Moral Framing in Politics (MFiP):
A new resource and models for moral framing

Ines Rehbein **Ines Reinig** **Simone Paolo Ponzetto**
Data and WEb Science Group
University of Mannheim, Germany
{rehbein|reinig|ponzetto}@uni-mannheim.de

Abstract

The construct of morality permeates our entire lives and influences our behavior and how we perceive others. It therefore comes at no surprise that morality also plays an important role in politics, as morally framed arguments are perceived as more appealing and persuasive. Thus, being able to identify moral framing in political communication and to detect subtle differences in politicians' moral framing can provide the basis for many interesting analyses in the political sciences. In the paper, we release MoralFramingInPolitics (MFiP), a new corpus of German parliamentary debates where the speakers' moral framing has been coded, using the framework of Moral Foundations Theory (MFT). Our fine-grained annotations distinguish different types of moral frames and also include narrative roles, together with the moral foundations for each frame. We then present models for frame type and moral foundation classification and explore the benefits of data augmentation (DA) and contrastive learning (CL) for the two tasks. All data and code will be made available to the research community.

1 Introduction

Recent years have seen a growing interest in detecting moral values in political communication, trying to identify moral frames used by political actors or parties to convey their messages (see Fulgoni et al. (2016); Roy and Goldwasser (2021); Johnson and Goldwasser (2018); Araque et al. (2020); Hoover et al. (2020); Beiró et al. (2023), *iter alia*).[1] Many of these studies are based on Moral Foundations Theory (MFT) (Haidt et al., 2009; Graham et al., 2013), a descriptive, pluralist theory of morality rooted in social psychology. MFT assumes the existence of a number of moral intuitions or "gut feelings" that drive moral reasoning and turn it into rationalisation. Knowing which "moral intuitions" are held by certain populations has been used to investigate a wide range of research questions, such as differences in moral values between cultures (Wu et al., 2023) or the driving factors behind human attitudes and behaviour, such as COVID-19 vaccine hesitancy (Weinzierl and Harabagiu, 2022).

Previous work has measured moral sentiment in text, using dictionary-based methods (Fulgoni et al., 2016; Jung, 2020; Weinzierl and Harabagiu, 2022; Wu et al., 2023; Stanier and Shin, 2024). Another common approach uses supervised ML and treats moral value prediction as a multi-label text classification task, where each text (such as tweets or Reddit posts) has been assigned one or more moral foundations (MFs) (Hoover et al., 2020; Trager et al., 2022).

While both approaches can easily be applied to large data, they come with certain limitations. Dictionary-based methods are insensitive to word meaning in context and also cannot handle negation. The document-based prediction of morality, on the other hand, is rather coarse-grained and neither provides information on *which passages* in the text carry the moral sentiment, nor on *who is the target* of the moral framing. This impairs the usefulness of the predictions for further analysis.

We are not the first to point out these shortcomings. For example, Roy et al. (2021a) discuss moral values in tweets by US politicians on the topic of abortion and note that both, Democrats and Republicans, tend to use the moral foundation Care-Harm to frame the topic (see §2.1 for more details on MFT). While they did not observe strong differences in the use of MFs, they noticed systematic differences regarding the *targets* of the moral sentiment, i.e., the entity in need of Care, with Republicans stressing the need to protect unborn life while

[1] Please note that the focus of this paper is *not* on aligning LLMs with human moral values, nor on investigating moral biases in LLMs. Instead, we are interested in using NLP techniques for analysing *moral rhetoric and framing in political text* to support analyses in the political and social sciences.

Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, pages 34643–34663
November 4-9, 2025 ©2025 Association for Computational Linguistics

Democrates mostly focus on women's needs.

In our work, we address these shortcomings by presenting a fine-grained annotation scheme for moral framing, capturing different frame types and their moral values. We first present our new annotation framework and then show how we can predict our annotations in a corpus of parliamentary debates. Our main contributions can be summarised as follows.

- We propose a new framework for the annotation of moral framing in political text.
- We release the `MFiP` corpus, a new resource for moral framing in German parliamentary debates.
- We investigate the benefits of data augmentation (DA) and contrastive learning (CL) for the task of MF prediction, with substantial improvements for both techniques.
- We present an evaluation on an out-of-domain testset and show that while DA and CL can help to better generalise, more work is needed on cross-domain modelling of morality.

2 Related Work

There is increased interest in modelling morality in NLP, evidenced by two recent surveys, one focussing on ethics in AI (Vida et al., 2023), the other on modelling morality for text analysis (Reinig et al., 2024). In this section, we start with some background on MFT and then discuss work on predicting moral values in text, with a focus on the political domain and on methods using DA and CL.

2.1 Moral Foundations Theory (MFT)

MFT is a descriptive, pluralist theory of morality that was developed in the field of social psychology (Haidt et al., 2009; Graham et al., 2013). In contrast to monist theories that explain morality in terms of one single principle or dimension, *right–wrong*, MFT believes that the concept of morality is based on more than one such dimension, or foundation. According to MFT, these foundations have been developed during evolution as responses to several adaptive challenges, e.g., the emergence of the PURITY foundation has been driven by the need to avoid pathogens. Moral foundations are seen as intuitions or feelings rather than conscious judgments, which is in contrast to other moral theories that describe moral intuitions as "strong, stable, immediate moral beliefs" (Sinnott-Armstrong et al., 2010) or as moral judgments (McMahan, 2000).

MFT assumes at least five moral intuitions that can be divided into *binding* foundations (ingroup LOYALTY, respect for AUTHORITY, and PURITY) and *individualising* foundations (CARE and FAIRNESS) (Graham et al., 2011). Newer work has proposed that ideas of fairness can be based on different notions of justice, and has further divided the FAIRNESS foundation into EQUALITY and PROPORTIONALITY (Atari et al., 2023) where EQUALITY favours an equal distribution of opportunities and resources while PROPORTIONALITY prefers a distribution in proportion to an individual's merit or contribution.

2.2 Moral framing in political text

Previous work has investigated morality in a variety of political text types, such as news articles (Fulgoni et al., 2016), politicians' tweets (Johnson and Goldwasser, 2018), or user-generated content that reflects different underlying ideologies (Araque et al., 2021). While many studies still make use of dictionaries (Lipsitz, 2018; Kraft, 2018; Jung, 2020; Husson and Palma, 2024), more recent work has argued for a frame-based approach where moral events and their participants are grounded in the text (Roy and Goldwasser, 2021; Zhang et al., 2024). This demand is supported by Fremann et al. (2023); Otmakhova et al. (2024) for the related topic of media framing.

Unfortunately, existing resources are sparse. To our best knowledge, the only available dataset modelling morality at the frame level is the English newswire corpus of Zhang et al. (2024). We address this gap by releasing a new resource for German, capturing moral framing in political debates, with more than 200,000 tokens and $>$5,000 encoded moral frames. Our operationalisation of moral framing also addresses Otmakhova et al. (2024)'s main criticism by modelling the frame targets and narrative roles and combining them with a theory-guided approach, based on MFT.

DA and CL for moral value prediction Previous work has used data augmentation (DA) and contrastive learning (CL) to improve the prediction of moral values. Kobbe et al. (2020) are the first to apply CL for learning better representations of morality, based on Wikipedia articles annotated with MFs in a weak supervision setup. However, their approach did not outperform the baseline, probably due to noise in the data. Zhang et al. (2024) exploit several large resources of moral scenarios for pretraining, reporting modest improvements (<2% F1) for MF prediction on gold frame spans over their best baseline system.

Zangari et al. (2025) use CL to learn representations of morality, integrating events and emotions. However, the data used for CL has been annotated automatically (Greco et al., 2024), using a classifier trained on the *same* data (the Moral Foundations Twitter Corpus of Hoover et al. (2020)) that was later used to evaluate the approach. It thus comes at no surprise that the approach outperforms most other baselines on this particular corpus but fails to do so on another dataset. Park et al. (2024) use SimCSE (Gao et al., 2021) to train morality-sensitive sentence embeddings but do not provide an evaluation of the learned representations.

3 Annotation

We now describe our annotation framework and the annotation process for our new resource, `MoralFramingInPolitics` (`MFiP`).

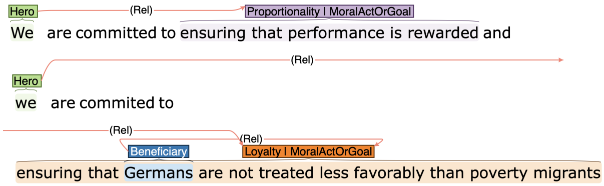
3.1 Annotation scheme

Our concept of a moral frame is inspired by the framing literature (in particular, Entman (1993)). By moral frame, we refer to text spans that express moral values or judge moral behaviour (including moral acts, goals, events, stances). Each moral frame has a frame type and a moral foundation, following MFT (see Table 1). In addition, each frame can have one or more narrative roles that describe the participants in a moral event and are linked to their respective moral frames, as shown in Figure 1.

While our work is similar in spirit to the entity annotations of Roy et al. (2021b), our moral frames substantially extend their framework which only highlights entities and their moral foundations but does not encode the frame span expressing the actual moral act or event.

Moral frames In contrast to previous work, we do not annotate moral values on the level of sentences or documents but, instead, encode textually anchored moral frames and their roles (see Figure 1). As a sentence or document (e.g., a tweet or reddit post) can include multiple moral frames, we argue that assigning MF labels to sentences or documents is suboptimal for at least 2 reasons:

1. Annotations are less informative, as it is unclear *which part* of a text has evoked the MF


We are committed to ensuring that performance is rewarded and

we are commited to

ensuring that Germans are not treated less favorably than poverty migrants

Figure 1: Example from our corpus (en translation) in the INCEpTION annotation platform (Klie et al., 2018).

2. In cases where annotators assign different MF labels to the same text, it remains unclear whether they disagree on the interpretation of the moral value or whether they simply considered different acts/events in the same text.

Grounding the annotations by anchoring moral frames and their participants to specific text spans addresses these issues and also results in more informative representations for analysis.

Specifically, our annotations encode abstract moral values as well as concrete acts and goals that are framed as (im)moral, using the four labels MORALVALUE, IMMORALVALUE, MORALACTORGOAL, and IMMORALACTORGOAL (see examples in Table 1). Additionally, we use the label POLITICALACTORGOAL to code text spans that describe more concrete laws and policy acts. Distinguishing between abstract values and concrete acts and goals will enable us to study how the two interact on a linguistic level.[2]

As shown in Table 1, moral values are typically expressed as NPs and describe abstract concepts (*freedom, injustice, traditional family values*) or symbols that transmit national and religious values (*the Statue of Liberty*). Descriptions of (im)moral acts or goals are typically expressed as VPs (e.g., *saving the planet*) but can also include nominalisations (e.g., *the fight against disposable packaging*). Whether a frame is coded as either moral or immoral depends on the speaker's framing, irrespective of the coder's moral preferences.

