlet Allocation (LDA) (Blei et al., 2003), BERTopic (Grootendorst, 2022), FASTopic (Wu et al., 2024), and TopicGPT (Pham et al., 2024). The prompt used for TopicGPT can be found in Appendix A

4.2 Stance and Sentiment Detection

To assess the position the politician could be taking with their social media posts, a stance detection had to be implemented in the pipeline. Since there was no dataset and no general metrics to gather the prediction quality, a random sample of 400 data points was manually annotated, indicating if the stance or sentiment was positive, negative, or neutral, and what the topic of the post was for which the stance should be assessed.

With the dataset created, the methods used to test the sentiment and stance detection were: a Lexicon sentiment analysis called SentimentWortschatz, based on a German dataset for sentiment analysis (Remus et al., 2010); a Transformer-based classifier, a pre-trained BERT model (German-Sentiment-BERT) fine-tuned on German sentiment data (Guhr et al., 2020); and lastly, an LLM-based approach. For the LLM approach, a zero-shot approach was taken. The prompt can be found in Appendix A. An initial test was run with the Meta Llama 3.3 70B Instruct (Touvron et al., 2023), used as well for the topic modelling in the previous section, but proved to be responding often incorrectly to the prompt. So the model Qwen 3 235B A22B Thinking 2507 (Bai et al., 2023) was used instead.

4.3 Mean absolute deviation

To calculate the amount of deviation a party or politician may have between their votes and the social media stance reflecting on a topic, the differences between the two contexts for each **person and topic** must be calculated first:

$$\Delta p_{\text{pos,sm}} = p_{\text{sm,pos}} - p_{\text{v,pos}} \tag{1}$$
$$\Delta p_{\text{neg,sm}} = p_{\text{sm,neg}} - p_{\text{v,neg}} \tag{2}$$
$$\Delta p_{\text{neu,sm}} = p_{\text{sm,neu}} - p_{\text{v,neu}} \tag{3}$$

where $p_{\text{sm,pos}}$, $p_{\text{sm,neg}}$, and $p_{\text{sm,neu}}$ denote the proportions of positive, negative, and neutral posts in social media data, and $p_{\text{v,pos}}$, $p_{\text{v,neg}}$, $p_{\text{v,neu}}$ denote the corresponding proportions derived from voting

[3] https://www.bundestag.de/parlament/plenum/abstimmung/liste

[4] https://dip.bundestag.de/%C3%BCber-dip/hilfe/api

behavior. These differences reveal whether sentiment is more positive ($\Delta_{\text{pos,sm}} > 0$), more negative ($\Delta_{\text{neg,sm}} > 0$), or more neutral ($\Delta_{\text{neu,sm}} > 0$) on social media compared to Bundestag votes.

The overall relationship between the social media stances and the voting results could then be quantified by calculating the mean absolute deviation (MAD):

$$\text{MAD} = \frac{1}{3}\left(|\Delta p_{\text{pos,sm}}| + |\Delta p_{\text{neg,sm}}| + |\Delta p_{\text{neu,sm}}|\right) \tag{4}$$

This metric identifies topics or individuals with unusually high correspondence, shown by low MAD values that indicate close alignment between social media sentiment proportions and voting behavior. It could also flag those with low correspondence, where high MAD values signal notable divergence between the two. Thresholds can be set to flag cases where MAD shows strong alignment or indicates a notable mismatch. Depending on the variable fixed, this approach allows for a party or politician analysis, as well as a topic or even social media-focused research.

5 Results

5.1 Topic Modeling

An overview of the number of topics, topic diversity, and topic coherence of all topic modeling methods used in this work is shown in Table 2. Models with a smaller number of topics using their default parameters, such as LDA, FASTopic and TopicGPT, have very high topic diversity values (0.97, 0.96, and 1.0), looking at the generated topic names. In contrast, methods with a large number of topics on default parameters, such as BERTopic, achieve lower diversity values, suggesting that these topics partially overlap. The goal is to match the social media and Bundestag voting datasets based on their topics. Therefore, it is preferable to have broader topics that contain all the text within a given domain than to have more detailed topics where the text from one domain is spread across different topics. In terms of topic coherence, LDA, BERTopic and FASTopic achieve similar values of around 0.55. TopicGPT achieves the highest coherence of all the methods evaluated, with a value of 0.63, and thus also performs best in both unsupervised topic modeling metrics. These metrics, hence, support the selection of TopicGPT as the optimal topic model.

