

Fine-Grained Detection of Solidarity for Women and Migrants in 155 Years of German Parliamentary Debates

Anonymous ACL submission

Abstract

In this study, we evaluate the potential of large language models (LLMs), particularly GPT-4, for detecting fine-grained and detailed notions of solidarity and anti-solidarity towards women and migrants in German parliamentary debates spanning from 1867 to 2022. We evaluate the capabilities of recent LLMs on their ability to detect and categorize nuanced expressions of solidarity and anti-solidarity using a fine-grained social solidarity framework and apply the best-performing models to conduct a longitudinal analysis, aiming to detect and interpret long-term trends in political discourse. Our findings reveal significant shifts in the representation of solidarity and anti-solidarity, corresponding with historical events and changing societal attitudes. However, challenges remain, particularly in the model’s sensitivity to the subtleties of political rhetoric and the limitations posed by partial dataset annotation.

1 Introduction

Large-scale, quantitative analyses of sociological data has the potential to offer deeper insights into social phenomena, such as subtle shifts in perceptions of social groups (Lazer et al., 2009). However, the sheer amount of available data makes human annotation infeasible. For example, just to annotate the training data for this work, 840 person-hours were required for annotation alone. Therefore, computational social science requires tools to automate annotations, using approaches from Natural Language Processing (NLP; Ziems et al., 2024; Wu et al., 2023).

In this work, we evaluate the potential of language model for the detection of fine-grained and detailed notions of solidarity and anti-solidarity towards women and migrants in political speech, namely German parliamentary debates from 1867 to 2022 (Walter et al., 2021). From an NLP view, this task is particularly challenging because (anti-

)solidarity is oftentimes not displayed on the surface level of speech but underlies the overt words (consider the examples in Table 1). Further, over such a long time frame, the topics discussed as well as the language used changes, leading to large variations in the data. From a sociological perspective, the annotation of (anti-)solidarity in German parliamentary debates is crucial because (i) it permits to study the development of (anti-)solidarity over a long period of time in which long-term shifts in perception may become apparent, (ii) it permits to study (anti-)solidarity in finer detail, distinguishing into subtypes of solidarity with different basis (such as shared identity, resource distribution, vulnerability, or diversity; refer to Fig. 1), and (iii) the German context is particularly interesting due to its massive shifts in political context (from empire to republic to dictatorship to partition to unified republic) and coinciding perceptions toward women and migrants (Thränhardt, 1993).

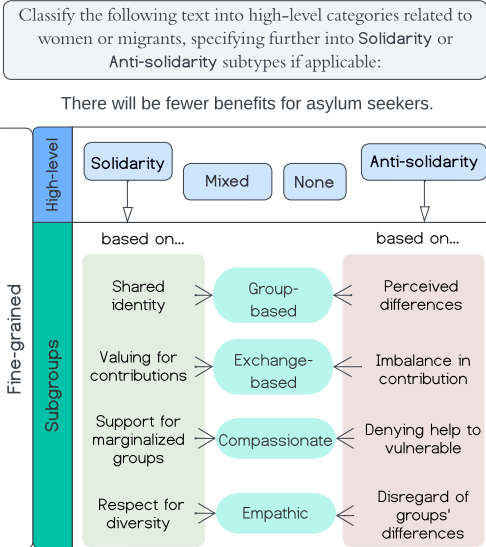


Figure 1: Annotation scheme based on Thijssen (2012).

We (i) manually annotate selected textual material for solidarity and anti-solidarity with our target

Gold Standard	Translation of the Original German Text	Explanation
(1) Compassionate solidarity towards women (June 29, 1961)	“In connection with § 1708 BGB, the Bundestag has set the age of 18 as the limit for the obligation to provide maintenance. In the transitional provisions, this stipulation has been repealed for those who had already reached the age of 16 on January 1, 1962. My faction finds this regulation unfair, as it would exempt significant groups of people from this maintenance obligation. Especially women who have made great efforts to send their children to higher education, for example, would have to bear these costs alone. [...] ”	The speaker is advocating for extended financial support for mothers and is emphasizing the unfairness of removing maintenance obligations.
(2) Exchange-based anti-solidarity towards migrants (Apr. 19, 2018)	“[...] Let me also add: Migration is not necessarily successful – you always act as if that is great – it can fail, and it fails in particular when the immigrants’ qualifications are low. In 2013, before the so-called refugee wave, 40 percent of immigrants from non-EU countries had no qualifications. Since the wave of refugees, stabbings have increased by 20 percent, and we have imported anti-Semitism in the country. Does this make for an outstandingly successful migration?”	The text criticizes migration for its negative economic impacts and the disproportionate burden placed by low-skilled immigrants who take more resources and social stability than they contribute.

Table 1: Example sentences from our dataset showing (anti-)solidarity towards women/migrants. Bold text is the main sentence, the other sentences are for context. Original German texts, as well as examples of mixed stance and none are available in Table 3 in the Appendix.

063 groups and then (ii) evaluate language models on
064 our human annotated data and (iii) we apply the
065 best ones large-scale to detect trends of solidarity
066 and anti-solidarity over time.

067 Our contributions are: (i) we provide a human
068 annotated dataset of 2864 text snippets (focus sen-
069 tence together with surrounding text) of support
070 and anti-support towards two vulnerable groups
071 (women and migrants) in German parliamentary
072 proceedings; (ii) we provide a comparative analysis
073 of LLMs on a complex sociological task using fine-
074 grained annotation framework; (iii) provide fine-
075 grained insights into shifting political discourse
076 with migrants in Germany in the last 155 years.

077 2 Related work

078 Our works connects to (i) computational social
079 science (CSS), (ii) analysis of political data (par-
080 liamentary debates) and (iii) the emergent field of
081 analysis of social solidarity using NLP approaches.

082 **NLP-based CSS.** Recent CSS studies have lever-
083 aged LLMs to tackle a variety of complex tasks.
084 [Ziems et al. \(2024\)](#) conduct a comprehensive eval-
085 uation of LLMs, noting their proficiency in tasks
086 like misinformation, stance, and emotion classifi-
087 cation, while also pointing out weaknesses in tasks
088 which require understanding of subjective expert
089 taxonomies that deviate from the training data of
090 LLMs (such as event argument extraction, implicit
091 hate and empathy classification). [Choi et al. \(2023\)](#)
092 explore capabilities of LLMs in understanding so-

cial knowledge through zero-shot evaluations, re- 093
vealing that pre-trained models have inherent 094
limitations in social language understanding and are 095
frequently outperformed by smaller, fine-tuned su- 096
pervised models. [Wu et al. \(2023\)](#) evaluate the 097
potential use of LLMs in addressing scaling and 098
measurement issues in social science and find that 099
LLMs can generate meaningful responses regard- 100
ing political ideology and significantly improve 101
text-as-data methods in social science. Investigat- 102
ing particularly German debates around the topic of 103
migration, [Blokker et al. \(2020\)](#) and [Zaberer et al. 104](#)
(2023) utilize fine-tuning of transformer-based lan- 105
guage models to classify claims in German news- 106
papers. [Chen et al. \(2022\)](#) apply LLM-based clas- 107
sification on German social media posts to study 108
public controversies over the course of one decade. 109
In contrast to these approaches, we apply LLMs to 110
longitudinal historical data and explore it for a new 111
challenging task, fine-grained detection of social 112
solidarity. 113

Analysis of parliamentary debates using NLP 114
tools. [Abercrombie and Batista-Navarro \(2020\)](#) 115
review 61 studies highlighting the use of dictionary- 116
based sentiment scoring, statistical machine learn- 117
ing, and other conventional NLP methods to ana- 118
lyze sentiment and position-taking within parlia- 119
mentary and legislative contexts. [Blätte and Wüst 120](#)
(2017), for instance, investigate differences in lan- 121
guage use in German parliamentary protocols be- 122
tween 1996 and 2013 comparing politicians with 123

and without a migrant background when debating migration issues. Müller-Hansen et al. (2021) explore dynamic topic modeling to investigate how discussions on coal have shifted in the German parliament in response to changes in energy and climate policy. Additionally, Walter et al. (2021) employ diachronic word embeddings to track anti-semitic and anti-communist biases in German parliamentary debates. More recently, Bornheim et al. (2023) apply Llama 2 to automate speaker attribution in German parliamentary debates from 2017-2021. Our research goes beyond this by adopting recent powerful LLMs to track changes of a specific social concept, solidarity, in plenary debates from three centuries.

Social solidarity in NLP. Previous studies of social solidarity in NLP have largely focused on social media platforms. For example, Santhanam et al. (2019) study how emojis are used to express solidarity in social media during Hurricane Irma in 2017 and Paris terrorist attacks from November 2015. Ils et al. (2021) also consider social media, and in particular analyze European solidarity in the context of COVID-19. Eger et al. (2022) extend this work, focusing on the influence of design choices (how to sample tweets, e.g., using which keywords and in which languages) on the assessment of how solidarity changes over time.

Compared to these works, we use a similar methodological setup (annotate data and infer trends), but differ along various content dimensions: we do not focus on social media but parliamentary debates; we analyze solidarity across much longer time periods (>150 years vs. few years or months) and with a particular focus on Germany. Most importantly, we apply a much more fine-grained solidarity frames following work from political science (Thijssen, 2012) using LLMs for systematic categorization and examination of solidarity types over time.

3 Data

We obtain data from two sources: (i) *Open Data* where the German parliament publishes all plenary protocols from the *Bundestag* (en.: federal diet); and (ii) *Reichstagsprotokolle* that contains all *Reichstag* (en.: imperial diet) protocols, distributed by the *Bayerische Staatsbibliothek*; we use the OCR-scanned version from Walter et al. (2021). Links to data, models, etc. used are in the [Appendix](#).

3.1 Parliament data

For the *Reichstag* data, we apply preprocessing steps similar to Walter et al. (2021) (e.g., removal of OCR artifacts), but keep German umlauts, capitalization, and punctuation. We then automatically split the data into individual sittings and collect metadata like the date, period and session number of each sitting, which we manually checked and corrected. We further remove interjections and split the data into individual sentences using NLTK (Bird et al., 2009). The resulting dataset consists of a total of 19.1M sentences. We release this dataset of plenary protocols from German political debates (DeuParl) consisting of 9,923 sittings from 1867 to 2022 on GitHub.¹

3.2 Data preprocessing

To select keywords, we train a word2vec model (Mikolov et al., 2013) on the dataset and manually select words with vector representations similar to *Migrant* (en.: migrant) and *Frau* (en.: woman) that are frequent enough; this resulted in 32 keywords for *Migrant* and 18 keywords for *Frau*. These include general terms like *Migrant*, *Emigrant* (en.: emigrant), *Immigrant* (en.: immigrant) and *Frau* to categories specific to historical periods and various social roles (see the full list of keywords, and further preprocessing in [Appendix](#)). For a detailed exploration of how these keywords distribute across the dataset, see [Figure 9](#) and [Figure 10](#) in [Appendix](#).