Narrative roles In addition to the moral frames, we annotated narrative roles, inspired by the Narrative Policy Framework (NPF) (Shanahan et al., 2017). We consider the following four roles: the *Victim* who is harmed, the *Villain* who is doing the

[2] We expect that the value categories correspond to what is typically referred to as *moralising speech acts*, i.e., concepts and values like *justice* that are presented as universally accepted so that no further justification is needed (Becker et al., 2024).

	Moral Frame	Example	Moral Foundation (MF)
MV	MoralValue	the statue of liberty	Liberty
MV	MoralValue	traditional family values	Authority
IV	ImmoralValue	the communist wall of shame	Purity
MA	MoralActOrGoal	save the planet and the people	Care
MA	MoralActOrGoal	strengthen our German economy	Loyalty
IA	ImmoralActOrGoal	impose draconian penalties for harmless offenses	Proportionality
MA	ImmoralActOrGoal	realise equal opportunities	Equality

Table 1: Examples for different moral frame types of (im)moral values, acts and goals and their corresponding MFs.

harm or causing a problem, the *Hero* who provides a solution to the problem and the *Beneficiary*, i.e., the person or group that benefits from a certain act or policy solution (Figure 1).

Moral Foundations (MF) We further augment our frames with moral foundations: On top of each moral frame, we encode its most relevant MF(s) (Care, Equality, Proportionality, Loyalty, Authority, Purity, Liberty; see A for an overview of MFT and Table 1 for some illustrative examples). While some studies tend to encode the poles of each moral foundation as separate classes (e.g., Care-Harm), MFT does not make any predictions based on this distinction, and psychological measurement tools for MFT like the Moral Foundations Questionnaire also do not measure it; see, e.g., Graham et al. (2009, 2011, 2012). We therefore do not follow this practice but encode each MF as one label only. In addition, guided by recent developments in MFT (Atari et al., 2023), we split the Fairness MF into the new foundations Equality and Proportionality. We also include Liberty as a new foundation, resulting in a set of seven MFs and the additional label General-Moral for general moral statements that do not fit any of the more specific MFs.

3.2 Data and sampling

The data we use in our study are parliamentary debates from the German Bundestag. This choice is motivated by our interest in studying how moral values are used to *frame political issues* and to achieve political goals. In addition, we augment our data with a supplementary dataset of political manifestos, as those include many statements about what ought to be done, often framed in moral terms.

The parliamentary data includes 244 speeches by 192 speakers from 6 parties (+ 4 speeches by non-inscript members of the parliament), with over 5,000 moral frames.[3] We sampled the data across topics (see B for details). To ensure that we cover the parties' main views on each topic, we decided to incude all speeches given by each party on a given topic, meaning that the data is not balanced across parties but reflects each party's speaking time in parliament (and thus the number of seats). The political manifestos are extracted from the Manifestos Project Database (Burst et al., 2022) and cover three controversially discussed topics, namely immigration, culture, and the media, with around 1,500 moral frames. For more details on dataset size and distribution across parties, see Tables 15 and 16 in the appendix.

3.3 Annotation of moral frames and roles

The identification of frames has been carried out by two trained coders, both MA students of linguistics. Each text has been annotated by both coders to ensure high recall. The coders were instructed to first read the whole speech, focussing on the moral values, goals and actions that are presented *by the speaker* as desirable (praiseworthy) as well as the ones framed as undesirable (blameworthy).[4]

3.4 Annotation of moral foundations (MF)

In the next step, we extracted the annotations and clustered the frames into semantically coherent frame groups (for details, see B.4 in the Appendix). Each instance is annotated by four trained coders: two MA students of linguistics, a PhD student and a postdoc, both with a background in computational linguistics. All annotators have received extensive training and feedback during the whole annotation cycle. The coders were presented with the clusters and were asked to assign moral foundations to each frame in the cluster. The motivation for this approach was to speed up the annotation and increase consistency by presenting the coders with sets of (more or less) thematically similar frames.

[3] The exact no. depends on whether we count overlapping but distinct frame spans as the same or different frames.

[4] The coders are instructed to *always encode the speaker's perspective*, not their own values. The detailed guidelines are available from `https://github.com/umanlp/mfip.git`.

Annotation of clusters with MFs We consider MF annotation as a multi-label task where each moral frame is assigned at most two MFs.[5] Figure 3 in the appendix shows our annotation interface for assigning moral foundation labels to clustered frames. Each frame is shown only once, however, the annotators can also expand the different contexts for each frame in the cluster by clicking on the Context column.

Moral frames that do not fit any of the seven MFs are annotated as GENERAL-MORAL. Importantly, we use step 2 of the annotation process to **validate the frames** collected in step 1 by our two coders. The four annotators are instructed to mark frames as NON-MORAL when they think that the annotated spans do not include a moral statement. We consider a moral frame as a false positive if at least two of the four coders mark it as NON-MORAL. Tables 14 and 15 (appendix) show the number of frames identified by the two coders and the total number of frames in the combined data. For the narrative roles and the frame distribution in the manifestos, please refer to Table 16.

3.5 Inter-annotator agreement (IAA)

IAA for moral frames As it is not straightforward to compute IAA for span-based annotations, we follow common practice for opinion role labelling (Marasović and Frank, 2018) and report strict match and binary token overlap. While strict match requires that the frame spans are identical, token overlap also considers annotations as a match if at least one of the tokens in the span annotated by the two coders overlap. We first consider A1's annotations as ground truth and compute how well they agree with A2's annotations, then we switch roles and do the same for A2. The lower scores for A2–A1 compared to A1–A2 reflect the higher number of frames identified by A2 (see Tables 2 and 14). Additionally, we report *oracle* agreement for frame labels where we only consider spans that have been identified by both coders.

We see that *strict* agreement for spans is rather low (43–46%) while results for *binary overlap* is much higher with 67–72%. This shows that our annotators often agree on which text passages include moral framing while much of the disagreement concerns the concrete frame spans. When also considering frame types, agreement is in the range of 63–66% overlap. This is mostly due to mismatches in the alignment of frame spans while oracle agreement for frames identified by both coders is high with 86% (2,682 out of 3,120 frames).

	A1–A2		A2–A1	
	strict	overlap	strict	overlap
spans only	43.7	67.2	46.9	72.1
spans + frames	39.7	57.8	42.6	62.1
frames on agreed spans: 86.0% (2,682 out of 3,120)				
Hero	38.6	68.0	38.8	54.4
Victim	59.0	74.1	47.1	65.6
Villain	49.6	76.8	54.3	66.1
Beneficiary	58.4	82.6	54.2	71.0
All roles	51.8	76.6	50.3	66.6

Table 2: Percentage agreement for frame annotation for strict match and token overlap, and frame label agreement for instances where coders agreed on the span.

Next, we look at frame spans that have been labelled by both coders, to identify the main reasons for disagreement. We notice that the coders often mark the same frames, however, there are differences regarding the exact span of the annotation (e.g., whether a modifier should be part of the annotation or not). Other differences between the annotations concern the question whether a moral frame should be coded as a (im)moral *value* or an *act or goal*, (e.g., *freedom of the press*), as moral values can also be framed as goals. The interested reader can find an analysis of the disagreements in C.2 in the appendix.

Reliability of MF annotation Above we have shown that the concept of morality cannot be easily grounded on the word level. For the annotation of moral foundations, low agreement has often been noted as a key problem. For example, Hoover et al. (2020) report Kappa scores in the range of .16 to .44 (Fleiss' κ), with a κ of .27 across all coded foundations. For comparison, we observe a Fleiss κ of .54 across all MFs assigned by the four annotators.[6] To provide an additional perspective on the reliability of the annotations, we revert to previous work on Bayesian models of annotation (Dawid and Skene, 1979; Passonneau and Carpenter, 2014; Paun et al., 2018) and augment our annotations with certainty measures that assess the probability of each label for every instance.

The Dawid-Skene model provides a Maximum Likelihood Estimation of the error rates in human annotation, using the EM algorithm, and determines the most probable label for each instance,

[5] The vast majority of the moral frames (>98%) are assigned one MF only.

[6] We use Jaccard as distance metric (Masi distance: $\kappa =$ .52). Please note that Hoover et al. (2020) consider a slightly different set of 10 MFs in their study.

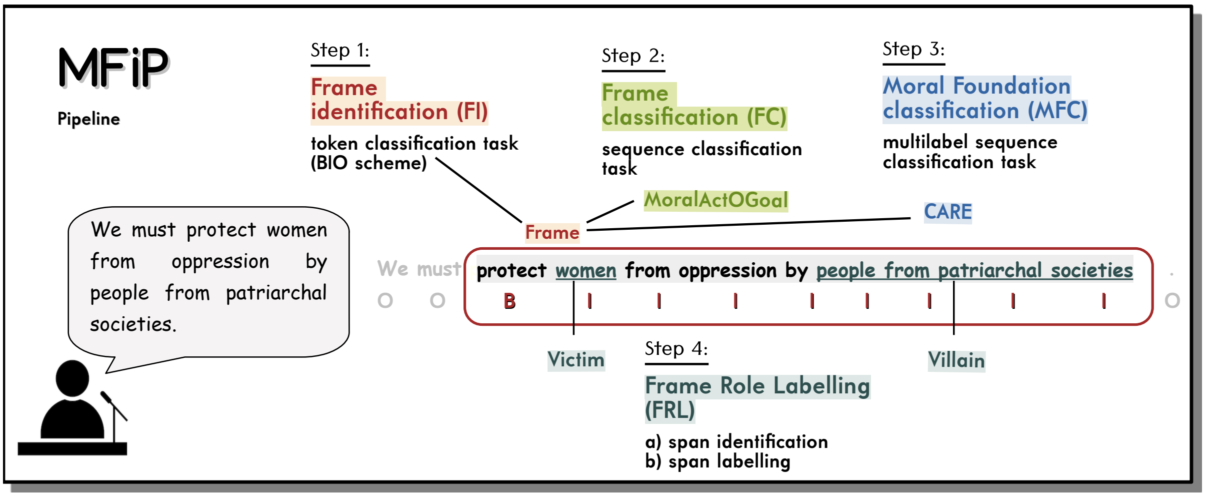

MFiP
Pipeline
Step 1:
Frame identification (FI)
token classification task (BIO scheme)
Step 2:
Frame classification (FC)
sequence classification task
Step 3:
Moral Foundation classification (MFC)
multilabel sequence classification task
We must protect women from oppression by people from patriarchal societies.
Frame
MoralActOGoal
CARE
We must protect women from oppression by people from patriarchal societies .
O O B I I I I I I I I O
Victim
Villain
Step 4:
Frame Role Labelling (FRL)
a) span identification
b) span labelling

Figure 2: Pipeline for modelling moral framing in text, including the task of moral frame identification (FI), frame classification (FC), moral foundation classification (MFC) and frame role labelling (FRL). Narrative roles can occur within or outside the frame (see Figure 1).

augmented with a confidence score. According to the model, 70% of the MF labels in our annotations have a confidence ≥96% (80% ≥ 88%; 90% ≥ 71%; 95% ≥ 62%). The model also provides us with scores that can be interpreted as reliability scores for the individual annotators, which are in the range of 69-73% for our trained coders. We take this as evidence that, despite the challenging task, our annotation scheme provides meaningful operationalisations of the concept of morality in text that can be annotated with sufficient reliability by trained human coders.