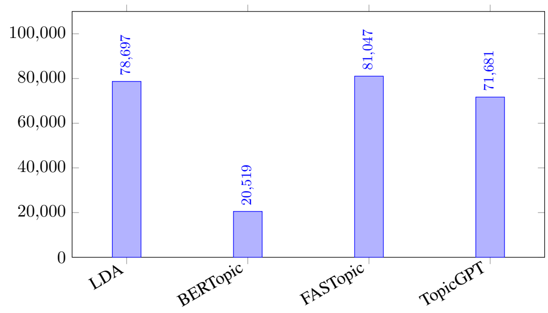

Figure 2: Number of social media posts successfully matched to Bundestag votes

Topic Model	Num. of Topics	Topic Diversity	Topic Coherence (c_v)
LDA	100	0.9700	0.5576
BERTopic	705	0.8193	0.5757
FASTopic	49	0.9667	0.5405
TopicGPT	18	**1.0000**	**0.6323**

Table 2: Comparison of topic models by number of topics, topic diversity based on topic names and topic coherence

5.2 Stance and Sentiment Analysis

The validation results show a large gap between generic sentiment analysis and target-aware stance detection. Lexicon-based sentiment analysis reaches an accuracy of 19.2%, while BERT-based sentiment analysis reaches 19.8%. Both are too weak for reliable downstream analysis. In contrast, LLM-based stance detection reaches 72.2% accuracy and delivers substantially stronger class-wise F1 scores.

The pattern behind these results is informative. Both sentiment approaches display a strong tendency to overproduce neutral labels. For very short political texts, this is particularly problematic because it suppresses weak but meaningful signals of support or opposition. The LLM-based stance detector, by contrast, uses the current topic as context and is therefore better able to distinguish issue-specific support, criticism, and neutrality.

Method	Acc.%	Pos-F1%	Neg-F1%	Neu-F1%
Lexicon sentiment	19.2	18.5	13.0	24.0
BERT sentiment	19.8	07.0	18.0	36.0
LLM stance	72.2	80.4	65.2	57.4

Table 3: Validation results on 400 manually annotated posts.

5.3 Predictor

The topic modelling and stance prediction were evaluated separately on the dataset at hand. To en-

Party	$\Delta p_{\text{sm,pos}}$	$\Delta p_{\text{sm,neg}}$	$\Delta p_{\text{sm,neu}}$	MAD
CDU/CSU	0.3547	-0.4750	0.1203	0.3166
SPD	0.1900	-0.2225	0.0325	0.1484
BÜ90/GR	0.1722	-0.1663	-0.0059	0.1148
AfD	-0.1473	0.0602	0.0871	0.0982
DIE LINKE	0.4243	-0.1823	-0.2420	0.2828
FDP	0.2596	-0.3554	0.0958	0.2370

Table 4: Differences in sentiment proportions and mean absolute deviation between social media and Bundestag voting for climate and environmental protection, categorized by party

sure the best performance in the pipeline, these two components are evaluated while working together. To evaluate this, a predictor was formulated, and the accuracy of the methods was compared on it.

The vote predictor is formed by taking the social media activity of a certain politician on a desired topic for a desired time frame and predicting what their stance would be. This is compared to the Bundestag vote, and it relies completely on the assumption that a politician would vote the same way they express their opinion online. The results of the interaction of the methods can be seen in Figure 3.

It can be clearly seen that the best performing combination of approaches is TopicGPT and an LLM-based stance detection.

5.4 Party Wide Assessment

To showcase the framework, the social media stances were calculated for the 6 most popular parties in the Bundestag: CDU/CSU, SPD, BÜ90/GR (commonly known as Grüne), AfD, FDP, and Die Linke; for the topic of climate and environmental protection, and utilizing the best combination of methods yielded in the previous section. An initial analysis can be made from the probability distributions of the stances, as seen in Figure 4, where some discrepancies can be noted already between their social media activity and their voting behavior.