Using these keywords, we extract 58k main sentences (*instances*) from DeuParl for migrant and 131k instances for women, which we expanded by adding three sentences before and three after each instance for context, resulting in a total of about (i) 463k sentences (9.79M tokens) for migrants and (ii) 1.58M sentences (32.82M tokens) for women. The number of instances in our dataset (i.e., that contain one of our keywords) over the years is visualized in [Figure 2](#). It should be noted that the dataset is sparse in the period from 1933 to 1949, i.e. during the NS dictatorship and the immediate after-war period until the first parliament after the war was elected in 1949.

4 Data annotation

To obtain ground truth data for model training and evaluation, we annotated 2864 instances with five annotators (all student assistants, with specializations in social science or computer science). The

¹removed for anonymity reasons

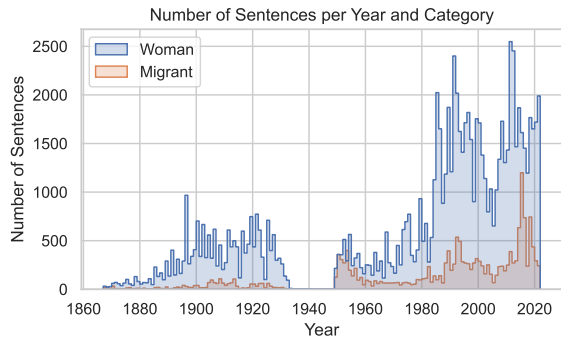


Figure 2: Number of instances in the Woman and Migrant dataset in each year.

annotation was performed over a duration of nine months. In the first three months, we iteratively refined the annotation guidelines and monitored the inter-rater agreement (measured by Cohen’s Kappa) on every 100 instances among three initial annotators. Once the guidelines were finalized and two new annotators had joined and completed the training, the team – now consisting of four members due to one of the annotators leaving the project – started to independently annotate new instances once improvements in inter-rater agreement stagnated (see Section 4.2 for exact scores).

4.1 Annotation task design

For the manual annotation, we take the target sentence and three preceding and following sentences for context into account. We first select a high-level category (*solidarity*, *anti-solidarity*, *mixed*, *none*). *Solidarity* or *anti-solidarity* cases are then further specified into frames as defined by Thijssen and Verheyen (2022): *group-based*, *compassionate*, *exchange-based*, and *empathic*². We describe each of the included variables below.

High-level categories Drawing from the definitions of solidarity by Lahusen and Grasso (2018) and Ils et al. (2021) we consider a sentence to show *solidarity* if the speaker expresses (i) willingness to share resources, directly or indirectly; (ii) implicit or explicit motivations for supporting our target groups. A sentence shows *anti-solidarity* if the speaker expresses (i) that resources should be limited for a particular target group, (ii) unwillingness to support the group or (iii) exclusionary implications towards the group. The label *mixed* is used for texts with both supporting and opposing expressions. Texts expressing neutral stance are labeled

²If no subtype for solidarity or anti-solidarity can be inferred, we label a case as (*anti*-)solidarity without a subtype

as *none*.

Group-based (anti-)solidarity *Group-based solidarity* is coded when a text emphasizes shared identity and common goals, values and rights among group members. In contrast, *group-based anti-solidarity* emphasizes out-group exclusion and difference, advocating for exclusion based on these perceived differences.

Compassionate (anti-)solidarity *Compassionate solidarity* is coded when a text expresses support for marginalized or vulnerable groups, highlighting their need for protection. Conversely, *compassionate anti-solidarity* dismisses these groups by considering them already in a good position, thereby minimizing their need for support or protection.

Exchange-based (anti-)solidarity *Exchange-based solidarity* is coded when speakers highlight the economic contributions of “exchange partners”, who may be rewarded or asked to contribute more for support. In contrast, *exchange-based anti-solidarity* advocates for punishment of groups perceived to receive more than they contribute or to cause problems.

Empathic (anti-)solidarity *Empathic solidarity* is coded when a speaker expresses support respect for individual differences, seeing social diversity as beneficial. Conversely, *empathic anti-solidarity* arises when differences are used as grounds for exclusion or neglect.

A detailed example for the annotation process is available in 13, which illustrates the full annotation pipeline with providing explanations for chosen labels.

4.2 Annotation results

While initial agreement levels were low, by the time annotators began working independently, they achieved a pairwise agreement with a Cohen’s Kappa of 0.42 on a fine-grained level and 0.62 on a high-level.

We observe three main disagreement issues in annotation. Firstly, *none* cases were frequently misclassified as expressing solidarity or anti-solidarity, and vice versa (see Figure 3); secondly, *mixed stance* was especially confused with *anti-solidarity*. Finally, there was considerable confusion within the subtypes of solidarity and anti-solidarity (see Figure 7 in the Appendix). This confusion is often due to overlapping characteristics or the presence

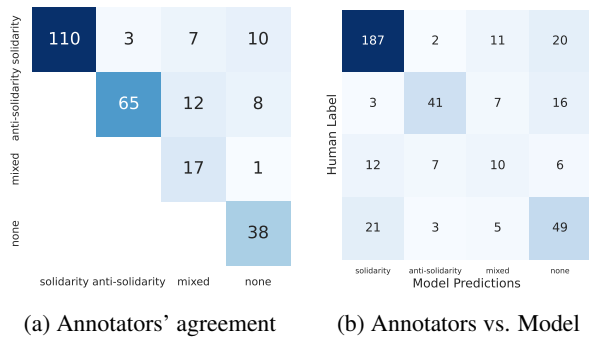


Figure 3: 3a shows the comparison of annotations between our annotators; 3b between the human annotations and our best model’s prediction (cf. Section 6) on a test set. The former is aggregated over all pairwise comparisons of annotators, thus the matrix is symmetric.

of multiple subtypes within the text. Table 5 in the Appendix lists some examples of divergence between our annotators, where we explain why several labels might be correct, which gives insight into more difficult instances. However, there was almost no confusion between solidarity and anti-solidarity.

Our dataset comprises 2864 annotated instances, 1437 for migrants and 1427 for women. 368 instances in our dataset (referred to as *curated*) were reviewed by a social science expert to provide a reliable comparison benchmark for evaluation of our models. Other consensus mechanisms to determine the final labels for the human-annotated dataset, and their distribution are shown in Table 4b.

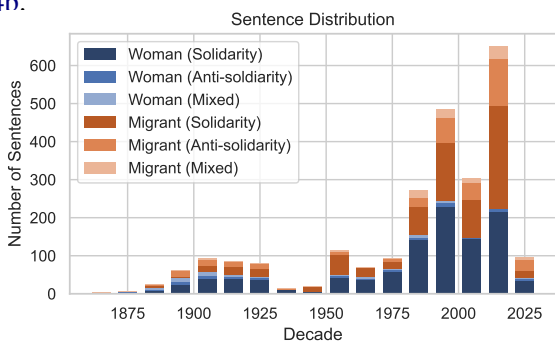


Figure 4: Distribution of instances in the human annotated dataset across time and target groups.

5 Models and experiments

With the goal of identifying the most effective model (both in terms of performance and costs) for our large-scale sociological analysis, we design a series of experiments to evaluate models’ performance at achieving human-level competency

across both high-level categories like notions of solidarity and anti-solidarity, and fine-grained categories specific to sociological theoretical frameworks. We first assess the performance of different models (gpt-4-1106-preview, base and instruction-finetuned gpt-3.5-turbo-0125³) by comparing them to several baselines and report results for both target groups. Once the quality of the models is assured, we apply the best performing model — GPT-4 — large-scale to determine trends in Section 7 (limiting it to the migrant dataset due to the costs associated with employing the GPT-4 model and the high degree of solidarity for women overall).

Data We use a 70/15/15 train/dev/test split for all Migrant and Woman annotated data, allocating 78% of *single* labels to training, along with 58% of *majority*, 43% of *curated*, and 31% of *most* labels. The test set includes 4% of *single*, 49% of *curated*, 31% of *majority* and 65% of *most* labels. This gives us 1539 train, 329 dev and 429 test instances. We create 3 random data splits, and calculate performance metrics as the average score of the 3 runs on the test sets. These sets are fully used for training and evaluating baseline models; for inference-based experiments with GPT-3.5 and GPT-4, only test sets are used (also averaged on 3 runs).

Metrics To evaluate our models, we report the **Macro F1 Score (Macro F1)** to account for class imbalance. We calculate these metrics for both high-level and fine-grained tasks. We also calculate the **F1 Score** for the classes individually.

5.1 Models

Baseline For the baseline, we use a BERT-based model (Devlin et al., 2019) with 110M parameters. We feed a target (i.e., *Frau* or *Migrant*) and a full text to the model (main sentence along with previous and next three sentences for the context), setting the maximum token length to 512. We add a fully connected layer atop the pooled output of the BERT-based model, with 4 output neurons for the high-level and 10 — for the fine-grained task. This layer utilizes softmax activation to determine the most probable class for a given input sentence. To address class imbalance, we oversample the

³We additionally evaluated the Llama-2 with 7B and 13B parameters, along with the LAION LeoLM models based on Llama-2 (<https://laion.ai/blog-de/leo-lm/>), but these experiments were discontinued due to unsatisfactory results.

minority classes to match the majority class by randomly duplicating entries until all classes have equal representation. The model is finetuned with the following hyperparameters: a batch size of 16, a learning rate of $4e-4$, a warmup ratio of 0.05, and linear decay. We train for 10 epochs and optimize using AdamW (Loshchilov and Hutter, 2017), and the categorical crossentropy as the loss function.

GPT-4 We design two prompts (one for each target group) that include several elements: (i) incorporating chain-of-thought reasoning (Wei et al., 2022); (ii) providing precise definitions and insights derived from annotation discussions (iii) introducing potentially problematic labels (such as *empathic solidarity* and *empathic anti-solidarity*) earlier in the prompt and (iv) implementing a two-step prompting strategy that initially categorizes texts at a high-level followed by detailed subtype classification (full prompts are provided in Figure 14 and Figure 15).⁴

Prompt-based fine-tuned GPT-3.5 In the following, using the prompt identified for GPT-4’s fine-grained classification, we proceeded to fine-tune GPT-3.5 on instances sampled from our initial train set (114 for migrants; 109 for women⁵), ensuring a balanced distribution across labels. The fine-tuning dataset was structured with the **system** initiating sequences by providing detailed instructions for classifying texts into high-level categories and requesting further sub-categorization into solidarity or anti-solidarity; **user** role presenting texts; and **assistant** providing classifications structured as per our two-step reasoning approach, along with explanations generated using GPT-4⁶.

For all GPT-experiments, we test the prompt under three conditions: zero-shot prompting, few-shot prompting, as well as removing or providing the context (three preceding and three succeeding sentences). We test the latter condition for BERT-based baseline as well.

⁴Two-step prompting was initially tested using two separate model calls but was later refined to a single, streamlined prompt to reduce inference costs.

⁵The fine-tuning guide by OpenAI recommends using 50 to 100 examples for training: <https://platform.openai.com/docs/guides/fine-tuning/preparing-your-dataset>

⁶We initially trained GPT-3.5 with the annotators’ brief explanations, but this led the model to replicate their concise style too closely during inference, resulting in inadequate reasoning and poor outcomes which we do not report here.