Validity of aggregated annotations We have shown that even trained and reliable coders can produce alternative codings for the same text. As we are mostly interested in making predictions about parties or speeches, we need to know how well the MF annotations on top of the moral frames from the two coders correlate at the party or speech level.

To investigate whether the annotations reflect different interpretations of the moral content, we aggregate the MF annotations per party, creating a set of annotations based on the frames identified by coder A1 and a separate set for coder A2. We then plot the number of MFs per party, normalised by speech length, and compute Pearson's correlation between the different sets (Figure 4 in the Appendix). All MFs show a strong correlation on the party level, for four of the MFs the correlation is near perfect while for the remaining two MFs that are rather rare in our corpus, the correlation is weaker (Purity: r=.93, p=.003 and Proportionality: r=.71, p=.11). When taking a more fine-grained view by looking at how well the frame annotations agree for individual speeches, we again see a strong correlation in the range of 0.75 to 0.91 with p values <.0001 (see Table 7, appendix).

This shows that even though the coders sometimes chose different text anchors to encode moral framing in the debates, the resulting moral values that we extracted, based on the individual frames identified by each coder, are **strongly correlated** both on the **party level** and on the **level of individual speeches**.

4 Experiments

The task of moral frame identification and labelling can be decomposed in the following subtasks (see Figure 2):

- `Frame Identification (FI)`: identify relevant frame spans in text
- `Frame Classification (FC)`: for each span, predict the frame type
- `Frame MF Classification (MFC)`: classify the moral foundations of a frame
- `Frame Role Labelling (FRL)`: identify and label the narrative roles for each frame (`Hero`, `Victim`, `Villain` and `Beneficiary`)

Due to space limitations, we focus on frame type and MF classification (`FC`, `MFC`) and leave `FRL` for future work. For `FI`, we use a simple BERT-based token classification model to identify moral frame spans in unlabelled data, which we then use to create additional training instances for DA and CL, as described below (see Sections 4.1 and 4.2).[7]

[7] Details and results for the `FI` baseline can be found in

Data preparation Motivated by the strong correlation between annotations (see above), we decided to include all moral frame spans coded by at least one of the annotators. This means that we can have overlapping but slightly different frame spans in the training data, which should help the model to focus on the relevant features for the task and ignore less important tokens.[8] For MFC, we consider the majority label as our target label. We decided against using the labels predicted by the Dawid-Skene model as gold standard, given that this would result in instances with low agreement being assigned to a MF with low-confidence. Instead, we prefer to label these cases as GENERAL-MORAL. We consider an MF annotation as part of the gold labels if it has been assigned by at least three of the four coders, otherwise we assign the GENERAL-MORAL label.

Negative sampling To create negative samples for training, we identify NPs and VPs in the sentences that do not contain any frame annotations. We want to roughly match the distribution in our data where ca. 90% of the moral frames are of the type 'ActOrGoal' (mostly expressed by VPs) and 10% are 'Values' (NPs). We use the Spacy[9] chunker and dependency parser to identify all noun chunks and VPs in the data and sample the category of the next negative sample from a binomial distribution with a prior of 0.1 for NPs. If the category of the next negative sample is an NP, we randomly select a noun chunk from the text, label it as 'non-moral' and add it to the training data. For VP samples, we proceed similarly, with the additional restriction that we discard VPs with more than 10 tokens, to avoid that the classifier learns instance length as a spurious feature for the class "non-moral". This results in 1,090 negative samples that we add to the training data.

4.1 Baseline: Moral Frame pipeline

Motivated by Zhang et al. (2024) who showed that a simple RoBERTa-based model outperforms larger models like Flan-T5 or Chat-GPT for the task of MF prediction, as well as by our own initial experiments, we decided on a transformer-based BERT model (Devlin et al., 2019) as our baseline. We train separate models for each task in a pipeline setting. The FI model uses a token classification setup to identify the frame spans in the text while the FC and MFC models classify the frame type and moral foundations for the predicted frames, using a sequence classification setup. We concatenate the frame span and its context and use the combined input to train the FC and MFC models. All reported results are for a 5-fold cross-validation setting, averaged over three runs with different random seeds.

4.2 DA and CL for moral frame prediction

Our extended systems explore the potential of data augmentation (DA) and contrastive learning (CL) for moral frame identification. For that, we use a large set of unlabelled debates from the German Bundestag (1949-2024). To avoid data leakage, we remove the complete 19th legislative term (2017-2021) from which we sampled our training and test data from the pool of unlabelled speeches.

We create additional training instances for FI, FC and MFC as follows. In the first step, we use our baseline FI classifier to identify moral frame spans in the unlabelled data. We consider these as frame candidates from which we sample additional training instances. To reduce noise, we filter instances based on their cosine similarities to instances in the training data, removing all instances with a similarity below a certain threshold θ.[10] We then add the predicted FC and MFC labels from the baseline models and use the filtered data for DA and CL.

DA setup For DA, we create a large, balanced dataset from the filtered pool by randomly sampling N=10,000 instances for each label class. For labels with less than N instances, we take all data available for this class. We train each model for one iteration on the data and then continue fine-tuning the models on the manually annotated MFiP. Please note that, in order to avoid data leakage, we do not include speeches from the same legislative term as the MFiP training/test sets in the pool.

CL setup To improve the input representations for the classification tasks, we use CL to train sentence embeddings that encode differences between moral frames on a large set of instances from the filtered pool. Through contrastive learning, the model learns to position representations of the same class closer together in the embedding space while representations for data points that belong to different classes are pushed further apart.

Section D in the appendix.

[8] We expect that this might have a similar regularizing effect as DA techniques like cutoff (Shen et al., 2020).

[9] https://spacy.io/, model: de-core-news-sm.

[10] We set θ to 0.9 and compute cosine similarity, using the langchain_huggingface library with HuggingFaceEmbeddings and the paraphrase-multilingual-MiniLM-L12-v2 embeddings model (Reimers and Gurevych, 2020).

We harvest the labels needed for creating the training samples, based on the predictions of our baseline classifiers, and create training instances as follows. To balance the data, we select a maximum number of N=2,000 instances to be included for each class. For each frame type or MF, we extract the list of all frames that have been predicted to belong to this class as positive samples and the list of all instances that do not belong to this class as negative samples. Then we shuffle both lists and extract training triples of "anchor", "positive" and "negative" frames where the ith instance of the positive list is the anchor, instance $i+1$ of the same list is a positive example and the ith instance of the list of negatives is a negative example. For classes with less than N instances, we set N to the number of instances for this class.

We then use the triplet objective function (eq.1) to train sentence embeddings (Reimers and Gurevych, 2019) that are sensitive to the different MF classes. d is a distance metric (here: Euclidean distance) and margin is set to 1.

$$L = max(d(s_{anchor} - s_{pos}) - d(s_{anchor} - s_{neg}) \\ +margin, 0) \tag{1}$$

We apply CL to both tasks, FC and MFC, and use the resulting embeddings to initialise our classifier.[11] Then we fine-tune the models on the MFiP corpus (for more details, see Appendix D).

4.3 Results

Frame type classification (FC) Table 3 shows results for frame type classification (FC). Overall, the results are quite high, with an average F1 of 82.0 across all classes. An exception is the ImmoralValue class for which only few training instances are available (see Table 15 in the appendix). There are no improvements for DA and only a minor increase in results for the CL setting, mostly for the detection of the NoFrame label. This might be due to the fact that for FC, we already have a large number of training instances with sufficient examples for all but the IMMORALVALUE class. This is reflected in the high baseline where results for the individual frame types are already in the range of 77%–86%, making it harder to achieve further improvements.

Moral foundation classification (MFC) For moral foundation classification, the task we are most interested in, results are mixed (Table 4). While we obtain high to moderate results for the more frequent MFs, F1 for the least frequent labels (PROPORTIONALITY, PURITY) is 0. However, we see substantial improvements of nearly 4 percentage points for DA and another increase for our morality aware sentence embeddings (CL, +6.4%).

Label	Prec	Rec	F1 micro
		FC	
MoralActOrGoal	82.9± 2.1	79.7± 3.0	81.2± 0.5
ImmoralActOrGoal	85.9± 1.3	86.5± 1.2	86.2± 0.3
MoralValue	77.9± 5.3	78.6± 5.1	78.0± 1.2
ImmoralValue	0.0± 0.0	0.0± 0.0	0.0± 0.0
PoliticalActOrGoal	75.1± 1.2	79.8± 2.4	77.4± 0.9
NoFrame	83.2± 4.6	85.1± 4.4	84.0± 0.8
Total	**82.0**± 0.2	**82.0**± 0.2	**82.0**± 0.2
		FC + DA	
MoralActOrGoal	80.7± 3.1	81.4± 3.2	80.9± 0.7
ImmoralActOrGoal	84.2± 1.0	89.0± 1.5	86.5± 0.9
MoralValue	77.1± 4.3	76.1± 6.3	76.4± 2.0
ImmoralValue	0.0± 0.0	0.0± 0.0	0.0± 0.0
PoliticalActOrGoal	77.7± 5.5	77.7± 4.2	77.5± 1.3
NoFrame	89.1± 1.5	80.1± 2.4	84.3± 1.2
Total	**82.0**± 0.6	**82.0**± 0.6	**82.0**± 0.6
		FC + CL	
MoralActOrGoal	82.0± 2.9	82.4± 4.6	82.1± 0.9
ImmoralActOrGoal	85.9± 4.6	87.8± 3.2	86.7± 1.0
MoralValue	79.0± 4.9	78.0± 6.0	78.4± 4.1
ImmoralValue	0.0± 0.0	0.0± 0.0	0.0± 0.0
PoliticalActOrGoal	77.9± 3.3	78.9± 2.6	78.4± 0.9
NoFrame	88.1± 0.5	84.6± 2.0	86.3± 1.1
Total	**82.9**± 0.7	**82.9**± 0.7	**82.9**± 0.7

Table 3: Avg. precision, recall and F1 (micro) for frame type classification (FC) (± shows standard deviations across the 3 runs).