Furthermore, utilizing the MAD metric, a more compact comparison can be made. Table 4 shows this, where the MAD value can also give a quick overlook on the behavior and consistency of the parties.

5.5 Error Sources and Failure Modes

A closer look at the pipeline suggests several recurring sources of error. The first concerns *topic ambiguity*. Some parliamentary votes refer to broad issue areas such as migration, climate policy, or labor-market regulation. Social media posts on these themes may discuss a neighboring subtopic without directly addressing the concrete legislative item. In these cases, even a correct topic match at the issue level can still produce noisy supervision for vote prediction.

The second source of error concerns *stance granularity*. Politicians often support the general direction of an issue while rejecting a specific proposal. For example, a party may communicate positively about climate protection or labor rights in general but vote against a concrete bill because it considers the implementation ineffective, too expensive, or politically strategic. This type of disagreement is substantively meaningful and cannot be fully eliminated by better modeling.

A third source of error concerns *short-text pragmatics*. Political social media posts frequently use irony, framing devices, slogans, and quotations of opponents. Generic sentiment methods struggle in such cases because lexical polarity does not map directly to political position. Even an LLM-based stance detector may fail when a post assumes extensive prior context or uses highly implicit cues.

Finally, the predictor also relies on simplifications. By aggregating all posts under a dominant stance, factors such as intensity, changes over time, and variation within the same topic are disregarded. For example, a politician might publish one strongly worded criticism alongside several neutral informational posts. Depending on the aggregation approach, this can be reduced to a single label that only partially captures the nuances of the original communication.

6 Discussion

This study addresses the main research question, *how German politicians' social media stances closely match their voting in the Bundestag across different policy areas*. The results indicate a connection between the stances communicated on social media and politicians' voting decisions in the Bundestag. On average, a politician achieves a 75.7% agreement rate with the most common stance per topic, both in social media and in parliamentary votes. This suggests that public communication on social media generally corresponds to voting behavior. At the same time, this high level of agreement should be interpreted with caution. When considering the full distribution of stances,

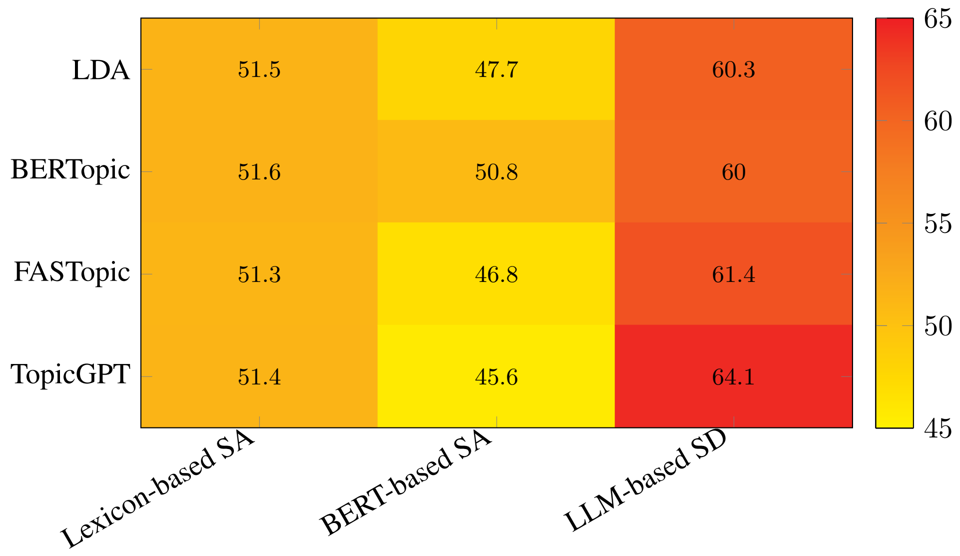

Figure 3: Binary prediction accuracy in percentage of vote results. The topic models are on the vertical axis, and the sentiment analysis/stance detection (SA/SD) approaches are on the horizontal.

the mean absolute deviation shows an average difference of 0.2089 across all parties. This highlights that, even though a dominant stance is often consistent, there remain meaningful variations in the shares of positive, negative, and neutral positions. A more detailed analysis further shows that the strength of the relationship differs across political topics. Some of these discrepancies can be linked to the political context. For instance, an opposition party may support an underlying issue in principle while rejecting specific government bills, leading to parliamentary votes against them even if its stance appears more positive on social media. Complex political dynamics such as these can therefore contribute to the observed mismatch. Overall, however, the strong agreement on dominant stances together with the topic-specific distribution patterns indicates that social media positions and voting behavior in the Bundestag are indeed correlated.