6 Results

Results on the test sets are shown in Table 2. GPT-4 consistently outperforms other models across both high-level and fine-grained tasks, both for women and migrants. Interestingly, it achieves similar performance in 0-shot and few-shot settings — 0.37 (0.60) and 0.42 (0.73) for women, and 0.37 (0.54) and 0.43 (0.63) for migrants, respectively. This can be attributed to the use of carefully crafted definitions, which eliminates the need for additional examples. Fine-tuned version of GPT-3.5 demonstrates only marginal improvement over the base model, generally falling short of GPT-4’s performance across most metrics.

Most importantly, GPT-4 leads in terms of F1 scores for specific categories. While all models effectively identify solidarity (achieving at least 0.7), GPT-4 also effectively identifies anti-solidarity for both target groups (0.65 for women and 0.87 for migrants in 0-shot setting), where other models often fail. However, although GPT-4 also leads in the mixed stance, it achieves lowest F1 scores for this category. These scores are particularly low compared to the human baseline, indicating that the model struggles with more complex and ambiguous categories.

Overall, the human baseline outperforms all models in high-level and fine-grained tasks, suggesting that while the models have reached a degree of proficiency, there is a gap compared to human understanding, especially in more complex categorizations. Given that GPT-4’s was the closest to human baseline and its consistent performance in both 0-shot and few-shot scenarios, we have opted to use the 0-shot approach for the Section 7 section.

6.1 Error analysis

For the error analysis, we compare the human annotations and 0-shot predictions of GPT-4 for both target groups on the test set using the confusion matrices for high-level labels shown in Figure 3, as well as for fine-grained level labels in Figure 7 provided in Appendix. We also consider explanations provided by GPT-4.

Overall, the errors we observe are consistent with confusion errors which we observed during the human annotation process. *Solidarity* and *anti-solidarity* are rarely confused (1% of cases). Most confusion is between (*anti*-)*solidarity* subtypes and *none*, as well as mixed stance and (*anti*-)*solidarity* subcategories, where the model seems to look for

Model	Method	Fine-grained (high-level)		Solidarity		Anti-solidarity		Mixed	
		W	M	W	M	W	M	W	M
GPT-4	0-shot	0.37 (0.60)	0.42 (0.73)	0.85	0.86	0.65	0.87	0.30	0.58
	Few-shot	0.37 (0.54)	0.43 (0.63)	0.85	0.83	0.50	0.75	0.18	0.40
GPT-3.5 fine-tuned	0-shot	0.18 (0.45)	0.27 (0.53)	0.80	0.74	0.12	0.61	0.28	0.27
	Few-shot	0.22 (0.47)	0.28 (0.48)	0.78	0.70	0.18	0.65	0.35	0.07
GPT-3.5 base	0-shot	0.15 (0.46)	0.19 (0.48)	0.75	0.66	0.38	0.68	0.17	0.11
	Few-shot	0.12 (0.41)	0.27 (0.50)	0.70	0.61	0.33	0.64	0.25	0.23
	BERT	0.02 (0.23)	0.19 (0.33)	0.87	0.78	0.00	0.38	0.00	0.00
	Human baseline	0.48 (0.72)	0.56 (0.78)	0.87	0.88	0.68	0.86	0.57	0.64

Table 2: Comparative performance (macro F1) of models vs. human baseline (calculated as an average macro F1 between annotators’ labels and the final label) on combined high-level (in parentheses) and fine-grained tasks for both women (W) and migrants (M), with further F-1 scores for the categories of solidarity, anti-solidarity, and mixed stance. Detailed scores for each category are provided, with the full breakdown available in the appendix.

more context and stronger indications of solidarity, despite being asked to consider even slight expressions of it (see examples 1 and 2 in Table 6).

There is also notable confusion between the solidarity subtypes, with the most frequent confusion occurring between *group-based* and *compassionate solidarity*, likely because of overlapping category elements within the texts (see example 3). Furthermore, there was confusion between *compassionate* and *empathic solidarity*, where the model incorrectly interprets the definition of the latter, as in example (4).

7 Analysis

In this section, we analyze the development of (anti-)solidarity over time. As basis for the analysis, we use annotations of the best-performing model (GPT-4). However, due to cost constraints, the annotations were limited to a) data concerning migrants, and b) a sample of 18,300 records from the overall 58k records concerning migrants. The sample was drawn proportionally for the time spans in the original data. This selection includes every record that has information about political parties (see Appendix for details on political parties data extraction and list of parties included in the analysis).

Figure 6 shows the percentage of instances classified by GPT-4 as solidarity, anti-solidarity, and mixed-stance towards migrants in 18,300 samples. Other records were not classified within these categories. Throughout the periods analyzed, solidarity consistently surpassed anti-solidarity. Soli-

arity increased from under 20% to 30% between 1880 and 1910, and surged to over 50% post-NS regime, stabilizing above 40% thereafter. Anti-solidarity initially rose from about 5% to over 15% between 1870 and 1890, maintained until 1920, then dropped below 5% in 1950, but began a steady climb back above 20% by 2020. Historically, the peak in solidarity in the 1950s coincides with an influx of people of German decent from Eastern Europe, which were generally viewed in a positive light in parliamentary debates (so-called “Vertriebene” or “expellees”; Fröhlich, 2023). The rise of anti-solidarity since 1960 coincides with rising anti-migrant sentiments in response to work migrants in the 1960s and 1970s (Faist, 1994), right-wing demonstrations against liberal asylum laws in the 1990s (Faist, 1994) and the immigration of refugees due to the Syrian war around 2015 with the subsequent rise of the extreme right-wing AfD party (Hertner, 2022).

In Figure 5, for solidarity (left), there is a noticeable peak in group-based solidarity (i.e., emphasis on shared national identity) of over 50% in 1870 with the founding of the German empire. It drops to below 20% by 1880, surges back above 50% in 1970, and then declines to below 30% by 2020. These trends align with periods of strong German nationalism (around 1870 and pre-World War II) and the influx of expellees in the 1950s and 1960s. For anti-solidarity (right), we observe particularly clear trends. Group-based anti-solidarity was predominant before World War II, increasing from about 70% to over 90%, and then sharply decreased to below 60% afterward, reflecting a decline in

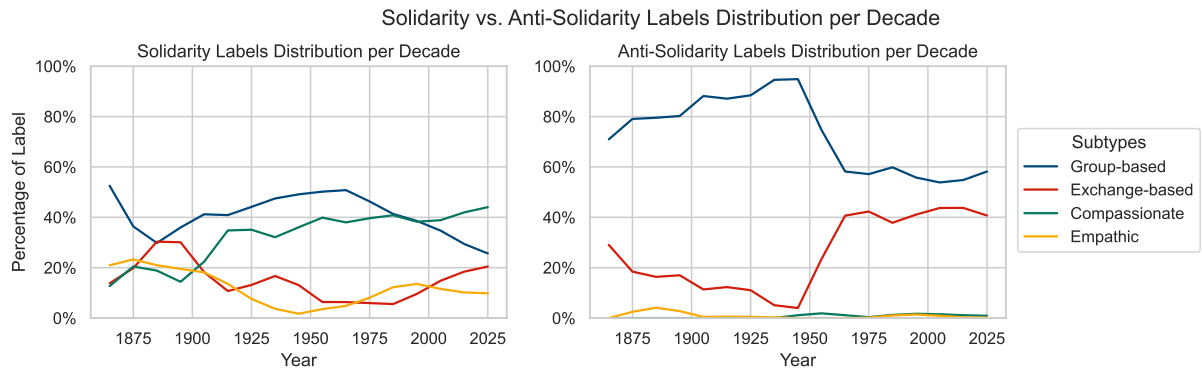


Figure 5: Fraction of solidarity (left) and anti-solidarity (right) subtypes according to GPT-4 in each decade.

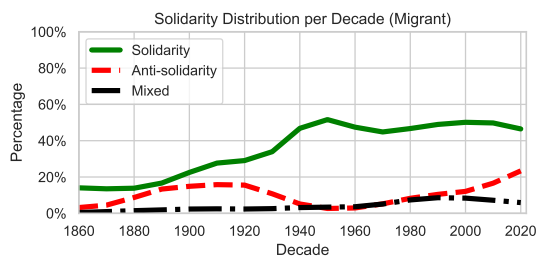


Figure 6: Fraction of solidarity, anti-solidarity, and mixed stance towards migrants according to GPT-4 in each decade.

530 opposition to migration based on national identity
 531 in parliamentary debates post-NS era. Instead,
 532 anti-solidarity arguments shifted to exchange-based
 533 anti-solidarity, which declines from about 30% to
 534 below 10% before World War II and then sharply
 535 rises to above 40% after the war, with arguments
 536 that stress that migrants do not contribute enough
 537 economically. Neither compassionate nor empathic
 538 anti-solidarity are frequent at any time, according
 539 to GPT-4.

540 Data subdivision by the speaker’s party shows
 541 that nearly all parties, except the extreme-right
 542 AfD, predominantly express solidarity over anti-
 543 solidarity (Figure 8 in Appendix, left). Compassion-
 544 ate and group-based solidarity are most common
 545 across parties. Left-leaning parties (Linke, Grüne)
 546 display higher levels of compassionate and empathic
 547 solidarity than centrist parties (SPD, FDP, CDU/CSU),
 548 which focus more on exchange-based solidarity, empha-
 549 sizing migrant contributions. Conversely, right-wing
 550 parties (CDU/CSU, AfD) predominantly engage in anti-
 551 solidarity speech, with CDU/CSU focusing on group-
 552 based and AfD on exchange-based anti-solidarity, sug-
 553 gesting migrants contribute less (Figure 8 in Appendix,
 554 right).

555 This distribution matches findings on partisan ideol-
 556 ogy in Flanders, Belgium (Thijssen and Verheyen,
 557 2022).

8 Concluding remarks 558

559 We find that large language models like GPT-4 are
 560 effective in identifying expressions of solidarity
 561 and anti-solidarity in parliamentary debates, which
 562 can be achieved with carefully crafted prompts in-
 563 formed by human expertise. However, they still fall
 564 short in handling the subtleties and complexities
 565 of human language, particularly when compared to
 566 human annotators. Secondly, our historical analysis
 567 of these debates offers insights into long-term shifts
 568 in political rhetoric and societal values, particu-
 569 larly concerning migrants. We observe that periods
 570 of increased solidarity often coincide with events
 571 that evoke national sympathy or shared hardships,
 572 while spikes in anti-solidarity align with times of
 573 economic pressures or political unrest.

Limitations 574

575 Our study faces several limitations that should be
 576 considered when interpreting the results. Firstly,
 577 parliamentary debates, while rich sources of po-
 578 litical rhetoric, are not necessarily representative
 579 of wider public opinion or broader societal atti-
 580 tudes. Secondly, the task of annotating political
 581 speech, particularly concepts such as solidarity and
 582 anti-solidarity, poses significant challenges. These
 583 concepts are inherently complex and laden with
 584 subtleties that are difficult to capture, both for hu-
 585 man annotators and automated models. Addition-
 586 ally, due to resource constraints, GPT-4 was only
 587 applied to annotate part of the dataset, which may
 588 limit the generalizability of our findings.