Interestingly, only DA helps to improve results for the low-frequency classes (PROPORTIONALITY: +27%, PURITY: +46%) while both the CL approach and the baseline are unable to learn these classes. This is probably due to the larger number of instances used for DA. Further increasing the training size for the CL setting, however, resulted in overfitting and did not yield any improvements.

Combining CL with DA also failed to improve results in terms of micro F1 but increases macro-F1 by another 1.6% as it manages to identify at least some of the low-frequency class instances.[12]

Out-of-domain results on manifestos Finally, we evaluate our models on the held-out manifestos data. While also from the political domain, the language of electoral manifestos is quite differ-

[11] To initialise our sentence embeddings, we use the German pretrained deepset/gbert-large model.

[12] We also experimented with using frame type labels as additional features for MFC. While the gold frame types resulted in substantial improvements, using predicted frame types did not increase results for MFC.

Label	Prec	Rec	F1	
			micro	macro
MFC				
Care	75.4± 1.0	67.9± 1.3	71.4± 0.3	
Equality	78.3± 1.2	51.7± 5.0	62.2± 3.3	
Proport.	0.0± 0.0	0.0± 0.0	0.0± 0.0	
Loyalty	66.4± 3.7	39.2± 1.6	49.2± 2.0	
Authority	89.7± 3.5	12.7± 5.5	22.0± 8.2	
Purity	0.0± 0.0	0.0± 0.0	0.0± 0.0	
Liberty	76.8± 1.6	54.9± 5.4	63.9± 3.0	
General	71.6± 0.9	65.4± 1.6	68.3± 0.5	
None	84.8± 2.3	85.3± 2.5	85.0± 0.4	
Total	**77.4**± 0.4	**67.1**± 1.3	**71.9**± 0.6	**46.9**± 1.3
MFC + DA				
Care	76.5± 3.4	73.5± 5.5	74.8± 1.3	
Equality	78.1± 5.7	67.6± 3.6	72.3± 1.4	
Proport.	53.2± 1.3	18.2± 1.3	27.1± 1.3	
Loyalty	70.5± 1.8	60.5± 5.9	64.9± 2.6	
Authority	69.3± 2.2	54.6± 7.2	60.9± 5.2	
Purity	45.0± 7.3	47.6± 7.4	46.3± 7.3	
Liberty	78.1± 2.5	70.3± 0.9	74.0± 1.6	
General	75.6± 0.4	69.1± 1.6	72.2± 0.9	
None	87.4± 1.7	82.1± 1.0	84.7± 0.5	
Total	**79.1**± 0.5	**72.5**± 1.3	**75.7**± 0.7	**64.1**± 1.7
MFC + CL				
Care	78.9± 1.4	76.2± 2.1	77.5± 0.5	
Equality	84.4± 0.4	65.6± 1.5	73.8± 0.8	
Proport.	0.0± 0.0	0.0± 0.0	0.0± 0.0	
Loyalty	78.0± 2.5	64.7± 1.8	70.7± 0.2	
Authority	75.8± 1.0	46.1± 3.7	57.2± 3.1	
Purity	0.0± 0.0	0.0± 0.0	0.0± 0.0	
Liberty	82.7± 1.3	69.4± 2.9	75.5± 1.5	
General	78.0± 0.2	73.4± 0.6	75.6± 0.3	
None	88.0± 0.7	85.8± 0.4	86.9± 0.3	
Total	**82.0**± 0.6	**75.0**± 0.6	**78.3**± 0.0	**57.5**± 0.5
MFC + DA + CL				
Care	75.4± 4.3	77.5± 4.4	76.3± 0.4	
Equality	76.2± 1.9	71.9± 1.2	74.0± 1.3	
Proport.	53.8± 5.4	20.8± 7.9	29.0± 8.2	
Loyalty	69.9± 2.0	68.0± 8.9	68.6± 3.9	
Authority	66.4± 1.9	60.7± 3.0	63.4± 2.0	
Purity	43.5± 8.7	46.4± 9.4	44.8± 8.8	
Liberty	78.1± 3.3	72.6± 4.9	75.1± 1.3	
General	79.2± 1.4	68.2± 2.2	73.3± 0.8	
None	86.6± 1.2	86.1± 2.1	86.3± 0.5	
Total	**79.6**± 0.4	**75.0**± 0.5	**77.3**± 0.4	**65.7**± 0.9

Table 4: Avg. precision, recall and F1 (mirco/macro) for MF classification over 3 runs.

ent from that of political speeches. While the debate speeches contain roughly the same number of moral and immoral frames, we observe around 4 times more references to moral acts or goals in the manifestos. It is therefore interesting to see how well our classifiers cope with the new text type and how well our data augmention and CL methods will help to adapt to the new genre.

Table 5 shows results for FC and MFC on the manifestos data (for detailed results for individual labels, see Table 12 in the appendix). We note that the FC classifier adapts well to the new domain, despite the differences in distribution, and our best results (DA) are only 2% lower than for the speeches.

Label	Prec	Rec	F1	
			micro	macro
FC	77.8± 0.9	77.8± 0.9	77.8± 0.9	61.1
FC+**DA**	80.0± 0.8	80.0± 0.8	80.0± 0.8	62.9
FC+**CL**	78.2± 3.0	78.2± 3.0	78.2± 3.0	60.3
MFC	72.9± 1.9	55.4± 3.2	62.9± 1.3	45.3
MFC+**DA**	70.1± 1.5	60.8± 2.4	65.1± 0.7	54.1
MFC+**CL**	70.4± 0.1	61.0± 0.3	65.4± 0.2	49.4
MFC+**DA+CL**	68.8± 1.7	62.0± 2.7	65.2± 2.1	53.7

Table 5: Avg. precision, recall and F1 for frame type classification (FC) and MFC on the held-out manifestos testset (± shows standard deviations across the 3 runs).

For MFC, however, we see a substantial decrease in results of around 10 percentage points (micro F1), giving evidence that moral framing in debates is different from the one in the manifestos. However, DA and CL are both able to mitigate the effect at least slightly and, as before, DA in particular helps to increase macro F1. Our results on the out-of-domain manifestos are in line with previous work showing a similar degradation of results for a classifier applied to tweets from different topical domains (Liscio et al., 2022). This points to future directions of research, showing the need for developing domain adaptation methods for moral framing.

5 Conclusions

We presented the MFiP corpus (MoralFramingInPolitics), a new resource for moral framing. The MFiP is, to our best knowledge, the first German benchmark for the prediction of moral values in text. The data has been manually annotated and includes moral frame types, narrative roles and moral foundations, offering new possibilities for modelling morality in political communication. We discussed the challenges of annotating morality in text and showed that our new schema results in reliable operationalisations of moral framing. Then we explored the potential of DA and CL for the automatic prediction of moral frames and values. Our results showed that both methods yield substantial and complementary improvements for MFC. While morality aware sentence embeddings trained with CL help to improve results for most but the rare classes, DA is especially suited to increase scores for classes with only a few instances.

In future work, we plan to extend our models with narrative roles, develop better models for moral frame identification (FI) and focus on domain adaptation for moral framing.

6 Limitations

Our resource includes German data only and can not be applied to other languages. We would like to argue that the contextual setting is even more important than the linguistic restriction. While the application of multilingual pretrained models allows us to train multilingual systems for the prediction of moral framing, the results might not be optimal, as the political issues and procedures are specific to the German political landscape and parliamentary system and might not adapt well to other countries with different political systems.

In addition, the training data was sampled from a recent legislative term. While we took great care to include a wide variety of topics, it is not yet clear how well the trained models will perform on less recent text, for example, debates from the first legislative terms of the German Bundestag, starting in 1949. We therefore advise researchers who want to apply our models for diachronic studies to add a further validation step, testing how well the trained models perform on the historical texts.

Another potential limitation is the number of coders for moral frame identification, which, due to limited funding, has been done by two coders only. This has been addressed during the annotation of moral foundations, where all instances in the data have been annotated by four coders who were also instructed to highlight incorrectly identified frames. However, this can only address precision (but not recall, as the four coders only see the frames that have been identified but not the ones that might have been missed).

The experiments presented in the paper report results on our new benchmark for German for the two tasks of frame type classification (FC) and moral foundations classification (MFC). Applying our models to large, unlabelled data for real-world analyses would require additional validation including the moral frame identification step (FI), which was out of scope for this work.

Acknowledgments

The work presented in this paper is funded by the German Research Foundation (DFG) under the UNCOVER project (PO1900/7-1 and RE3536/3-1). We would like to thank our annotators, Lilly Brauner and Florian Ertz, for their dedicated work. We would also like to thank the anonymous reviewers for their constructive feedback.

References

Hassan Alhuzali and Sophia Ananiadou. 2021. SpanEmo: Casting multi-label emotion classification as span-prediction. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1573–1584, Online. Association for Computational Linguistics.

Oscar Araque, Lorenzo Gatti, and Kyriaki Kalimeri. 2020. Moralstrength: Exploiting a moral lexicon and embedding similarity for moral foundations prediction. *Knowledge-Based Systems*, 191:105184.

Oscar Araque, Lorenzo Gatti, and Kyriaki Kalimeri. 2021. The language of liberty: A preliminary study. In *Companion Proceedings of the Web Conference 2021*, WWW '21, page 623–626, New York, NY, USA. Association for Computing Machinery.

M. Atari, J. Haidt, J. Graham, S. Koleva, S. T. Stevens, and M. Dehghani. 2023. Morality beyond the weird: How the nomological network of morality varies across cultures. *Journal of Personality and Social Psychology*, 5(125):1157–1188.

Maria Becker, Ekkehard Felder, and Marcus Müller. 2024. Moralisierung als sprachliche praxis. In Ekkehard Felder, Friederike Nüssel, and Jale Tosun, editors, *Moral und Moralisierung: Neue Zugänge*, pages 123–151. Berlin, Boston: De Gruyter.

Mariano Beiró, Jacopo D'Ignazi, Victoria Bustos, María Prado, and Kyriaki Kalimeri. 2023. Moral narratives around the vaccination debate on facebook. In *Proceedings of the ACM Web Conference 2023*, WWW'23), pages 4134–4141.

Shaun Bevan. 2019. Gone Fishing: The Creation of the Comparative Agendas Project Master Codebook. In Frank R. Baumgartner, Christian Breunig, and Emiliano Grossman, editors, *Comparative Policy Agendas: Theory, Tools, Data*. Oxford: Oxford University Press.