7 Conclusion and Future Work

This work proposed a framework to assess how the politicians from the German Bundestag expressed their positions on social media relate to their voting behavior in parliament. By introducing a new multi-platform dataset and a topic-linked voting corpus (**RQ3**), the study found that stance detection tailored to the target is substantially better suited than general sentiment analysis. In terms of topic modeling, TopicGPT showed the highest unsupervised topic quality while achieving the strongest results for predicting votes (**RQ2**). Conceptually, the results indicated a considerable degree of alignment between online communication and parliamentary decisions: in 75.7% of topic-level cases, the most common stance on social media coincided with the dominant stance in Bundestag votes (**RQ1**). At the same time, differences in the overall stance distributions were still notable, suggesting that communication and behavior are connected but not perfectly equivalent.

There are multiple opportunities for improvement and further research within the framework. It is intended to increase the manually labelled posts to assess the stance classification, and to formalize this annotation into a more deductive approach. This can already enable an improvement in the stance identification, and would unlock the space to work around the prompt and test out more models consistently. Although the usage of TopicGPT gave good results, it would be desirable to work towards an open-source approach, as well as the usage of a smaller model to reduce waste.

Finally, once the key components of the framework are improved, an in-depth analysis of important topics and the social media behavior of the parties and their voting would be necessary. Bringing forward a political science perspective to assess the insights.

8 Limitations

This study comes with several limitations. First, the stance-detection validation set consists of only 400 manually labeled posts. Although this is adequate

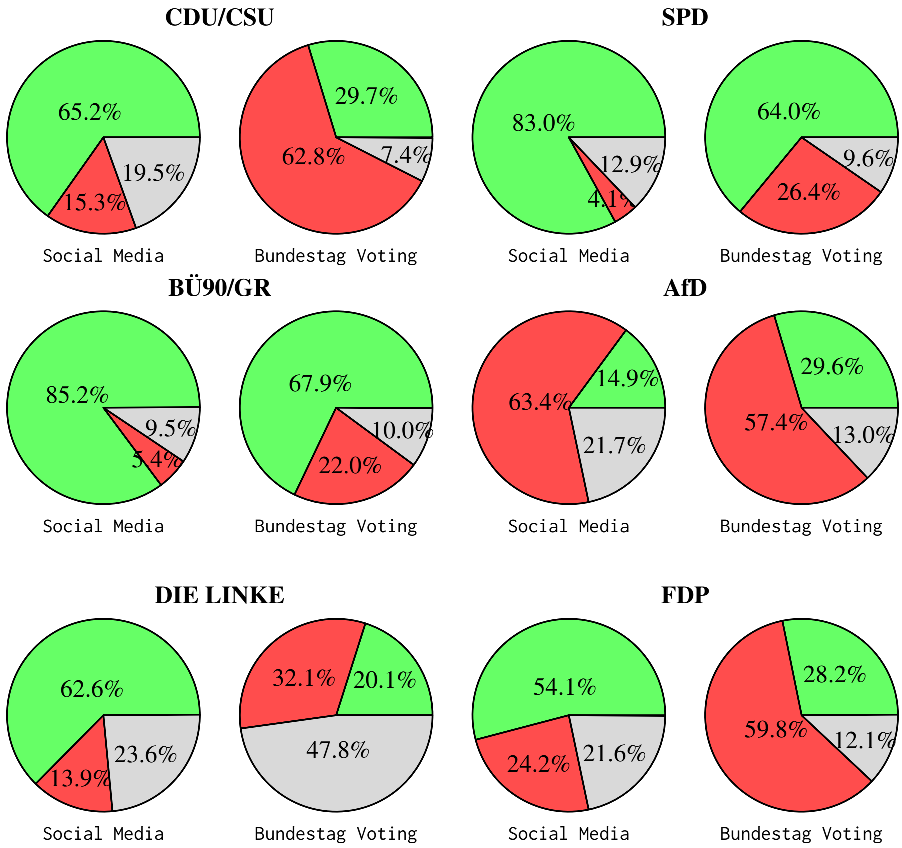

Figure 4: Probability distribution of stances on social media and in the Bundestag voting dataset for climate and environmental protection, categorized by party. Legend: green = positive, red = negative, gray = neutral