589
590
591
592
593

594
595
596
597

598
599
600
601
602
603
604

605
606
607
608
609
610

611
612
613
614
615

616
617
618
619

620
621
622
623
624

625
626
627
628
629
630
631
632
633

634
635
636

637
638
639
640

641
642
643
644

References

Gavin Abercrombie and Riza Batista-Navarro. 2020. Sentiment and position-taking analysis of parliamentary debates: a systematic literature review. *Journal of Computational Social Science*, 3(1):245–270.

Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc."

Nico Blokker, Erenay Dayanik, Gabriella Lapesa, and Sebastian Padó. 2020. [Swimming with the tide? positional claim detection across political text types](#). In *Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science*, pages 24–34, Online. Association for Computational Linguistics.

Andreas Blätte and Andreas M. Wüst. 2017. [Der migrationspezifische einfluss auf parlamentarisches handeln: Ein hypothesentest auf der grundlage von redebeiträgen der abgeordneten des deutschen bundestags 1996–2013](#). *Politische Vierteljahresschrift*, 58(2):205–233.

Tobias Bornheim, Niklas Grieger, Patrick Gustav Blaneck, and Stephan Bialonski. 2023. Speaker attribution in german parliamentary debates with qlora-adapted large language models. *arXiv preprint arXiv:2309.09902*.

Yiyi Chen, Harald Sack, and Mehwish Alam. 2022. Analyzing social media for measuring public attitudes toward controversies and their driving factors: a case study of migration. *Soc. Netw. Anal. Min.*, 12(1):135.

Minje Choi, Jiaxin Pei, Sagar Kumar, Chang Shu, and David Jurgens. 2023. Do llms understand social knowledge? evaluating the sociability of large language models with socket benchmark. *arXiv preprint arXiv:2305.14938*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Steffen Eger, Dan Liu, and Daniela Grunow. 2022. Measuring social solidarity during crisis: The role of design choices. *Journal of Social Computing*.

Thomas Faist. 1994. [How to define a foreigner? the symbolic politics of immigration in german partisan discourse, 1978–1992](#). *West European Politics*, 17(2):50–71.

Christiane Fröhlich. 2023. [Migration as crisis? german migration discourse at critical points of nation-building](#). *American Behavioral Scientist*, early access.

Isabelle Hertner. 2022. [Germany as ‘a country of integration’? the cdu/csu’s policies and discourses on immigration during angela merkel’s chancellorship](#). *Journal of Ethnic and Migration Studies*, 48(2):461–481.

Alexandra Ils, Dan Liu, Daniela Grunow, and Steffen Eger. 2021. [Changes in European solidarity before and during COVID-19: Evidence from a large crowd- and expert-annotated Twitter dataset](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1623–1637, Online. Association for Computational Linguistics.

Christian Lahusen and Maria T. Grasso, editors. 2018. [Solidarity in Europe: Citizens’ Responses in Times of Crisis](#). Palgrave Studies in European Political Sociology. Springer International Publishing.

David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. 2009. [Computational social science](#). *Science*, 323(5915):721–723.

Ilya Loshchilov and Frank Hutter. 2017. [Decoupled weight decay regularization](#). *arXiv preprint arXiv:1711.05101*.

Tomas Mikolov, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. In *ICLR*.

Finn Müller-Hansen, Max W Callaghan, Yuan Ting Lee, Anna Leipprand, Christian Flachsland, and Jan C Minx. 2021. [Who cares about coal? analyzing 70 years of german parliamentary debates on coal with dynamic topic modeling](#). *Energy Research & Social Science*, 72:101869.

Sashank Santhanam, Vidhushini Srinivasan, Shaina Glass, and Samira Shaikh. 2019. [I stand with you: Using emojis to study solidarity in crisis events](#). *ArXiv*, abs/1907.08326.

Peter Thijssen. 2012. [From mechanical to organic solidarity, and back: With honneth beyond durkheim](#). *European Journal of Social Theory*, 15(4):454–470.

Peter Thijssen and Pieter Verheyen. 2022. [It’s all about solidarity stupid! how solidarity frames structure the party political sphere](#). *British Journal of Political Science*, 52(1):128–145.

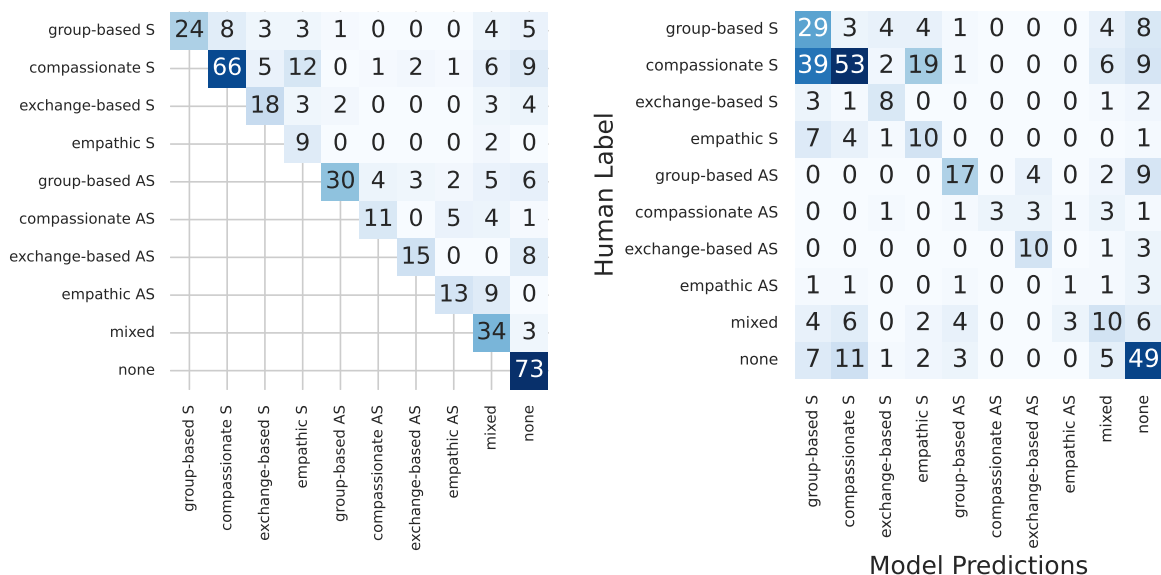
Dietrich Thränhardt. 1993. [Die ursprünge von rassistismus und fremdenfeindlichkeit in der konkurrenzdemokratie: Ein vergleich der entwicklungen in england, frankreich und deutschland](#). *Leviathan*, 21(3):336–357.

Tobias Walter, Celina Kirschner, Steffen Eger, Goran Glavavs, Anne Lauscher, and Simone Paolo Ponzetto. 2021. [Diachronic analysis of german parliamentary](#)

700	proceedings: Ideological shifts through the lens of	752
701	political biases. <i>2021 ACM/IEEE Joint Conference</i>	753
702	<i>on Digital Libraries (JCDL)</i> , pages 51–60.	754
703	Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten	755
704	Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou,	756
705	et al. 2022. Chain-of-thought prompting elicits rea-	757
706	soning in large language models. <i>Advances in neural</i>	758
707	<i>information processing systems</i> , 35:24824–24837.	759
708	Patrick Y Wu, Joshua A Tucker, Jonathan Nagler, and	760
709	Solomon Messing. 2023. Large language models	761
710	can be used to estimate the ideologies of politi-	762
711	cians in a zero-shot learning setting. <i>arXiv preprint</i>	763
712	<i>arXiv:2303.12057</i> .	764
713	Urs Zaberer, Sebastian Pado, and Gabriella Lapesa.	765
714	2023. Political claim identification and categoriza-	766
715	tion in a multilingual setting: First experiments. In	767
716	<i>Proceedings of the 19th Conference on Natural Lan-</i>	768
717	<i>guage Processing (KONVENS 2023)</i> , pages 219–228,	769
718	Ingolstadt, Germany. Association for Computational	770
719	Linguistics.	771
720	Caleb Ziems, William Held, Omar Shaikh, Jiaao Chen,	772
721	Zhehao Zhang, and Diyi Yang. 2024. Can large lan-	773
722	guage models transform computational social sci-	774
723	ence? <i>Computational Linguistics</i> , pages 1–55.	775
724	Appendix	776
725	List of keywords	777
726	Keywords for <i>Migrant</i> (en.: migrant) we use	778
727	Flüchtlinge, Ausländer, Flüchtlingen, Zuwan-	779
728	derung, Vertriebenen, Ausländern, Asylbewerber,	780
729	Migranten, Migration, Heimatvertriebenen,	781
730	Aussiedler, Einwanderung, Ansiedler, Ver-	782
731	triebene, Zuwanderer, Asylbewerber, Flüchtling,	783
732	Heimatvertriebene, Sowjetzonenflüchtlinge,	
733	Aussiedlern, Einwanderer, Asylsuchenden, Asyl-	
734	suchende, Bürgerkriegsflüchtlinge, Zuwanderern,	
735	Ansiedlern, Migrantinnen, Vertriebener, Emi-	
736	granten, Kriegsflüchtlinge, Ausländerinnen, and	
737	Immigranten.	
738	When doing stability tests over the chosen key-	
739	words, we make sure to choose sufficiently many	
740	keywords, i.e., at least 5 (out of 16/32) keywords	
741	and at least 10% of the data, such that enough data	
742	is present to create the plots. For the analysis of	
743	frequency of keywords over time, we calculate the	
744	percentages normalized for each keyword, i.e., a	
745	value of $p\%$ in year y implies that in year y $p\%$	
746	of all sentences with this keyword occurred. The	
747	trends are shown in the Appendix in Figure 9 for	
748	women and in Figure 10 for migrants.	
749	Parties	
750	We identified political parties by searching for party	
751	names within parentheses, a conventional nota-	
	tion within parliamentary records to denote the	
	speaker’s party affiliation, as seen in examples like	
	" <i>Benjamin Strasser (FDP): Sehr geehrter Präsi-</i>	
	<i>dent!...</i> ". Following automated extraction, we con-	
	ducted a manual review to verify the correctness	
	of the party associations, which resulted in 3,499	
	out of 58k records with party information spanning	
	from 1940 to 2022.	
	List of the parties included in the dataset, along	
	with the variations of their names or abbrevi-	
	ations as they have been recorded: AfD (AL-	
	ternative for Germany); Die Linke (The Left)	
	with variations such as PDS, Gruppe der PDS;	
	Bündnis 90/Die Grünen (Alliance 90/The Greens);	
	CDU/CSU (Christian Democratic Union/Christian	
	Social Union); SPD (Social Democratic Party of	
	Germany); FDP (Free Democratic Party); DP	
	(German Party) with variations such as DP/DPB,	
	DP/FVP, FVP; GB/BHE (All-German Bloc/League	
	of Expellees and Deprived of Rights); KPD (Com-	
	munist Party of Germany); BP (Bavarian Party);	
	WAV (Economic Reconstruction Union).	
	Links to data and code	
	<i>Open Data:</i> https://www.bundestag.de/	
	services/opendata ;	
	<i>Reichstagsprotokolle:</i>	
	https://www.reichstagsprotokolle.de/ ;	
	OCR-scanned version of Walter et al. (2021):	
	https://tudatalib.ulb.tu-darmstadt.de/	
	handle/tudatalib/2889 ; F1 score implemen-	
	tation: https://scikit-learn.org/stable/	
	modules/generated/sklearn.metrics.f1_	
	score.html ;	