Tobias Burst, Werner Krause, Pola Lehmann, Jirka Lewandowski, Theres Matthieß, Nicolas Merz, Sven Regel, and Lisa Zehnter. 2022. Manifesto corpus. version: 2022-1.

Alexander Philip Dawid and Allan M. Skene. 1979. Maximum likelihood estimation of observer error-rates using the em algorithm. *Applied Statistics*, 1(28):20–28.

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.

Robert M. Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4).

Lea Frermann, Jiatong Li, Shima Khanehzar, and Gosia Mikolajczak. 2023. Conflicts, villains, resolutions: Towards models of narrative media framing. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8712–8732, Toronto, Canada. Association for Computational Linguistics.

Dean Fulgoni, Jordan Carpenter, Lyle Ungar, and Daniel Preoţiuc-Pietro. 2016. An empirical exploration of moral foundations theory in partisan news sources. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3730–3736, Portorož, Slovenia. European Language Resources Association (ELRA).

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. SimCSE: Simple contrastive learning of sentence embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6894–6910, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Jesse Graham, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, Sean P. Wojcik, and Peter H. Ditto. 2012. Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. *Advances in Experimental Social Psychology*, pages 55–130.

Jesse Graham, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, Sean P. Wojcik, and Peter H. Ditto. 2013. Chapter two - moral foundations theory: The pragmatic validity of moral pluralism. In Patricia Devine and Ashby Plant, editors, *Advances in Experimental Social Psychology*, volume 47, pages 55–130. Academic Press.

Jesse Graham, Jonathan Haidt, and Brian A Nosek. 2009. Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96(5):1029–1046.

Jesse Graham, Brian A Nosek, Jonathan Haidt, Ravi Iyer, Spassena Koleva, and Peter H Ditto. 2011. Mapping the moral domain. *Journal of Personality and Social Psychology*, 101(2):366–385.

Candida M. Greco, Lorenzo Zangari, Davide Picca, and Andrea Tagarelli. 2024. E2mocase: A dataset for emotional, event and moral observations in news articles on high-impact legal cases. *Preprint*, arXiv:2409.09001.

Jonathan Haidt, Jesse Graham, and Conrad Joseph. 2009. Above and below left–right: Ideological narratives and moral foundations. *Psychological Inquiry*, 20(2-3):110–119.

Joe Hoover, Gwenyth Portillo-Wightman, Leigh Yeh, Shreya Havaldar, Aida Mostafazadeh Davani, Ying Lin, Brendan Kennedy, Mohammad Atari, Zahra Kamel, Madelyn Mendlen, Gabriela Moreno, Christina Park, Tingyee E. Chang, Jenna Chin, Christian Leong, Jun Yen Leung, Arineh Mirinjian, and Morteza Dehghani. 2020. Moral foundations twitter corpus: A collection of 35k tweets annotated for moral sentiment. *Social Psychological and Personality Science*, 11(8):1057–1071.

Minghao Hu, Yuxing Peng, Zhen Huang, Dongsheng Li, and Yiwei Lv. 2019. Open-domain targeted sentiment analysis via span-based extraction and classification. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 537–546, Florence, Italy. Association for Computational Linguistics.

Clara Husson and Nicola Palma. 2024. Broadening the study of morality in multiparty settings through a novel dictionary translation and validation methodology. *Political Psychology*.

R. Iyer, S. Koleva, J. Graham, P. Ditto, and J. Haidt. 2012. Understanding libertarian morality: The psychological dispositions of self-identified libertarians. *PLoS ONE*, 8(7).

Kristen Johnson and Dan Goldwasser. 2018. Classification of moral foundations in microblog political discourse. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 720–730, Melbourne, Australia. Association for Computational Linguistics.

Jae-Hee Jung. 2020. The mobilizing effect of parties' moral rhetoric. *American Journal of Political Science*, 64(2):341–355.

Jan-Christoph Klie, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. 2018. The inception platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9. Association for Computational Linguistics. Event Title: The 27th International Conference on Computational Linguistics (COLING 2018).

Jonathan Kobbe, Ines Rehbein, Ioana Hulpuș, and Heiner Stuckenschmidt. 2020. Exploring morality in argumentation. In *Proceedings of the 7th Workshop on Argument Mining*, pages 30–40, Online. Association for Computational Linguistics.

Patrik W Kraft. 2018. Measuring morality in political attitude expression. *Journal of Politics*, 3(80):1028–33.

Keena Lipsitz. 2018. Playing with emotions: The effect of moral appeals in elite rhetoric. *Political Behavior*, 40:57–78.

Enrico Liscio, Alin Dondera, Andrei Geadau, Catholijn Jonker, and Pradeep Murukannaiah. 2022. Cross-domain classification of moral values. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 2727–2745, Seattle, United States. Association for Computational Linguistics.

Ana Marasović and Anette Frank. 2018. SRL4ORL: Improving opinion role labeling using multi-task learning with semantic role labeling. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 583–594, New Orleans, Louisiana. Association for Computational Linguistics.

Jeff McMahan. 2000. Moral intuition. In Hugh LaFollette -, editor, *The Blackwell Guide to Ethical Theory*, pages 92–110. Blackwell.

Yulia Otmakhova, Shima Khanehzar, and Lea Frermann. 2024. Media framing: A typology and survey of computational approaches across disciplines. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15407–15428, Bangkok, Thailand. Association for Computational Linguistics.

Jeongwoo Park, Enrico Liscio, and Pradeep Murukannaiah. 2024. Morality is non-binary: Building a pluralist moral sentence embedding space using contrastive learning. In *Findings of the Association for Computational Linguistics: EACL 2024*, pages 654–673, St. Julian's, Malta. Association for Computational Linguistics.

Rebecca J. Passonneau and Bob Carpenter. 2014. The Benefits of a Model of Annotation. *Transactions of the Association for Computational Linguistics*, 2:311–326.

Silviu Paun, Bob Carpenter, Jon Chamberlain, Dirk Hovy, Udo Kruschwitz, and Massimo Poesio. 2018. Comparing Bayesian Models of Annotation. *Transactions of the Association for Computational Linguistics*, 6:571–585.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3980–3990, Hong Kong, China. Association for Computational Linguistics.

Nils Reimers and Iryna Gurevych. 2020. Making monolingual sentence embeddings multilingual using knowledge distillation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4512–4525, Online. Association for Computational Linguistics.

Ines Reinig, Maria Becker, Ines Rehbein, and Simone Ponzetto. 2024. A survey on modelling morality for text analysis. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 4136–4155, Bangkok, Thailand. Association for Computational Linguistics.

Shamik Roy and Dan Goldwasser. 2021. Analysis of nuanced stances and sentiment towards entities of US politicians through the lens of moral foundation theory. In *Proceedings of the Ninth International Workshop on Natural Language Processing for Social Media*, pages 1–13, Online. Association for Computational Linguistics.

Shamik Roy, Maria Leonor Pacheco, and Dan Goldwasser. 2021a. Identifying morality frames in political tweets using relational learning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9939–9958, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Shamik Roy, Maria Leonor Pacheco, and Dan Goldwasser. 2021b. Identifying Morality Frames in Political Tweets using Relational Learning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9939–9958, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

E. Shanahan, M. Jones, M. Mcbeth, and C. Radaelli. 2017. The narrative policy framework. In C.M. Weible and P.A. Sabatier, editors, *The Theories of the Policy Process*, pages 173–213. Boulder, CO: Westview Press.

Dinghan Shen, Mingzhi Zheng, Yelong Shen, Yanru Qu, and Weizhu Chen. 2020. A simple but tough-to-beat data augmentation approach for natural language understanding and generation. *arXiv preprint arXiv:2009.13818*.

Walter Sinnott-Armstrong, Liane Young, and Fiery Cushman. 2010. 246Moral Intuitions. In *The Moral Psychology Handbook*. Oxford University Press.

Tessa Stanier and Hagyeong Shin. 2024. Polarization and morality: Lexical analysis of abortion discourse on reddit. *Preprint*, arXiv:2407.00455.

Jackson Trager, Alireza S. Ziabari, Aida Mostafazadeh Davani, Preni Golazizian, Farzan Karimi-Malekabadi, Ali Omrani, Zhihe Li, Brendan Kennedy, Nils Karl Reimer, Melissa Reyes, Kelsey Cheng, Mellow Wei, Christina Merrifield, Arta Khosravi, Evans Alvarez, and Morteza Dehghani. 2022. The moral foundations reddit corpus. *Preprint*, arXiv:2208.05545.

Karina Vida, Judith Simon, and Anne Lauscher. 2023. Values, ethics, morals? on the use of moral concepts in NLP research. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 5534–5554, Singapore. Association for Computational Linguistics.

Maxwell A. Weinzierl and Sanda M. Harabagiu. 2022. From hesitancy framings to vaccine hesitancy profiles: A journey of stance, ontological commitments and moral foundations. *Proceedings of the International AAAI Conference on Web and Social Media*, 16(1):1087–1097.

Winston Wu, Lu Wang, and Rada Mihalcea. 2023. Cross-cultural analysis of human values, morals, and biases in folk tales. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 5113–5125, Singapore. Association for Computational Linguistics.

Lorenzo Zangari, Candida M. Greco, Davide Picca, and Andrea Tagarelli. 2025. ME2-BERT: Are events and emotions what you need for moral foundation prediction? In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 9516–9532, Abu Dhabi, UAE. Association for Computational Linguistics.

Xinliang Frederick Zhang, Winston Wu, Nicholas Beauchamp, and Lu Wang. 2024. MOKA: Moral knowledge augmentation for moral event extraction. 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 4481–4502, Mexico City, Mexico. Association for Computational Linguistics.

A Moral Foundations Theory (MFT)

Below we provide a short description of the moral foundations, adapted from the MFT website.[13]

Care: This foundation is related to our long evolution as mammals with attachment systems and an ability to feel (and dislike) the pain of others. It underlies the virtues of kindness, gentleness, and nurturance.

Fairness: This foundation is related to the evolutionary process of reciprocal altruism. It underlies the virtues of justice and rights.

In 2023, Atari et al. (2023) was split into two new foundations, Equality and Proportionality, as it was found that politically left-leaning individuals more strongly endorse values of Equality while more conservative individuals prefer the notion of proportionality.

Equality: Equality is defined as “Intuitions about equal treatment and equal outcome for individuals.”

Proportionality: Proportionality is defined as “Intuitions about individuals getting rewarded in proportion to their merit or contribution.”

Loyalty: This foundation is related to our long history as tribal creatures able to form shifting coalitions. It is active anytime people feel that it’s “one for all and all for one.” It underlies the virtues of patriotism and self-sacrifice for the group.

Authority: This foundation was shaped by our long primate history of hierarchical social interactions. It underlies virtues of leadership and followership, including deference to prestigious authority figures and respect for traditions.