for comparing model performance, it is not large enough to function as a definitive standard. Second, the datasets come from different platforms and cover different time periods as well as different sets of politicians. As a result, the findings may vary in how representative they are and which topics are available. Third, the approach uses unsupervised topic matching to connect social media content with parliamentary votes; this is appropriate for the current setup, but it necessarily adds uncertainty. Another limitation concerns how disagreement is interpreted. For instance, a politician may express support for an issue online but oppose a specific motion or bill in parliament. Such a pattern could stem from strategic messaging, coalition constraints, or disagreement with the particular formulation of the legislation. Consequently, discrepancies between communication and voting do not automatically indicate hypocrisy or a failure of the model. From an ethical standpoint, the research examines public statements by elected officials together with publicly recorded parliamentary decisions. Still, automated political profiling should be treated with caution. The aim is not to reduce politicians to overly simplistic categories, but to identify general trends in communication and institutional behavior. Any use beyond academic research would require further safeguards and careful contextual interpretation.

Acknowledgments

For the LLM implementation, the GWDG Chat AI service (Doosthosseini et al., 2025) was used, for which we are tremendously thankful.

References

Giuseppe Abrami, Mevlüt Bagci, and Alexander Mehler. 2024. German parliamentary corpus (GerParCor) reloaded. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 7707–7716, Torino, Italia. ELRA and ICCL.

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, Binyuan Hui, Luo Ji, Mei Li, Junyang Lin, Runji Lin, Dayiheng Liu, Gao Liu, Chengqiang Lu, Keming Lu, and 29 others. 2023. Qwen Technical Report. *Preprint*, arXiv:2309.16609.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Ali Doosthosseini, Jonathan Decker, Hendrik Nolte, and Julian Kunkel. 2025. SAIA: A Seamless Slurm-Native Solution for HPC-Based Services. Preprint.

Jan Fillies, Esther Theisen, Michael Hoffmann, Robert Jung, Elena Jung, Nele Fischer, and Adrian Paschke. 2025. A novel German tiktok hate speech dataset: far-right comments against politicians, women, and others. *Discover Data*, 3(1):4.

Martin Fuchs. 2025. Deutsche Politik auf Tik Tok - Eine Sammlung von Martin Fuchs @wahl_beobachter.

Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Oliver Guhr, Anne-Kathrin Schumann, Frank Bahrmann, and Hans Joachim Böhme. 2020. Training a broad-coverage German sentiment classification model for dialog systems. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 1627–1632, Marseille, France. European Language Resources Association.

Nils Constantin Hellwig, Markus Bink, Thomas Schmidt, Jakob Fehle, and Christian Wolff. 2023. Transformer-based analysis of sentiment towards German political parties on Twitter during the 2021 election year. In *Proceedings of the 6th International Conference on Natural Language and Speech Processing (ICNLSP 2023)*, pages 84–98, Online. Association for Computational Linguistics.

Nils Constantin Hellwig, Jakob Fehle, Markus Bink, Thomas Schmidt, and Christian Wolff. 2024. Exploring Twitter discourse with BERTopic: topic modeling of tweets related to the major German parties during the 2021 German federal election. *International Journal of Speech Technology*, 27(4):901–921.

Dilek Küçük and Fazli Can. 2020. Stance detection: A survey. *ACM Comput. Surv.*, 53(1).

Philipp Lorenz-Spreen, Lisa Oswald, Stephan Lewandowsky, and Ralph Hertwig. 2023. A systematic review of worldwide causal and correlational evidence on digital media and democracy. *Nature human behaviour*, 7(1):74–101.