Gold Standard	Original Text	Explanation
(1) Compassionate solidarity towards women (June 29, 1961)	“Im Zusammenhang mit § 1708 BGB hat das Hohe Haus das 18. Lebensjahr als Grenze für die Unterhaltspflicht festgelegt. In den Übergangsvorschriften ist diese Bestimmung für diejenigen, die am 1. Januar 1962 schon das 16. Lebensjahr vollendet haben, aufgehoben worden. Diese Regelung erscheint meiner Fraktion ungerecht, denn dadurch würden beträchtliche Personengruppen aus dieser Unterhaltspflicht herausgenommen. Gerade die Frauen, die unter großen Mühen ihre Kinder z. B. auf die höhere Schule geschickt haben, müßten diese Unkosten ganz allein tragen. ”	The speaker is advocating for extended financial support for mothers and is emphasizing the unfairness of removing maintenance obligations.
(2) Exchange-based anti-solidarity towards migrants (Apr. 19, 2018)	“[...] Lassen Sie mich noch anfügen: Migration ist nicht zwingend erfolgreich – Sie tun immer so, als sei das super –, sie kann scheitern, und sie scheitert vor allem dann, wenn die Qualifikation der Einwanderer niedrig ist. 2013, also vor der sogenannten Flüchtlingswelle, hatten 40 Prozent der Zuwanderer aus dem Nicht-EU-Ausland keinen Abschluss. Seit der Flüchtlingswelle haben die Messerstechereien um 20 Prozent zugenommen, und wir haben importierten Antisemitismus im Land. Ist das eine hervorragend erfolgreiche Migration?”	The text criticizes migration for its negative economic impacts and the disproportionate burden placed by low-skilled immigrants who take more resources and social stability than they contribute.
(3) Mixed stance towards migrants (Feb. 2, 1982)	“[...] Wir müssen akzeptieren, daß wir in wenigen Jahren auch wieder eine höhere Zahl ausländischer Arbeitnehmer in der Bundesrepublik brauchen werden, wie Herr Urbaniak vorhin angedeutet hat. Wir haben also in Wirklichkeit zu einer wirksamen Integration, die allerdings voraussetzt, [...] daß es in der Frage des Anwerbestopps und des Verhinderns der illegalen Einwanderung keine Ausnahmen geben darf, keine Alternative. [...]”	This text acknowledges the economic need for foreign workers and the importance of their integration, yet simultaneously emphasizing strict controls on illegal immigration.
(2) None case (women) (June 17, 2015)	“[...] ‘Wir wollen freie Menschen sein!’ Es gibt wohl keinen besseren Satz, um die heutige Debatte hier im Deutschen Bundestag über den Volksaufstand von 1953 zu eröffnen. [...] Wir erinnern an Frauen und Männer, die vor 62 Jahren viel Mut bewiesen, weil sie der Entwicklung ihres Landes und ihrem eigenen Leben eine andere Richtung geben wollten, weil sie freie Menschen sein wollten. ”	The mention of women is integrated into the broader remembrance of the collective effort of people fighting against oppression without emphasizing any specific women’s issues or needs.

Table 3: Original German texts for the examples from our dataset in 1 showing solidarity/anti-solidarity towards women/migrants. Bold text is the main sentence, the other sentences are for context.



(a) Annotators' agreement

(b) Annotators vs. Model

Figure 7: 7a shows the comparison of annotations between our annotators on a fine-grained level; 7b between the final label from the human annotated dataset and our best model’s prediction (cf. Section 6) on a test set. The former is aggregated over all pairwise comparisons of annotators, thus the matrix is symmetric.

	Women	Migrant	Total per label
Group-based solidarity	112 (3.9%)	188 (6.6%)	300 (10.5%)
Exchange-based solidarity	54 (1.9%)	56 (2%)	110 (3.8%)
Empathic solidarity	125 (4.4%)	21 (0.7%)	146 (5.1%)
Compassionate solidarity	732 (25.6%)	466 (16.3%)	1198 (41.8%)
Solidarity (no subtype)	41 (1.4%)	53 (1.9%)	94 (3.3%)
Total for solidarity	1064 (37.2%)	784 (27.4%)	1848 (64.5%)
Group-based anti-solidarity	10 (0.3%)	197 (6.9%)	207 (7.2%)
Exchange-based anti-solidarity	0 (0%)	48 (1.7%)	48 (1.7%)
Empathic anti-solidarity	17 (0.6%)	3 (0.1%)	20 (0.7%)
Compassionate anti-solidarity	8 (0.3%)	80 (2.8%)	88 (3.1%)
Anti-solidarity (no subtype)	5 (0.2%)	19 (0.7%)	24 (0.8%)
Total for anti-solidarity	40 (1.4%)	347 (12.1%)	387 (13.5%)
Mixed	60 (2.1%)	101 (3.5%)	161 (5.6%)
None	273 (9.5%)	195 (6.8%)	468 (16.3%)
Instances in total	1437 (50.2%)	1427 (49.8%)	2864

(a) Distribution of labels by target group.

Label	Curated	Majority	Most	Single
Group-based solidarity	57	34	11	198
Exchange-based solidarity	19	20	2	69
Empathic solidarity	28	12	2	104
Compassionate solidarity	119	190	12	877
Solidarity (no subtype)	5	6	2	81
Total for solidarity	228	262	29	1329
Group-based anti-solidarity	20	32	21	134
Exchange-based anti-solidarity	11	10	5	22
Empathic anti-solidarity	1	13	5	1
Compassionate anti-solidarity	1	25	4	58
Anti-solidarity (no subtype)	2	0	3	19
Total for anti-solidarity	35	80	38	234
Mixed	21	24	16	100
None	84	79	19	286
Instances per label level (out of 2864)	368	445	102	1949

(b) Distribution of instances per label level. *Curated*: labels established by manual revision by an expert; *majority*: labels assigned when more than half of annotators agree on the same label; *most*: labels established by the highest number of selections, even if it is not chosen by the majority; *single*: instances with only one annotation.

Table 4: Human annotated dataset statistics.

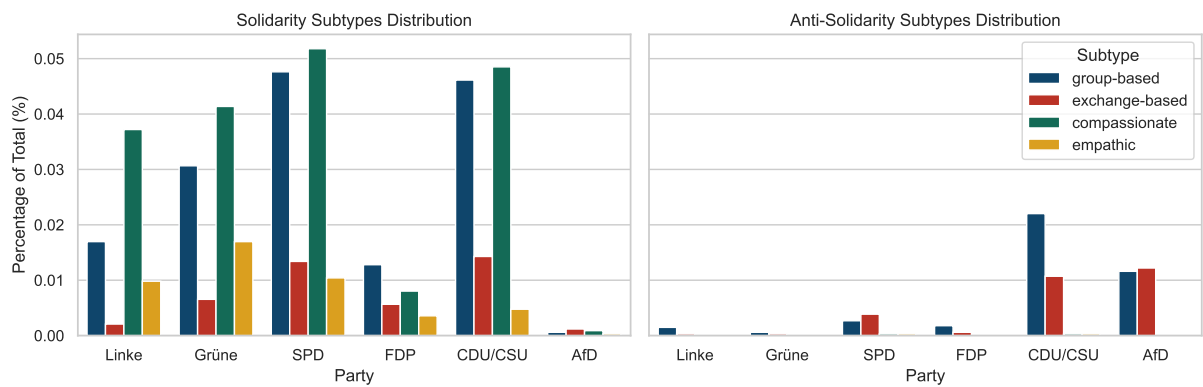


Figure 8: Distribution of *(anti-)solidarity* subtypes across selected political parties, ordered from the most left-wing to the most right-wing. Each subtype's percentage represents its share of the total dataset.

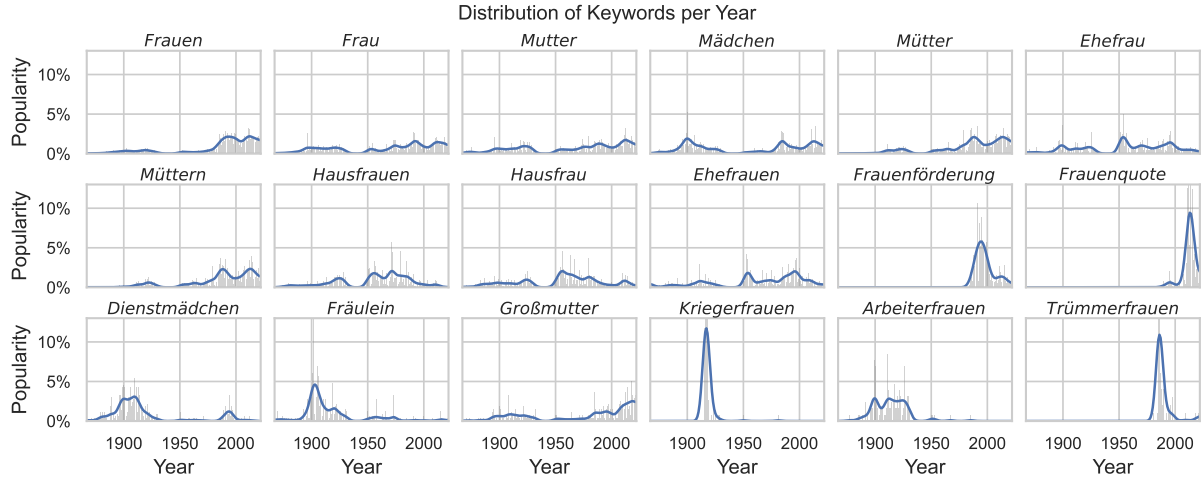


Figure 9: Distribution of all Woman keywords over the years, normalized per keyword. The keywords are sorted by frequency, which means that the reliability decreases towards the bottom-right.