Purity: This foundation was shaped by the psychology of disgust and contamination. It underlies notions of striving to live in an elevated, less carnal, more noble, and more “natural” way (often present in religious narratives). This foundation underlies the widespread idea that the body is a temple that can be desecrated by immoral activities and contaminants (an idea not unique to religious traditions). It underlies the virtues of self-discipline, self-improvement, naturalness, and spirituality.

Liberty: This foundation is about the feelings of reactance and resentment people feel toward those who dominate them and restrict their liberty. Its intuitions are often in tension with those of the authority foundation. The hatred of bullies and dominators motivates people to come together, in solidarity, to oppose or take down the oppressor.

The last foundation is not yet considered as part of the moral foundations but often discussed as a plausible candidate (Iyer et al., 2012). LIBERTY is often used to frame political arguments, we therefore include it in our annotations.

B Data and sampling

The transcripts used in our dataset are freely available from `https://www.bundestag.de/services/opendata`. The `MoralFramingInPolitics` corpus is made available under the Open Data Commons Attribution License: `http://opendatacommons.org/licenses/by/1.0/`.

B.1 Bundestag debates corpus

The `MFiP` corpus includes speeches sampled from the 19th legislative term (2017–2021) of the German Bundestag. The distribution of topics in the data is not representative of the larger data but has been sampled to cover a more diverse range of topics, with contributions from all parties distributed over the whole legislative term. Below, we describe the sampling procedure in more detail.

B.2 Sampling procedure

We extracted a dataset of parliamentary debates from the German Bundestag, covering a time

[13] `https://moralfoundations.org/`.

No.	Major topic
1	Cultural Policy Issues
2	Defense
3	Domestic Macroeconomic Issues
4	Education
5	Environment
6	Health
7	Immigration and Refugee Issues
8	Law, Crime, and Family Issues

Table 6: Major topics from the Comparative Agendas Project that we sampled to be included in our data set.

period from the 19th legislative term (2017 to 2021).[14] The corpus includes speeches by 807 different speakers, with over 900,000 sentences and over 16 mio tokens. From this corpus, we selected individual speeches for annotation as follows. Our goal was to create a gold standard, controlled for topic and including speeches for each of the political parties. In addition, we wanted the texts to be evenly distributed over the time span of the legislative term (2017–2021). To achieve this goal, we selected specific agenda items that covered a range of topics, and then sampled all speeches that belong to this specific agenda item, to increase the comparability of the contributions made by speakers from different parties.

CAP topics We based our topic selection on the coding scheme developed in the Comparative Agendas Project (CAP) (Bevan, 2019). The coding scheme includes 21 major topics (see Table 6) and more than 200 fine-grained subtopics. The topics we selected have been annotated as major CAP topics, which allowed us to use the annotated CAP data to train a topic classifier (a transformer-based text classifier). For the 21 major topics, our classifier achieves a micro F1 of 72.9% on the indomain interpellation data.

Sampling based on predicted CAP topics We then used the classifier to predict topics for each speech in the parliamentary debates, after applying the same preprocessing steps to the data. This gives us topic predictions for each individual speech. To guide our sampling process, we aggregated the predictions for all speeches belonging to the same agenda item. We call the topic based on a "majority vote" for each agenda item the *major topic* of the agenda. Our assumption is that all speeches given on the same agenda item should belong to the same major topic. As a result, we obtained a distribution of topics over all speeches for each respective agenda item. We sorted the predictions and *manually selected and validated* agenda items for each of the CAP topics in Table 6, where the majority of the speeches for this agenda item have been predicted as belonging to this topic.

We only selected agenda items where each of the political parties participated in the debate, and also aimed at selecting items that are roughly evenly distributed over the time period of the legislative term, to *ensure that our data set is as representative as possible, covering a range of different topics, distributed over the whole legislative term and including speeches from all different parties on the same set of topics*.

B.3 Manifestos subcorpus

The manifestos in the `MFiP` corpus have been extracted from the Manifestos Project Database (Burst et al., 2022). We downloaded the manifestos for the German election of the Bundestag in 2021 for all parties that were part of the Bundestag at the time. See below for the list of included parties.

- Alternative für Deutschland (AfD; Alternative for Germany)
- Bündnis 90/Die Grünen (Green party)
- Christlich-Demokratische Union/Christlich-Soziale Union in Bayern (CDU/CSU; Christian Democratic Union/Christian Social Union in Bavaria)
- Freie Demokratische Partei (FDP; Liberal Democratic Party)
- Die Linke (The Left)
- Sozialdemokratische Partei Deutschlands (SPD; Social Democratic Party Germany)

B.4 Clustering frames into moral themes

We applied the fast clustering algorithm[15] provided in the S-BERT library (Reimers and Gurevych, 2019). Specifically, we use the `German_Semantic_STS_V2` model[16] and extract clusters with a minimum community size of $\{25, 25, 15, 5\}$ and a threshold of

[14] The data is freely available from https://www.bundestag.de/services/opendata, the Open Data service of the German Bundestag.

[15] See the documentation at https://sbert.net/examples/sentence_transformer/applications/clustering.

[16] For documentation, see https://huggingface.co/aari1995/German_Semantic_STS_V2.

MF	coefficient	p-value
Care	r=0.884	p<.0001
Equality	r=0.914	p<.0001
Proportionality	r=0.710	p<.0001
Loyalty	r=0.831	p<.0001
Authority	r=0.801	p<.0001
Purity	r=0.897	p<.0001
Liberty	r=0.872	p<.0001
General-Moral	r=0.751	p<.0001

Table 7: Pearson's correlation for the Moral Foundations annotated on top of the frames identified by coder1 and coder2. The strong correlation shows that even though the span agreement for individual frame spans is not high (67-77% overlap) as the coders sometimes grounded the annotation of moral values on different text spans, the aggregated Moral Foundations for each speech strongly correlate.

$\{0.7, 0.7, 0.7, 0.6\}$ for $\{$*MoralActOrGoal, ImmoralActOrGoal, MoralValue, ImmoralValue*$\}$, respectively. We also experimented with other settings but found that the ones above gave us a good balance between cluster coherence and coverage.

Not all frames could be assigned to a cluster in the first clustering round. We therefore ran a second round of clustering where we subsequently decreased the threshold until nearly every frame had been assigned to a cluster. The remaining frames that could not be clustered were considered as their own group.

C Annotation process and validation

C.1 Annotation interface for MF annotation

Figure 3 shows our annotation interface for assigning moral foundation labels to clustered frames. The displayed cluster mostly includes MORALVALUE frames related to values of freedom and self-determination. We include English translations for the original German frames in the figure.

C.2 Analysis of disagreements

Frame spans The most frequent causes for mismatches regarding the frame spans included modifiers and coordination. While the guidelines instructed the coders to focus on the arguments and exclude modification, we found that annotators sometimes deviated from this rule when they felt that excluding the modifier did not accurately capture the meaning of the frame (Ex. C.1). Other mismatches include prepositional modifier phrases and relative clauses.

Ex. C.1. (further (promote dialog between religions, world views and cultures)$_{A1}$)$_{A2}$

Party	# speeches	# tokens	# speakers
CDU	71	68,249	56
SPD	55	46,681	43
AfD	37	27,970	30
FDP	31	21,121	23
LEFT	26	18,672	21
GREENS	24	16,701	17
non-inscrit	4	2,111	2
Total	248	201,505	192

Table 8: Distribution of speeches/speakers in the `MFiP`.

Regarding coordination, we find that sometimes one annotator includes the whole coordinate phrase as one frame while the other split it up into several frames (Ex. C.2).

Ex. C.2. ((decent training)$_{A1}$, working conditions and pay)$_{A2}$

Frame labels We notice that the largest part of the disagreements concerning the frame labels is due to one annotator chosing to annotate the frame as a MORALVALUE while the second coder annotated an overlapping span as an act or goal. For example, the frame *protect freedom* has been annotated as a MORALACTORGOAL by coder 1 while coder 2 chose to only mark *freedom* as a MORALVALUE.

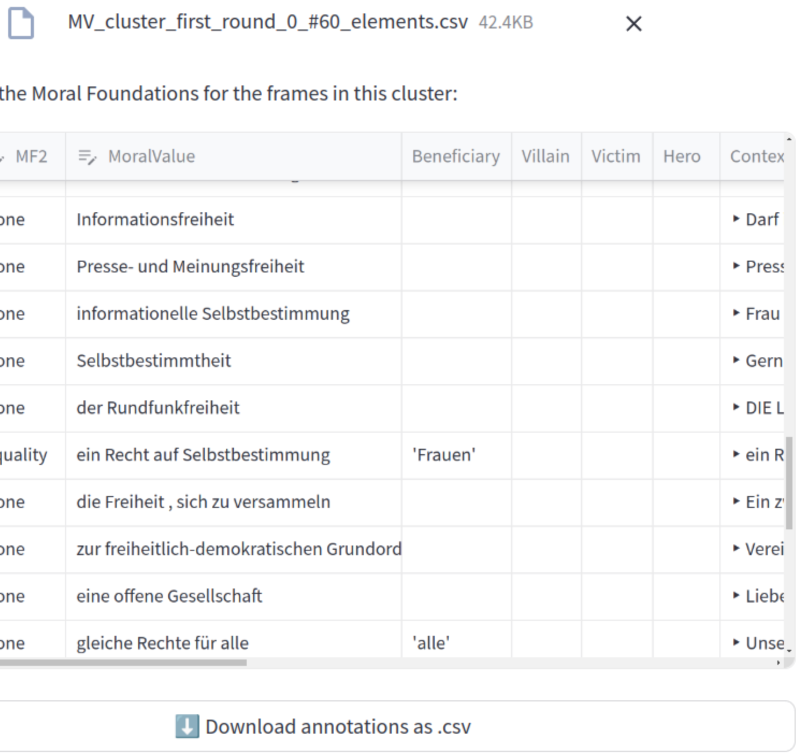
We also found instances that have been identified by one coder only while the other coder did not consider this instance as a case of moral framing. These included strong evaluative statements that, however, did not include strong moral rhetoric.

To our surprise, we also encountered cases labelled as *moral* by one coder while the other coder annotated the same instance as *immoral*. An example is shown in Ex. C.3 below.

Ex. C.3. (Strict punishment for (false statements in the asylum procedure)$_{A2}$)$_{A1}$

This frame expresses a political demand by the far-right party AfD which coder1 chose to annotate as a moral goal while coder2 took a different, but equally valid perspective and only annotated the subspan "false statements in the asylum procedure", framed as an immoral act by the speaker.