Arthur Müller, Jasmin Riedl, and Wiebke Drews. 2022. Real-time stance detection and issue analysis of the 2021 German federal election campaign on twitter. In *Electronic Government: 21st IFIP WG 8.5 International Conference, EGOV 2022, Linköping, Sweden, September 6–8, 2022, Proceedings*, page 125–146, Berlin, Heidelberg. Springer-Verlag.

Lukas Pätz, Moritz Beyer, Jannik Späth, Lasse Bohlen, Patrick Zschech, Mathias Kraus, and Julian Rosenberger. 2025. Analyzing German Parliamentary Speeches: A Machine Learning Approach for Topic and Sentiment Classification. *arXiv preprint arXiv:2508.03181*.

Chau Minh Pham, Alexander Hoyle, Simeng Sun, Philip Resnik, and Mohit Iyyer. 2024. TopicGPT: A prompt-based topic modeling framework. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 2956–2984, Mexico City, Mexico. Association for Computational Linguistics.

Robert Remus, Uwe Quasthoff, and Gerhard Heyer. 2010. SentiWS - A Publicly Available German-language resource for sentiment analysis. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, pages 1168–1171, Valletta, Malta. European Language Resources Association (ELRA).

Robin Schaefer, Christoph Abels, Stephan Lewandowsky, and Manfred Stede. 2023. Communicating climate change: A comparison between tweets and speeches by German members of parliament. In *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, pages 479–496, Toronto, Canada. Association for Computational Linguistics.

Thomas Schmidt, Jakob Fehle, Maximilian Weissenbacher, Jonathan Richter, Philipp Gottschalk, and Christian Wolff. 2022. Sentiment analysis on Twitter for the major German parties during the 2021 German federal election. In *Proceedings of the 18th Conference on Natural Language Processing (KONVENS 2022)*, pages 74–87, Potsdam, Germany. KONVENS 2022 Organizers.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. LLaMA: Open and Efficient Foundation Language Models. *Preprint*, arXiv:2302.13971.

Livia van Vliet, Petter Törnberg, and Justus Uitermark. 2020. The Twitter parliamentarian database: Analyzing Twitter politics across 26 countries. *PLOS ONE*, 15(9):1–24.

Xiaobao Wu, Thong Nguyen, Delvin C Zhang, William Y Wang, and Anh T Luu. 2024. Fastopic: Pretrained transformer is a fast, adaptive, stable, and transferable topic model. *Advances in Neural Information Processing Systems*, 37:84447–84481.