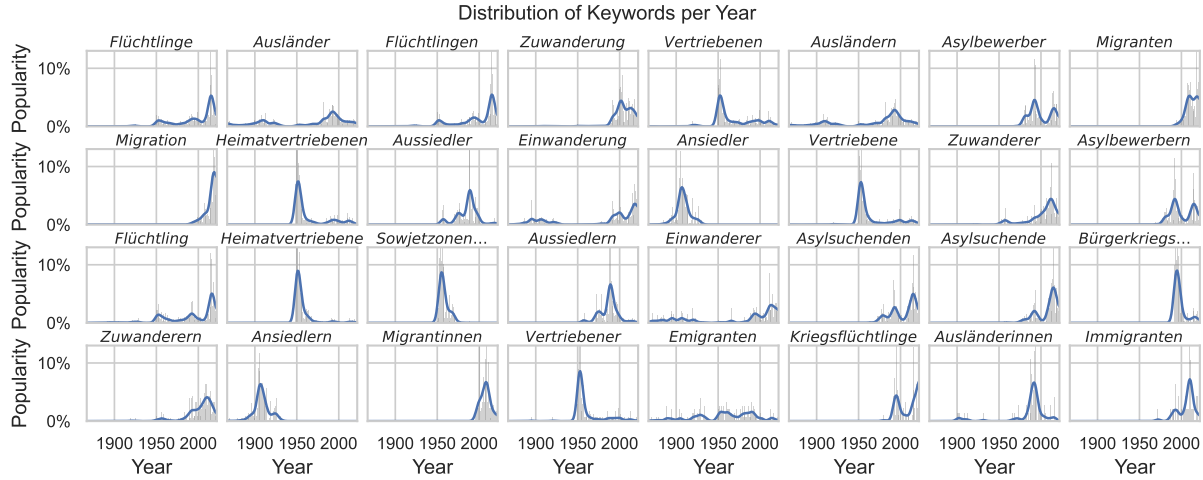


Figure 10: Distribution of all Migrant keywords over the years, normalized per keyword. The keywords are sorted by frequency, which means that the reliability decreases towards the bottom-right.

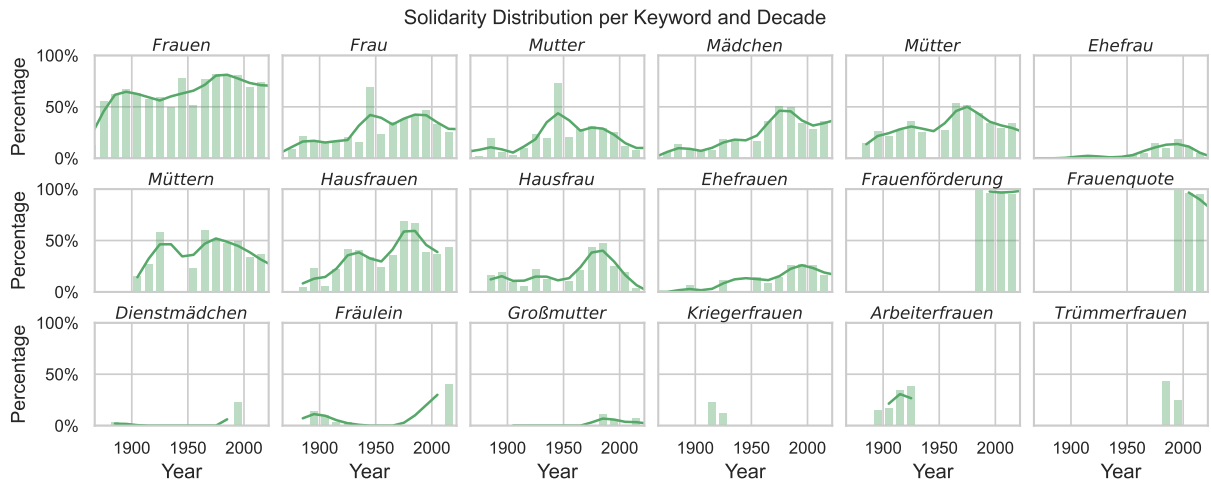


Figure 11: Percentage of sentences showing solidarity per decade for all Woman keywords. The keywords are sorted by frequency, which means that the reliability decreases towards the bottom-right.

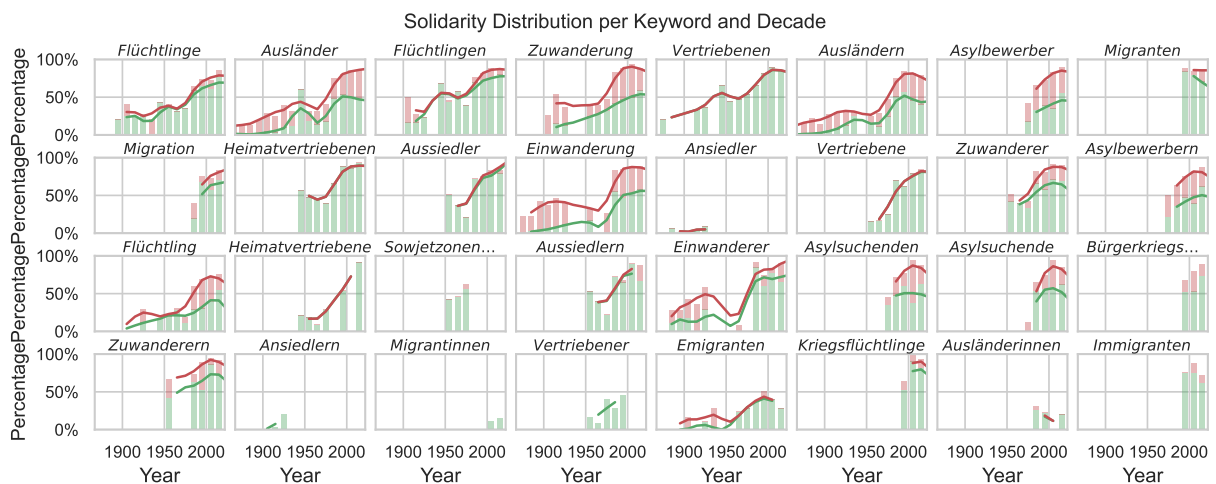


Figure 12: Percentage of sentences showing solidarity/anti-solidarity per decade for all Migrant keywords. The keywords are sorted by frequency, which means that the reliability decreases towards the bottom-right.

Text	Gold Standard, Alternative Label & Explanation
<p>“[...] wenn nachher irgendwelche Schwierigkeiten bei der Rückzahlung der Darlehne entstehen, man nicht nach der Strenge des Gesetzes auf dem Schein bestehen und die Rückzahlung unter allen Umständen fordern müsse. Ich habe aber noch eins ganz kurz zu bemerken: die Ausländer sollen auch mitberücksichtigt werden; nach unsern Beschlüssen in der Kommission würden die Ausländer in derselben Weise behandelt werden. Auch da würde man erst fragen: bist du arm und hilfsbedürftig geworden, dann bekommst du eine Beihilfe als Geschenk; willst du dich weiter ansiedeln im Lande, dann bekommst du ein Darlehn. Ich möchte hier wiederholen: wir haben bei den deutschen Reichsangehörigen, die in andern Ländern geschädigt worden sind durch Revolution, uns nicht damit begnügt, daß sie eine Unterstützung bekommen haben, sondern wir haben erklärt: der Mann ist geschädigt, und er muß daher eine Entschädigung für seine Verluste erhalten, wenn er seiner Pflicht strengster Neutralität genügt hat, und diese Ansprüche haben wir nicht nur für Hilfsbedürftige erhoben, sondern auch für recht reiche Leute. [...]”</p>	<p>Compassionate solidarity or group-based solidarity towards migrants (Apr. 22, 1904)</p> <p>On the one hand, this text expresses compassionate solidarity by offering assistance to those in need, such as financial aid and loans, without requiring reciprocation. On the other hand, the speaker proposes treating foreigners and nationals equally in hardship, expressing group-based solidarity through a unified approach to support.</p>
<p>“Dann bitte ich um genaue Nennung. Ich finde es in der Tat nicht sonderlich sinnvoll, daß wir amtliche deutsche Dokumente in einer nichtamtlichen Sprache abfassen. Wenn das geschehen ist, werden wir gerne darauf hinwirken, daß das geändert wird. Ich möchte nur darauf aufmerksam machen, daß aus Ihrer Frage der völlig gegenteilige Sinn herauszulesen war, nämlich daß Sie offensichtlich wünschsten, Ausländer, die in Deutschland Examen ablegen, sollten diese Diplome in ihrer eigenen Sprache ausgefertigt bekommen, was sicherlich nicht unsere Aufgabe sein kann und auch nicht sehr sinnvoll wäre. [...]”</p>	<p>Empathic anti-solidarity, compassionate anti-solidarity or none towards migrants (Sept. 26, 1974)</p> <p>The text suggests empathic anti-solidarity by expecting foreign students to conform to German norms and shows compassionate anti-solidarity by deeming diplomas in native languages unnecessary. However, its focus on administrative details without strong bias classifies it as none.</p>
<p>“Die Zieglerarbeit ist eine schwere, sogar eine sehr schwere; das wird allgemein anerkannt, auch von allen Gewerbeinspektoren. Es wird deshalb auch ziemlich häufig für wünschenswert erklärt, daß die Arbeiterinnen aus diesem Produktionszweige mehr und mehr verdrängt werden. [...] Wir sind sicherlich der Ansicht, daß die Arbeit auf den Ziegeleien im allgemeinen für Frauen nicht geeignet ist. Deshalb können wir uns auch durchaus damit einverstanden erklären, daß man die Arbeit der Frauen auf den Ziegeleien erheblich eingeschränkt hat. Wir wünschen, daß man darin weiter fortfährt, auch selbst dann, wenn dadurch vielleicht zunächst ein gewisser Widerstand nicht nur bei den Unternehmern, sondern auch sogar bet den Arbeitern selber erzeugt werden wird. Denn darüber sind die Berichte ziemlich einig, daß, wenn die Frauenarbeit nicht erheblich eingeschränkt würde, dann die Frauen gesundheitliche und sittliche Schädigungen davontragen.”</p>	<p>Empathic anti-solidarity or mixed stance towards women (Jan. 13, 1897)</p> <p>The text can be classified as empathic anti-solidarity by suggesting that women be excluded from brickmaking, which supports traditional roles that limit their opportunities. It also presents a mixed stance by recognizing the job’s difficulty and proposing to restrict women’s employment for their protection, which can be viewed as conditional support and simultaneously – an imposed restriction.</p>
<p>“Wir wünschen das nicht im Interesse der Frauen, wir haben uns mit den Frauenrechtlerinnen noch niemals auf eine Stufe gestellt. Wir wünschen das nicht im Namen der Frauen und tm Interesse der Frauen, sondern im Gesamtinteresse des deutschen Volkes, weil wir der Meinung sind, daß bei Mitwirkung der Frauen mehr Verständnis für die Angeklagten und ein sozialerer Geist sich in der Rechtsprechung durchsetzen wird. Deshalb bitten wir Sie, bei den Beschlüssen zweiter Lesung zu bleiben. [...]”</p>	<p>Group-based solidarity or exchange-based solidarity towards women (March 8, 1921)</p> <p>This text shows group-based solidarity by promoting women’s participation in the judiciary to improve legal proceedings for societal benefit. It can also be interpreted as exchange-based solidarity, as it highlights the reciprocal advantages of women’s inclusion.</p>

Table 5: Examples of divergence between our annotators. We mark the gold label bold and add explanations of why two or more labels could be correct, to illustrate the difficulty of this task. Bold text is the main sentence, the other sentences are for context.