This illustrates some of the challenges for the annotation of morality in text and shows that different and overlapping moral annotations with opposing polarity can exist at the same time. This, however, does not so much reflect different moral beliefs or biases held by the coders but rather shows that morality is a compositional construct that cannot be captured by assigning labels to sentences or docu-



MV_cluster_first_round_0_#60_elements.csv 42.4KB

Please annotate the Moral Foundations for the frames in this cluster:

MF	MF2	MoralValue	Beneficiary	Villain	Victim	Hero	Contex
Liberty	None	Informationsfreiheit					▸ Darf
Liberty	None	Presse- und Meinungsfreiheit					▸ Press
Liberty	None	informationelle Selbstbestimmung					▸ Frau
Liberty	None	Selbstbestimmtheit					▸ Gern
Liberty	None	der Rundfunkfreiheit					▸ DIE L
Liberty	Equality	ein Recht auf Selbstbestimmung	'Frauen'				▸ ein R
Liberty	None	die Freiheit , sich zu versammeln					▸ Ein z
Liberty	None	zur freiheitlich-demokratischen Grundord					▸ Verei
Liberty	None	eine offene Gesellschaft					▸ Liebe
Equality	None	gleiche Rechte für alle	'alle'				▸ Unse

German frame text	English translation
Informationsfreiheit	Freedom of information
Presse- und Meinungsfreiheit	Freedom of the press and expression
informationelle Selbstbestimmung	informational self-determination
der Rundfunkfreiheit	freedom of broadcasting
ein Recht auf Selbstbestimmung	the right to self-determination
die Freiheit, sich zu versammeln	the freedom to assemble
zur freiheitlich-demokratischen Grundordnung	to the free and democratic basic order
eine offene Gesellschaft	an open society
gleiche Rechte für alle	equal rights for everybody

Figure 3: Annotation interface for the annotation of Moral Foundations (MF) on clustered frames. The MoralValue column shows the clustered frames, the next four columns show the annotated roles. The last column (Context) shows the context(s) for each frame and can be expanded by clicking on it. The English translations are shown in the Table above.

ments. Instead, annotations need to be grounded on the frame level to be transparent and interpretable.

C.3 MFiP: corpus statistics

Below we show some statistics for our new corpus. Table 8 shows the number of speeches and token counts per party. The number of moral frame types per party is shown in Table 15. Please note that the number of moral frames in the MFiP is higher than the number of frames identified by each coder as we combined the annotated frames (removing duplicates). The number of frames individually annotated by each coder is shown in Table 14.

Table 9 illustrates the total number of combined frames and the number of moral foundations in our corpus, and Table 10 details the number of narrative roles for each party in the MFiP.

FC	freq.	MF	freq.
ImmoralValue	45	Purity	28
MoralValue	475	Proportionality	77
PoliticalActOrGoal	1,314	Authority	152
ImmoralActOrGoal	2,262	Loyalty	435
MoralActOrGoal	2,386	Equality	449
NoFrame	1,090	Liberty	480
		Care	1,205
		General-Moral	2,457
		None	2,404
Total	**7,572**		**7,687**

Table 9: Distribution of moral foundations in the MFiP corpus (note that each frame can have more than one MF; PoliticalActOrGoal frames have MF "None").

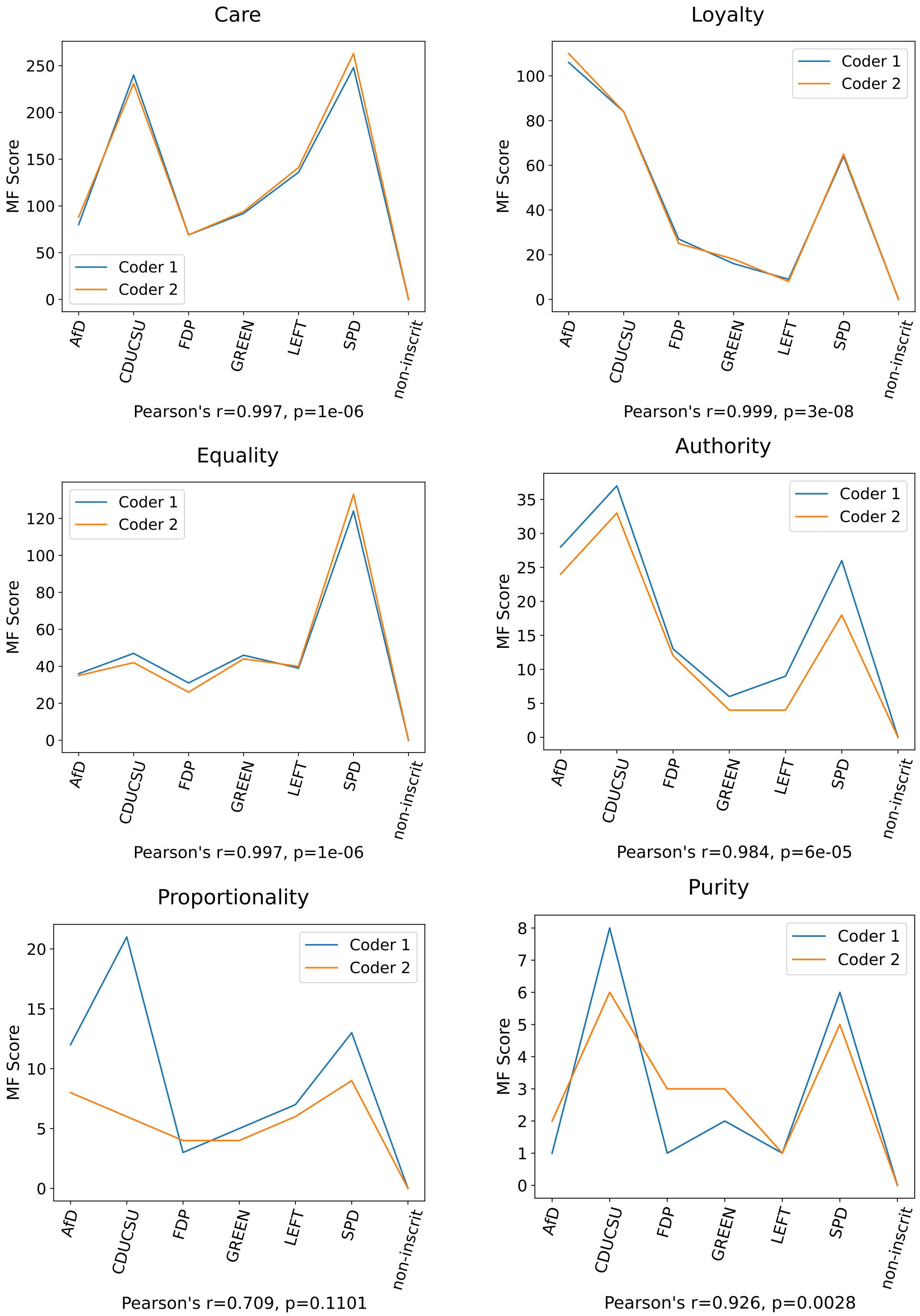


Figure 4: Moral Foundations (MFs) per political party and Pearson's correlation for the MFs resulting from frames identified by the different coders. All MFs show a very strong positive correlation (r≥.93) with the exception of Proportionality (r=.65 with p≥.1).

A1	Hero	Victim	Villain	Benef.	Total
AfD	36	96	224	59	415
CDU/CSU	149	60	158	158	525
FDP	16	30	61	45	152
GREEN	12	29	89	50	180
LEFT	22	58	142	44	266
SPD	86	86	95	148	415
non-inscrit	3	0	11	6	20
Total	324	359	780	510	1,973
A2					
AfD	39	97	210	81	427
CDU/CSU	127	79	120	148	474
FDP	16	40	49	51	156
GREENS	16	40	74	41	171
LEFT	25	85	148	60	318
SPD	89	109	102	160	460
non-inscrit	5	4	11	5	25
Total	317	454	714	546	2,031

Table 10: Distribution of narrative roles per party in the German parliamentary debates identified by each coder.

D Model details

Moral frame identification (FI) Our span identification system is a BERT-based transformer[17] trained in a token classification setup. We use the BIO schema to encode the frame spans (see Ex.13). Our model uses the AdamW optimiser and a learning rate of 2.693154582157772e-05, and is trained for 8 epochs. For FI, we remove the negative samples during training while for FC and MFC, we include all instances in training.

Overall results on the token level show a weighted avg. F1 of around 85%. This, however, is misleading as it reflects the high accuracy for non-frame tokens (label: O) with around 93% while the results for the B and I classes are much lower, with roughly 57% (B) and 67% (I). This shows that the token classification architecture might not be suitable for this task. We also experimented with span extraction models that predict the start and end position of the span and jointly learn the span boundaries and labels (Hu et al., 2019; Alhuzali and Ananiadou, 2021) but failed to improve results. This is probably due to the fact that our spans are much longer and less lexically grounded than the ones in the cited work.

Frame type classification (FC) Our FC model is a BERT-based text classifier, initialised with the pretrained deepset/gbert-large model. We use the AdamW optimiser, set the learning rate to 2.693154582157772e-05 and weight decay to 0, use a batch size of 16 and a maximum input length of 512. We train the model for 10 epochs and save the best model (the model with the lowest validation loss).

Moral Foundation classification (MFC) Our MFC model is a BERT-based text classifier, initialised with the pretrained deepset/gbert-large model. We use the AdamW optimiser, set the learning rate to 2.693154582157772e-05 and weight decay to 0.001. For Moral Foundation classification, we use a batch size of 8 and a maximum input length of 256. Learning rate is set to 3e-6. We train the model for 15 epochs and save the best model (the model with the lowest validation loss).

D.1 Validation on the manifestos dataset

Table 11 shows detailed results per class for FC on the out-of-domain manifestos data and Table 12 shows results for each moral foundation for MFC.