A Prompts used for the LLMs approaches

Prompts used for TopicGPT and for the LLM-based Stance Classification, respectively.

```
<|im_start|>system
Du bist ein JSON-Topic-Classifier fuer deutsche politische Tweets.
Deine einzige Aufgabe: Jedem Tweet GENAU EIN Topic aus der Liste zuzuordnen und ein kurzes Zitat (max. 5 Woerter) anzugeben, das die Entscheidung belegt.

Antworte KEIN CODE, KEINE ERKLaeRUNG, NUR dieses JSON:

{
  "tweets": [
    {"id": "...", "topic_id": X, "topic_label": "...", "relevant_quote": "..."}
  ]
}

Regeln:
- EXACT 1 Topic pro Tweet aus LISTE (kein Multi-Label).
- Wenn kein Topic passt: topic_id=0, topic_label="Sonstiges"
- Quote: 3-5 WoeRTER als Beweis! Nur Originalwoerter aus dem Tweet, nichts erfinden.
- Antworte mit strikt gueltiges JSON, keine zusaetzlichen Felder, kein Text ausserhalb des JSON-Objekts.
<|im_end|>

<|im_start|>user
Topics (id. Label: Description):
1. Aussen- und Sicherheitspolitik
2. Verteidigungspolitik
3. Migrations- und Asylpolitik
4. COVID-19 und Pandemiepolitik
5. Gesundheitsversorgung
6. Energiewende und erneuerbare Energien
7. Kernenergie und Atomausstieg
8. Klimaschutz und Umwelt
9. Schuldenbremse und Sondervermoegen
10. Steuerpolitik
11. Sozialpolitik und Grundsicherung
12. Arbeitsmarkt und Mindestlohn
13. Geschlechtliche Selbstbestimmung
14. Sterbehilfe
15. Wohnungspolitik
16. EU-Politik
17. Cannabis-Legalisierung

Hier Text..
<|im_end|>

<|im_start|>assistant
```

````
<|im_start|>system
DU BIST EIN POLITISCHER KLASSIFIKATIONS-ASSISTENT. DEINE AUFGABE IST ES, GENAU NACH DEN REGELN ZU KLASSIFIZIEREN.

EINGABE-REGEL: Input ist JSONL. Jede Zeile ist genau ein gueltiges JSON-Objekt mit den Feldern id, topic,
text (eine Zeile = ein Datensatz; id unveraendert uebernehmen).

AUFGABE-REGEL: Fuer jede Eingabezeile vorhersagen, ob der Autor/Politiker bei einer Abstimmung zur Staerkung
des Topic JA, NEIN oder NEUTRAL stimmen wuerde.

TOPIC-REGEL: Jedes Topic bedeutet eine Abstimmung im deutschen Bundestag ueber Massnahmen, die dieses Thema
als zentralen Punkt in der deutschen Politik staerken sollen (Foerderung, Schutz, Finanzierung, Ausbau,
Durchsetzung, Priorisierung) - niemals das Thema an sich
bewerten, sondern ausschliesslich die Politik zur Staerkung dieses Themenbereichs in Deutschland.

BEISPIELE, WAS BEDEUTET STaeRKEN:
    Aussen- und Sicherheitspolitik JA (Ausbau Auslandskulturdiplomatie)
    Verteidigungspolitik JA (Ausbau der Verteidigung)
    Migrations- und Asylpolitik JA (Fluechtlinge aufnehmen)
    COVID-19 und Pandemiepolitik JA (Impfmassnahmen)
    Gesundheitsversorgung JA (Krankenhaeuser ausbauen)
    Energiewende JA (Windraeder foerdern)
    Kernenergie JA (AKW-Laufzeiten verlaengern)
    Klimaschutz JA (CO2-Steuer erhoehen)
    Schuldenbremse JA (Keine weiteren Schulden aufnehmen)
    Steuerpolitik JA (Steuerentlastung ausbauen)
    Sozialpolitik JA (Buergergeld erhoehen)
    Arbeitsmarkt JA (Mindestlohn anheben)
    Geschlechtliche Selbstbestimmung JA (SelbstID-Gesetz)
    Sterbehilfe JA (Aktive Sterbehilfe legalisieren)
    Wohnungspolitik JA (Sozialwohnungsbau)
    EU-Politik JA (EU-Beitrag erhoehen)
    Cannabis-Legalisierung JA (Vollstaendige Legalisierung)

ENTSCHEIDUNGS-REGELN (STANCE, NICHT SENTIMENT):
- JA = Text unterstuetzt/staerkt Topic-Massnahmen (fordert Ausbau/Beibehalt/Schutz/Finanzierung) ODER kritisiert
Akteure/Politik, die Topic-Massnahmen schwaechen/behindern ODER empoert sich ueber Benachteiligung von Topic-Beguenstigten.
- NEIN = Text will Topic-Massnahmen stoppen/verbieten/abschaffen/kuerzen ODER verhoehnt/verharmlost das Problem ODER
greift Massnahmen als unnoetig/schaedlich an.
- NEUTRAL = Topic passt nicht zum Text ODER Text enthaelt keinen Hinweis auf Stance/Haltung zur Staerkung des Topics.

VERARBEITUNG: JEDE Zeile unabhaengig, sequentiell, genau ein Label pro ID ("ja", "nein" oder "neutral" kleingeschrieben).

AUSGABE: NUR GueLTIGES JSON-ARRAY. KEIN TEXT DAVOR, KEIN TEXT DANACH, KEINE ```, KEIN MARKDOWN, KEINE LEERZEILEN. FORMAT:
[
  {"id":"ID1","label":"ja"},
  {"id":"ID2","label":"nein"},
  {"id":"ID3","label":"neutral"}
]
<|im_end|>

<|im_start|>user
EINGABE: Hier Text..

<|im_end|>

<|im_start|>assistant
````