	Original Text	Translation
<p>(1) Gold standard: group-based solidarity towards migrants</p> <p>Predicted label: none (Sept. 14, 1989)</p>	<p>“[...] Dr. Hirsch: Herr Minister Möllemann, es ist doch wohl so, daß es sich nicht um das Geld des Bundes oder um das Geld eines Landes handelt, sondern immer um das Geld des Steuerzahlers und daß dementsprechend verlangt werden kann, daß Bund und Länder gemeinsam das tun, was im Interesse der Bevölkerung, zu der dann ja auch die Aussiedler gehören, notwendig und richtig ist. [...]”</p>	<p>“[...] Dr. Hirsch: Mr. Minister Möllemann, it is indeed the case that it is not about the federal government’s money or a state’s money, but always about the taxpayer’s money, and accordingly, it can be demanded that the federal and state governments together do what is necessary and right in the interest of the population, which then also includes the expellees. [...]”</p>
<p>Model’s Explanation</p>	<p>[...] The appropriate high-level category for this text is <i>NONE</i>, as it neither promotes support nor opposition towards migrants but rather discusses financial governance regarding a subset of migrants without an emotional or ideological angle.</p>	
<p>(2) Gold standard: exchange-based solidarity towards migrants</p> <p>Predicted label: mixed stance (May 26, 1869)</p>	<p>“[...] Ich glaube wohl nicht näher aus die materiellen Unzuträglichkeiten eingehen zu sollen, die damit verbunden sind, wenn man die Gegenseitigkeit in der Gesetzgebung dem Ausländer gegenüber streng ausrecht erhält. Es liegt ja doch wirklich im Interesse der Staaten des Norddeutschen Bundes, daß man diejenigen Ausländer, die uns etwas Gutes bringen können, zum Gewerbebetrieb zuläßt und uns nicht selbst damit straft, daß man die Gewerbetreibenden nicht zuläßt, in deren Heimath vielleicht noch sehr schlechte Gewerbe Gesetze existiren. [...]”</p>	<p>“[...] I believe I should not delve into the material inconveniences associated with maintaining strict reciprocity in legislation towards foreigners. It is indeed in the interest of the states of the North German Confederation to allow those foreigners who can bring us benefits to engage in business, and not to punish ourselves by denying business people who may come from countries with very poor trade laws. [...]”</p>
<p>Model’s Explanation</p>	<p>[...] This could be interpreted as exchange-based solidarity to some extent. However, the core of the discussion is legal and pragmatic, rather than being based on an ethos of solidarity or anti-solidarity as defined. [...] It is important to note that without the broader context of the discussion or the surrounding text, it is difficult to ascertain the overall stance regarding migrants beyond this specific legislative detail.</p>	
<p>(3) Gold standard: compassionate solidarity towards women</p> <p>Predicted label: group-based solidarity (March 13, 2003)</p>	<p>“[...] Wir werden es nicht zulassen, dass künftig ausländische Ehefrauen vier Jahre in unzumutbaren Ehen verharren, um nicht aus Deutschland ausgewiesen zu werden. Das werden Sie nicht erreichen. Ich will es überhaupt nicht verhehlen: Wir haben noch unendlich viel zu tun, bevor wir gerade im Erwerb-leben und bei der traditionellen Rollen- und Aufgabenverteilung eine faktische Gleichstellung von Frauen und Männern erreicht haben. [...]”</p>	<p>“[...] We will not allow foreign wives to remain in unbearable marriages for four years to avoid being deported from Germany. You will not achieve this. I do not want to conceal it at all: we still have an immense amount of work to do before we achieve actual equality between women and men in the workforce and in the traditional distribution of roles and tasks. We take this responsibility seriously. [...]”</p>
<p>Model’s Explanation</p>	<p>[...] The most fitting subtype of solidarity in this context is <i>GROUP-BASED SOLIDARITY</i> because the text discusses the collective responsibility to promote equality and the shared efforts needed to achieve substantial change in societal structures concerning the roles and treatment of women.</p>	
<p>(4) Gold standard: compassionate solidarity towards women</p> <p>Predicted label: empathic solidarity (Apr. 5, 1922)</p>	<p>“[...] Wenn man von den Leuten fordert, daß sie sich anmelden sollen, wenn alle möglichen Einschränkungen für den Aufenthalt getroffen werden, soll man diese Ausländer auch wissen lassen, nach welchen Vorschriften sie sich zu richten haben. Es ist schikanös und heimtückisch, die Leute mit solchen Vorschriften zu überfallen. Wie das in der Praxis vor sich geht, brauche ich Ihnen nicht lange zu schildern. [...]”</p>	<p>“[...] When people are required to register, and all kinds of restrictions on their stay are imposed, these foreigners should also be informed about the regulations they must follow. It is harassing and treacherous to ambush people with such regulations. I do not need to describe to you at length how this happens in practice. [...]”</p>
<p>Model’s Explanation</p>	<p>[...] The text seems to transmit a sense of <i>EMPATHIC SOLIDARITY</i> because it addresses the unfairness and challenges that foreigners face and advocates for fairness and clear guidelines, which shows a degree of empathy for their situation.</p>	

Table 6: Example predictions by GPT-4 for both target groups in 0-shot setting for migrants and women. Bold text is the main sentence, the other sentences are for context.

Target group of the text		Text to annotate in the column "Sentences" consisting of "Previous", "Middle" and "Next". The surrounding sentences (<i>previous</i> and <i>next</i>) are provided for better understanding of the context.			Column for high-level categorization	Categorization within (anti-)solidarity
Category	Sentences			Main category	(Anti-)solidarity subcategory	
	Previous	Middle	Next			
Migrant	Ich denke, einige unserer Möglichkeiten in diesem Bereich sind längst noch nicht ausgeschöpft. Nein, die politisch Verfolgten sind erkennbar nicht das Problem in Deutschland. Das Problem ist die ungesteuerte, gegenwärtig zu massive Einwanderung nach Deutschland.	Es ist nicht nur das Recht der Politik, es ist die Pflicht der Politik, die Zuwanderung nach Deutschland zu reduzieren, sie steuerbar zu machen und dabei die — durchaus beachtliche — Aufnahme- und Integrationskraft der deutschen Gesellschaft nicht zu überschätzen.	Im Bereich der Zuwanderung befinden sich die Bundesregierung und die Koalition leider noch nicht auf der Höhe des Problems. In den kommenden Jahren werden jährlich 200 000 Spätaussiedler und mehr als 100 000 Menschen im Wege der Familienzusammenführung zu uns kommen. Diese Menschen sind Einwanderer.	Anti-solidarity	Group-based	

(a) Columns for high-level and (anti-)solidarity categorizations.

Resource on the basis of which solidarity or anti-solidarity is expressed		Indicators for a specific type of (anti-)solidarity (given in drop-down boxes for a respective subtype)			A free-form explanation for choosing a label (a short comment in 1-2 sentences)
Resource	Indicator - group-based	Indicator - compassion	Indicator - exchange-based	Indicator - empathic	Free text
access to Germany	'resistance to integration'				The speaker points out that the government has the duty to restrict and limit the number of migrants and to respect the wish of the people to not allow to many people to migrate.

(b) Columns for providing explanations.

Figure 13: Example of the annotation process from the annotation file. 13a illustrates the step where annotators choose a high-level label and an (anti-)solidarity subcategory, if applicable. 13b shows columns for detailed explanations, including the choice of a resource, an indicator, and providing free-text commentary.

Analyze the following German text and classify it into one of the high-level categories regarding migrants (refugees — Flüchtlinge, expellees — Vertriebene, asylum seekers — Asylbewerber, immigrants — Einwanderer, and other migrant categories within Germany): SOLIDARITY, ANTI-SOLIDARITY, MIXED, or NONE. If applicable, further specify by choosing the most appropriate subtype (EMPATHIC, EXCHANGE-BASED, GROUP-BASED, COMPASSIONATE) within SOLIDARITY or ANTI-SOLIDARITY. Begin your response by providing the high-level category and then the subtype, if applicable.

- **SOLIDARITY:** Involves expressions that promote understanding, support, and unity with different groups or individuals (migrants in our case), often emphasizing shared goals, compassion, mutual assistance, and empathic understanding. Consider cases with even slight expressions of solidarity, regardless of the main topic of the text.
- **ANTI-SOLIDARITY:** Entails expressions that show opposition, disregard, or exclusion towards certain groups or individuals (migrants in our case). This includes emphasizing differences, denying the need for support or assistance, highlighting unequal exchanges between groups, and disregarding the unique characteristics or needs of certain groups. Even slight expressions of anti-solidarity should be considered, irrespective of the primary focus of the text.
- **MIXED:** A mixed stance toward migrants is characterized by the presence of both supportive and opposing expressions within the same text. This stance emerges in discussions where acknowledgment of migrants' rights, contributions, or needs is juxtaposed with limitations, conditions, or reservations that counteract or diminish the initial support. Key features of a mixed stance include (but are not limited with): conditional hospitality and selective support; balanced policies (e.g. improve the situation of migrants already within the country, while simultaneously seeking to regulate or limit further influx); expressions of empathy or concern for migrants' hardships, contrasted with discussions on practical constraints, such as societal integration challenges, or national security concerns.
- **NONE:** Texts which neither express solidarity nor anti-solidarity toward migrants in Germany, reflecting a neutral position or the absence of any specific stance. The absence of overt support or opposition does not automatically lead to a NONE classification; subtle cues or implicit messages may still align with solidarity or anti-solidarity categories.

If the text falls into SOLIDARITY or ANTI-SOLIDARITY, please specify further by choosing the most appropriate subtype from the following, after the initial high-level classification.

For SOLIDARITY: EMPATHIC SOLIDARITY, EXCHANGE-BASED SOLIDARITY, GROUP-BASED SOLIDARITY, COMPASSIONATE SOLIDARITY. Definitions:

- **EMPATHIC SOLIDARITY:** Is coded when a group is different from others and this should be recognized, supported, valued. In applying empathic solidarity to migrants, this can be expressed by (but not limited with): recognition of diversity and individuality; emphasis on the importance of preserving migrants' identities when integrating them into new communities; advocating for the right to live authentically without fear of persecution or discrimination; challenging stereotypes, prejudices against migrants.
- **EXCHANGE-BASED SOLIDARITY:** Is coded when a speaker refers to the usefulness of 'exchange partners' in terms of their actual or future contributions (economic, cultural, or social, etc.) or willingness to contribute. In applying exchange-based solidarity to migrants, this can be expressed by (but not limited with): emphasis on the importance of migrants' work, skills, and cultural diversity as essential for the host society; support for migrants which is framed as an investment in individuals who contribute to the community.
- **GROUP-BASED SOLIDARITY:** Is coded when solidarity is based on the idea of unity and support among members of a group, driven by shared characteristics, goals, interests, values and norms, or common rights and duties. The support might be driven by shared characteristics or challenges aiming at broader societal change; fostering inclusivity, equality, societal cohesion (difference from compassionate solidarity). In applying group-based solidarity to migrants, this can be expressed by (but not limited with): a unified effort to address and advocate for migrants' rights, equality, and representation; advocacy aimed at ensuring migrants' full integration; active stance against discrimination and xenophobia.
- **COMPASSIONATE SOLIDARITY:** Emphasizes providing support to marginalized, disadvantaged, or vulnerable groups, focusing on aid without expecting anything in return. It involves recognizing vulnerabilities, advocating for assistance to alleviate hardships, and offering support purely based on need. While not all indicators must be present, the core of compassionate solidarity lies in acknowledging and addressing the needs of those in vulnerable positions. The support is aimed at addressing vulnerabilities and needs without expectation of reciprocity; alleviating suffering and providing assistance based on need (difference from group-based solidarity).