Label	**Prec**	**Rec**	**F1**
	FC		
MoralActOrGoal	85.3± 2.7	80.0± 4.2	82.5± 1.2
ImmoralActOrGoal	76.5± 3.4	87.6± 3.4	81.6± 0.6
MoralValue	83.5± 5.4	67.4± 10.0	74.0± 4.4
ImmoralValue	0.0± 0.0	0.0± 0.0	0.0± 0.0
PoliticalActOrGoal	61.2± 5.7	77.1± 5.7	67.9± 1.8
NoFrame	62.6± 21.5	67.1± 20.4	60.9± 5.5
Total	**77.8**± 0.9	**77.8**± 0.9	**77.8**± 0.9
	FC + DA		
MoralActOrGoal	83.4± 2.2	85.3± 4.2	84.3± 1.0
ImmoralActOrGoal	78.9± 6.3	88.0± 3.6	83.0± 1.9
MoralValue	81.2± 1.9	71.5± 1.8	76.0± 0.6
ImmoralValue	0.0± 0.0	0.0± 0.0	0.0± 0.0
PoliticalActOrGoal	68.8± 3.9	71.1± 8.1	69.6± 1.7
NoFrame	74.8± 4.4	57.6± 12.3	64.4± 7.5
Total	**80.0**± 0.8	**80.0**± 0.8	**80.0**± 0.8
	FC + CL		
MoralActOrGoal	80.1± 5.7	87.7± 2.8	83.6± 2.3
ImmoralActOrGoal	83.2± 5.5	80.5± 5.2	81.6± 2.1
MoralValue	86.5± 6.2	56.4± 24.3	65.5± 17.3
ImmoralValue	0.0± 0.0	0.0± 0.0	0.0± 0.0
PoliticalActOrGoal	65.7± 7.1	71.3± 5.4	68.0± 2.2
NoFrame	69.5± 8.6	60.0± 11.9	63.3± 3.9
Total	**78.2**± 3.0	**78.2**± 3.0	**78.2**± 3.0

Table 11: Avg. precision, recall and F1 (micro) for frame type classification (FC) on held-out manifestos testset (± shows standard deviations across the 3 runs).

[17] We use the deepset/gbert-large model, available from https://huggingface.co/deepset/gbert-large.

Label	Prec	Rec	F1
MFC			
Care	78.8± 4.7	68.0± 5.0	72.8± 1.0
Equality	74.4± 5.3	63.3± 6.1	68.1± 2.5
Proportionality	0.0± 0.0	0.0± 0.0	0.0± 0.0
Loyalty	57.6± 3.6	33.3± 11.2	41.3± 8.2
Authority	90.6± 10.7	21.7± 3.3	34.7± 3.5
Purity	0.0± 0.0	0.0± 0.0	0.0± 0.0
Liberty	81.9± 3.0	54.3± 5.3	65.1± 3.0
General-Moral	70.6± 2.4	48.5± 3.2	57.5± 2.3
None	69.0± 4.9	68.8± 7.1	68.6± 1.4
Total	72.9± 1.9	55.4± 3.2	62.9± 1.3
MFC + DA			
Care	79.4± 7.7	63.0± 10.3	69.5± 3.3
Equality	64.2± 5.2	79.3± 7.7	70.6± 0.2
Proportionality	68.3± 2.7	29.4± 5.9	40.9± 5.7
Loyalty	51.0± 7.3	53.5± 7.1	51.6± 1.1
Authority	59.1± 5.5	56.7± 11.7	57.1± 5.7
Purity	0.0± 0.0	0.0± 0.0	0.0± 0.0
Liberty	79.9± 3.9	57.1± 3.2	66.6± 3.2
General-Moral	73.4± 4.5	51.3± 8.0	59.9± 3.7
None	72.3± 5.0	69.7± 9.8	70.5± 2.8
Total	70.1± 1.5	60.8± 2.4	65.1± 0.7
MFC + CL			
Care	75.0± 1.0	72.3± 1.3	73.6± 0.2
Equality	69.1± 0.8	73.7± 1.7	71.3± 0.4
Proportionality	0.0± 0.0	0.0± 0.0	0.0± 0.0
Loyalty	50.6± 1.0	50.0± 2.0	50.3± 1.5
Authority	60.0± 0.8	43.3± 2.9	50.3± 2.2
Purity	0.0± 0.0	0.0± 0.0	0.0± 0.0
Liberty	75.0± 0.4	66.8± 1.2	70.7± 0.5
General-Moral	76.7± 0.1	44.1± 0.5	56.0± 0.4
None	66.1± 0.9	80.2± 0.4	72.5± 0.4
Total	70.4± 0.1	61.0± 0.3	65.4± 0.2
MFC + DA + CL			
Care	75.4± 8.1	67.3± 16.9	69.6± 7.3
Equality	59.9± 1.7	84.1± 2.7	69.9± 0.4
Proportionality	65.4± 10.3	31.4± 6.8	41.9± 6.7
Loyalty	51.6± 13.9	51.9± 9.0	50.1± 2.5
Authority	65.7± 8.1	43.3± 9.3	51.3± 5.1
Purity	0.0± 0.0	0.0± 0.0	0.0± 0.0
Liberty	83.2± 8.9	59.8± 3.5	69.2± 1.5
General-Moral	76.2± 1.8	46.8± 3.7	57.9± 3.1
None	67.9± 1.5	79.4± 2.8	73.1± 0.7
Total	68.8± 1.7	62.0± 2.7	65.2± 2.1

Table 12: Precision, recall and F1 (micro) for MF classification on the held-out manifestos dataset (avg. over 3 runs).

With	a	coal phase-out	continue	we	the	deindustrialisation	Germany’s	`particle`	.
Mit	einem	Kohleausstieg	setzen	wir	die	Deindustrialisierung	Deutschlands	fort	.
O	O	O	B	I	I	I	I	I	O

“By phasing out coal, we are continuing the deindustrialisation of Germany.”

Table 13: Example input for segmentation, using the BIO schema.

A1			**Moral**	**Moral**	**Immoral**	**Immoral**	**Political**
Party	**# Tokens**	**# Frames**	**Value**	**Act**	**Value**	**Act**	**Act**
AfD	71,275	740	30	215	2	411	82
CDU/CSU	161,447	1,289	58	617	2	357	255
FDP	47,565	475	26	222	0	151	76
GREEN	40,956	406	21	154	2	171	58
LEFT	46,308	557	24	179	0	297	57
SPD	119,938	1,052	66	537	2	272	175
non-inscr	6,562	50	1	23	0	19	7
Total (A1)	494,051	4,569	226	1,947	8	1,678	710

A2			**Moral**	**Moral**	**Immoral**	**Immoral**	**Political**
Party	**# Tokens**	**# Frames**	**Value**	**Act**	**Value**	**Act**	**Act**
AfD	71,275	701	62	164	7	353	115
CDU/CSU	161,447	1,171	132	407	5	300	327
FDP	47,565	441	38	161	2	122	118
GREEN	40,956	393	50	99	10	141	93
LEFT	46,308	513	27	127	2	277	80
SPD	119,938	1,064	117	392	15	286	254
non-inscr	6,562	41	3	10	2	17	9
Total (A2)	494,051	4,324	429	1,360	43	1,496	996

Table 14: Number of moral frames and frame types identified by each coder and distribution across parties. Table 15 shows the merged set of frames after removing duplicates and validating the frame annotations (see Section 3.4).

			Moral	**Moral**	**Immoral**	**Immoral**	**Political**	**No**
Party	**# Tokens**	**# Frames**	**Value**	**Act**	**Value**	**Act**	**Act**	**Frame**
AfD	71,275	1,193	68	272	8	541	157	147
CDU	161,447	2,218	145	747	6	467	446	407
FDP	47,565	799	46	282	1	204	153	113
GRUENE	40,956	653	53	180	10	222	115	73
LINKE	46,308	868	36	220	2	396	107	107
SPD	119,938	1,771	123	664	16	403	324	241
fraktionslos	6,562	70	4	21	2	29	12	2
Total	494,051	7,572	475	2,386	45	2,262	1,314	1,090

Table 15: Distribution of frames and frame types per party in the parliamentary debates train/dev/test data. NoFrames are negative instances that have been created through negative sampling (see Section 4).

Party	**# Tokens**	**# Frames**	**MoralValue**		**MoralAct**		**ImmoralAct**		**PoliticalAct**	
			#	%	#	%	#	%	#	%
Migration – Coder 1										
AfD	2093	103	2	0.96	45	21.50	38	18.16	18	8.60
CDU/CSU	872	47	3	3.44	29	33.26	9	10.32	6	6.88
FDP	1503	79	10	6.65	53	35.26	4	2.66	12	7.98
GRUENE	1815	102	5	2.75	67	36.91	27	14.88	3	1.65
LINKE	2368	178	8	3.38	125	52.79	27	11.40	18	7.60
SPD	679	49	4	5.89	34	50.07	4	5.89	7	10.31
Migration – Coder 2										
AfD	2093	99	10	4.78	43	20.54	33	15.77	13	6.21
CDU/CSU	872	41	5	5.73	22	25.23	8	9.17	6	6.88
FDP	1503	55	17	11.31	25	16.63	6	3.99	7	4.66
GRUENE	1815	111	6	3.31	62	34.16	27	14.88	16	8.82
LINKE	2368	179	12	5.07	102	43.07	38	16.05	27	11.40
SPD	679	49	6	8.84	33	48.69	3	4.42	7	10.31
Media – Coder 1										
AfD	344	31	4	11.63	6	17.44	20	58.14	1	2.91
CDU/CSU	337	14	1	2.97	9	26.71	4	11.87	0	0.00
FDP	496	38	3	6.05	23	46.37	5	10.08	7	14.11
GRUENE	223	15	2	8.97	12	53.81	0	0.00	1	4.48
LINKE	774	52	6	7.75	38	49.10	4	5.17	4	5.17
SPD	381	22	4	10.50	11	28.87	6	15.75	1	2.62
Media – Coder 2										
AfD	344	25	6	17.44	4	11.63	14	40.70	1	2.91
CDU/CSU	337	20	8	23.74	7	20.77	5	14.84	0	0.00
FDP	496	35	7	14.11	17	34.27	4	8.06	7	14.11
GRUENE	223	19	7	31.39	11	49.33	0	0.00	1	4.48
LINKE	774	65	23	29.72	34	43.93	5	6.46	3	3.88
SPD	381	27	11	28.87	12	31.50	3	7.87	1	2.62
Culture – Coder 1										
AfD	679	37	4	5.89	17	25.04	14	20.62	2	2.95
CDU/CSU	592	38	9	15.20	21	35.47	0	0.00	8	13.51
FDP	921	51	2	2.17	30	32.57	6	6.51	13	14.12
GRUENE	1288	87	4	3.11	72	55.90	2	1.55	9	6.99
LINKE	1719	95	6	3.49	72	41.88	9	5.24	8	4.65
SPD	797	46	8	10.04	29	36.39	5	6.27	4	5.02
Culture– Coder 2										
AfD	679	45	19	27.98	10	14.73	15	22.09	1	1.47
CDU/CSU	592	45	15	25.34	22	37.16	3	5.07	5	8.45
FDP	921	59	9	9.77	36	39.09	4	4.34	10	10.86
GRUENE	1288	89	22	17.08	63	48.91	0	0.00	4	3.11
LINKE	1719	119	19	11.05	78	45.38	9	5.24	13	7.56
SPD	797	62	15	18.82	37	46.42	5	6.27	5	6.27
Total/avg.										

Table 16: Distribution of moral frames in the German manifestos on the topics Migration, Media and Culture.