For ANTI-SOLIDARITY: GROUP-BASED ANTI-SOLIDARITY, EXCHANGE-BASED ANTI-SOLIDARITY, EMPATHIC ANTI-SOLIDARITY, COMPASSIONATE ANTI-SOLIDARITY. Definitions:

- **EMPATHIC ANTI-SOLIDARITY:** Refers to the cases when a speaker refers to a group indicating that this group is different from everyone else, but this difference should not be respected or recognised. This can be expressed in: an overt or covert refusal to acknowledge the validity of a group's diverse identities, roles, and choices. Opposition to the group's efforts to define their identity, roles, and aspirations on their own terms. Promotion of a return to or maintenance of traditional societal roles and structures that restrict the group's opportunities.
- **EXCHANGE-BASED ANTI-SOLIDARITY:** Focuses on the idea that some groups take more than they give. It brings attention to groups that seem to receive a lot of resources or support, but do not contribute much or cause problems. This perspective supports punishing such groups, giving them less help, or making them contribute more. This can be expressed in suggesting that this group should be punished, or get less than they are currently getting; calling for the group to give or do more in return for help. Remember: exchange-based criticisms focus on the economic implications and perceived unfairness in resource distribution (difference from compassionate anti-solidarity).
- **GROUP-BASED ANTI-SOLIDARITY:** Points to an out-group characterised by pronounced differences, and an expression of exclusion. This can be expressed in: a stronger emphasis on one's own group's interests (e.g., We must stand up for our own interests.); resistance to support of, assimilation or integration with other groups. With migrants, this can be expressed as an emphasis on protecting interests of the native population at the expense of including migrants; portrayal of migrants as fundamentally different and unable to integrate, thus excluding them; a call for migrants to conform to existing norms without offering them full participation in society.
- **COMPASSIONATE ANTI-SOLIDARITY:** It is about disregarding and excluding individuals or subgroups, refusing to support them either on the basis that they are perceived to be in a good position already, or are viewed as undeserving of help. When applied to migrants, this can be expressed by (but not limited with): denial of systemic issues (protections or affirmative actions are unnecessary or overreaching); security concerns, skepticism about the motives and legitimacy of asylum seekers. Remember, compassionate criticisms might question the necessity of certain protections or the legitimacy of migrants (difference from exchange-based anti-solidarity).

Think step by step. Begin by analyzing the text to identify its high-level category related to migrants: SOLIDARITY, ANTI-SOLIDARITY, MIXED, or NONE. If the text falls under SOLIDARITY or ANTI-SOLIDARITY, further specify by identifying the most appropriate subtype: EMPATHIC, EXCHANGE-BASED, GROUP-BASED, or COMPASSIONATE. After your comprehensive analysis, conclude with 'Label:' followed by the most fitting category name. If identifying a subtype, combine it with the high-level category, such as GROUP-BASED SOLIDARITY, EXCHANGE-BASED SOLIDARITY, EMPATHIC SOLIDARITY, COMPASSIONATE SOLIDARITY, GROUP-BASED ANTI-SOLIDARITY, EXCHANGE-BASED ANTI-SOLIDARITY, EMPATHIC ANTI-SOLIDARITY, COMPASSIONATE ANTI-SOLIDARITY. Please choose just one label.

Figure 14: 2-step Prompt for Migrants used for GPT 0-shot experiments

Analyze the following German text and classify it into one of the high-level categories regarding women: SOLIDARITY, ANTI-SOLIDARITY, MIXED, or NONE. If applicable, further specify by choosing the most appropriate subtype (EMPATHIC, EXCHANGE-BASED, GROUP-BASED, COMPASSIONATE) within SOLIDARITY or ANTI-SOLIDARITY. Begin your response by providing the high-level category and then the subtype, if applicable.

- **SOLIDARITY:** Involves expressions that promote understanding, support, and unity with different groups or individuals (women in our case), often emphasizing shared goals, compassion, mutual assistance, and empathic understanding. Consider cases with even slight expressions of solidarity, regardless of the main topic of the text.
- **ANTI-SOLIDARITY:** Entails expressions that show opposition, disregard, or exclusion towards certain groups or individuals (women in our case). This includes emphasizing differences, denying the need for support or assistance, highlighting unequal exchanges between groups, and disregarding the unique characteristics or needs of certain groups. Even slight expressions of anti-solidarity should be considered, irrespective of the primary focus of the text.
- **MIXED:** A mixed stance toward women is characterized by the presence of both supportive and opposing expressions within the same text. This stance emerges in discussions where acknowledgment of women's rights, contributions, or needs is juxtaposed with limitations, conditions, or reservations that counteract or diminish the initial support. Key features of a mixed stance include: acknowledgment followed by restriction; conditional support which is predicated on certain traditional or societal expectations; statements of recognition of women's contributions or rights followed by contradicting actions or policies.
- **NONE:** Texts that neither express solidarity nor anti-solidarity with women, reflecting a neutral position or absence of any specific stance towards women's issues. Remember, discussions that touch upon women's familial roles (e.g., as mothers, wives, or daughters) and their societal implications are relevant and should be analysed for expressions of solidarity or anti-solidarity. The absence of overt support or opposition doesn't automatically lead to a NONE classification; subtle cues or implicit messages may still align with solidarity or anti-solidarity categories.

If the text falls into SOLIDARITY or ANTI-SOLIDARITY, please specify further by choosing the most appropriate subtype from the following, after the initial high-level classification:

- **EMPATHIC SOLIDARITY:** is coded when a group is different from others and this should be recognized, supported, valued. In applying empathic solidarity to women, this can be expressed by (but not limited with): celebrating the varied roles and achievements of women; supporting women's freedom to express their views and make personal and professional choices; challenges traditional stereotypes about women's roles and abilities, promoting a broader and more inclusive understanding of women's capabilities and choices.
- **EXCHANGE-BASED SOLIDARITY:** is coded when a speaker refers to the usefulness of 'exchange partners' in terms of their actual or future contributions or willingness to contribute. In applying exchange-based solidarity to women, this can be expressed by (but not limited with): focus of contributions and the expectation of receiving something in return; mentions of rewarding or supporting based on past, present, or anticipated contributions; discussions about the need for exchange partners to contribute more for support.
- **GROUP-BASED SOLIDARITY:** is coded when solidarity is based on the idea of unity and support among members of a group, driven by shared characteristics, goals, interests, values and norms, or common rights and duties. The support might be driven by shared characteristics or challenges aiming at broader societal change; fostering inclusivity, equality, societal cohesion (difference from compassionate solidarity). In applying group-based solidarity to women, this can be expressed by (but not limited with): a unified effort to address and advocate for women's rights, equality, and representation; advocacy might be aimed at ensuring women's full integration into all aspects of society.
- **COMPASSIONATE SOLIDARITY:** emphasizes providing support to marginalized, disadvantaged, or vulnerable groups, focusing on aid without expecting anything in return. It involves recognizing vulnerabilities, advocating for assistance to alleviate hardships, and offering support purely based on need. The support is aimed at addressing vulnerabilities and needs without expectation of reciprocity; alleviating suffering and providing assistance based on need (difference from group-based solidarity).

For ANTI-SOLIDARITY: GROUP-BASED ANTI-SOLIDARITY, EXCHANGE-BASED ANTI-SOLIDARITY, EMPATHIC ANTI-SOLIDARITY, COMPASSIONATE ANTI-SOLIDARITY. Definitions:

- **EMPATHIC ANTI-SOLIDARITY:** Refers to the cases when a speaker refers to a group indicating that this group is different from everyone else, but this difference (needs, characteristics, beliefs, opinions) should not be respected or recognised. In applying empathic anti-solidarity to women, this can be expressed by (but not limited with): an overt or covert refusal to acknowledge the validity of women's diverse identities, roles, and choices; opposition to women's efforts to define their identity, roles, and aspirations on their own terms; promotion of a return of traditional societal roles and structures that restrict women's opportunities.
- **EXCHANGE-BASED ANTI-SOLIDARITY:** focuses on the idea that some groups take more than they give. It brings attention to groups that seem to receive a lot of resources or support, but do not contribute much or cause problems. This perspective supports punishing such groups, giving them less help. In applying exchange-base anti-solidarity to women, this can be expressed by (but not limited with): talking about women as a group that has received a lot, but has not given much back or has caused harm; suggesting that women should be punished, or get less than they are currently getting; calling for women to give or do more in return for help.
- **GROUP-BASED ANTI-SOLIDARITY:** often points to exclusion and lack of support for a particular group based on certain differences. A strong emphasis is made on one's own group's interests (e.g., We must stand up for our own interests...); resistance to assimilation or integration with other groups. In applying group-based anti-solidarity to women, this can be expressed by (but not limited with): resistance to changing societal norms and integrating women into roles beyond traditional ones; excluding women from certain professional or public roles based on perceived gender norms or capabilities, maintaining a societal structure that marginalizes women's broader participation.
- **COMPASSIONATE ANTI-SOLIDARITY:** it is about disregarding and excluding individuals or subgroups, refusing to support them either on the basis that they are perceived to be in a good position already, or are viewed as undeserving of help. In applying compassionate anti-solidarity to women, this can be expressed by (but not limited with): denial of systemic issues (protections are unnecessary or overreaching); refusal to acknowledge the specific needs or challenges faced by women; emphasizing women's choice in scenarios where systemic barriers limit those choices; discourse shifts focus from societal or structural solutions to individual responsibility; reinforcing stereotypes or generalizations about women.

Think step by step. Begin by analyzing the text to identify its high-level category related to women: SOLIDARITY, ANTI-SOLIDARITY, MIXED, or NONE. If the text falls under SOLIDARITY or ANTI-SOLIDARITY, further specify by identifying the most appropriate subtype: EMPATHIC, EXCHANGE-BASED, GROUP-BASED, or COMPASSIONATE. After your comprehensive analysis, conclude with 'Label:' followed by the most fitting category name. If identifying a subtype, combine it with the high-level category, such as GROUP-BASED SOLIDARITY, EXCHANGE-BASED SOLIDARITY, EMPATHIC SOLIDARITY, COMPASSIONATE SOLIDARITY, GROUP-BASED ANTI-SOLIDARITY, EXCHANGE-BASED ANTI-SOLIDARITY, EMPATHIC ANTI-SOLIDARITY, COMPASSIONATE ANTI-SOLIDARITY. Please choose just one label.

Figure 15: 2-step Prompt for Women used for GPT 0-shot experiments