Journal of the Brazilian Computer Society, 2023, 30:1, doi: 10.5753/jbcs.2024.3932

A benchmark for Portuguese zero-shot stance detection

Matheus Camasmie Pavan [**University of São Paulo** | ***matheus.pavan@usp.br***]
Ivandré Paraboni [**University of São Paulo** | ***ivandre@usp.br***]

Escola de Artes, Ciências e Humanidades (EACH), Universidade de São Paulo (USP), Av. Arlindo Bettio 1000, São Paulo, SP, 03828-000, Brazil.

Received: 13 December 2023 • **Accepted:** 12 July 2024 • **Published:** 15 October 2024

Abstract Stance detection is the task of inferring for/against attitudes towards a particular target from text. As targets are in principle unlimited, however, research in the field has moved from so-called in-domain classification (which assume the availability of a sufficient number of stances towards the intended target for training purposes) to more realistic zero-shot scenarios. However, regardless of which - or how much - training data is taken into account, most existing zero-shot approaches are devoted to the English language, in stark opposition to alternatives devoted to Portuguese. As a means to overcome some of these difficulties, this article presents a benchmark (hereby understood as the combination of a dataset, baseline systems and their results) for zero-shot Portuguese stance detection that is, to the best of our knowledge, the first of it kind. More specifically, we adapt a number of existing models available for the English language to Portuguese, and introduce novel approaches to the task based on more recent prompt engineering methods and off-task labelling, achieving SOTA results that are, in some cases, even superior to in-domain classification.

Keywords: Natural Language Processing, Stance detection, Zero-shot

1 Introduction

In Natural Language Processing (NLP), stance detection [Kucuk and Can, 2020; Aldayel and Magdy, 2021; Alturayeif *et al*., 2023; Pavan *et al*., 2023] is the computational task of inferring for or against attitudes towards a particular target (e.g., a company, a piece of legislation, etc.) from text data. For instance, '*This company should not be allowed to be on the stock market*' conveys a stance against a particular target company. As an NLP field, stance detection has grown considerably since the seminal *SemEval-2016* stance detection shared task in Mohammad *et al*. [2016], with potential applications including fake news [Oshikawa *et al*., 2020] or hate speech detection [Schmidt and Wiegand, 2017; da Silva *et al*., 2020], among others.

Computational stance detection shares similarities with sentiment analysis but, as pointed out in Kucuk and Can [2020], sentiment and stance do not generally correlate. For instance, a stance against the target 'Christmas' may express a positive sentiment, as in '*I just love all these people spending a fortune on pretty gifts*'. Moreover, as a stance towards a given target may be expressed in a wide range of linguistic formulations, stance detection is arguably more semantics-oriented than sentiment analysis and, accordingly, the task has often been addressed in a fully supervised fashion with the aid of purpose-built labelled datasets [Taulé *et al*., 2017; Hosseinia *et al*., 2020; Jaziriyan *et al*., 2021].

The availability of labelled data conveying a sufficient number of stances towards the intended target enables the development of *in-domain* stance classifiers (e.g., [Zarrella and Marsh, 2016; Kochkina *et al*., 2017; Yang *et al*., 2019; Cignarella *et al*., 2020; Flores *et al*., 2021]), that is, models that are trained and tested on a set of labelled examples of stance towards the target of interest. Models of this kind may arguably achieve optimal results for the task as long as suitable (e.g., compatible) labelled train and test datasets are available but, as the number of stance targets is in principle unlimited, building a new labelled dataset for every target of interest renders general stance detection unattainable.

Given these difficulties, the focus of the field in recent years has gradually shifted from in-domain stance detection to more practical *zero-shot* settings, that is, stance detection models that are to be evaluated on target topics unseen during training [Allaway and McKeown, 2022]. The latter, in their most common implementation, often amounts to cross-target stance detection [Xu *et al*., 2018; Zhang *et al*., 2020], that is, models that detect stance towards a particular target based on labelled examples of stance towards other, unseen targets [Allaway and McKeown, 2020; Allaway *et al*., 2021; Liu *et al*., 2021]. Thus, for instance, we may in principle train a model to detect stance towards Donald Trump from a collection of labelled stances towards , e.g., Hillary Clinton [Mohammad *et al*., 2016]. Given the lack of suitable (i.e., fully compatible) training data, however, cross-target results are usually below those observed in (arguably optimal) in-domain settings, and many studies of this kind do not even provide a comparison with in-domain alternatives, perhaps under the assumption that in-domain results could not be surpassed if simply using less or no training data.

In addition to standard cross-target stance detection, recent advances in large language models (LLMs) have provided yet another fresh research perspective for the field. LLMs may be prompted to assign stance labels to an input text with no training data at all and, more importantly, have been shown to outperform previous cross-target methods [Zhang *et al*., 2023]. However, regardless of which – or how much – training data is taken into account, we notice that most existing zero-shot approaches are devoted to the English language (e.g., [Liu *et al*., 2021; Liang *et al*., 2022; Chunling *et al*., 2023]). Our target language - Portuguese - by contrast, still lacks considerably behind, and adaptation may be time-consuming or even unfeasible. This may be the case, for instance, when attempting to adapt models that resort to external knowledge provided by language-specific resources (e.g., English ConceptNet [Speer *et al*., 2017]) to mitigate the lack of training data.

Based on these observations, this article intends to contribute to the field of Portuguese NLP by presenting a benchmark - hereby understood as the combination of (i) a labelled corpus, (ii) baseline systems and (iii) reference results - for zero-shot Portuguese stance detection that is, to the best of our knowledge, the first of it kind. Using a large social media stance corpus that has been previously assessed mainly for in-domain classification [Pereira *et al*., 2023], we adapt a number of existing models available for the English language to Portuguese, and introduce novel approaches based on more recent prompt engineering methods and off-task labelling, in some cases outperforming even in-domain classification results. Put together, these efforts are intended to foster further investigation in the field, addressing not only the issue of how different zero-shot methods for Portuguese compare to each other, but also presenting alternatives that may indicate future directions of research in the field.

The main contributions made in this work are as follows.

- Standard zero-shot stance detection methods from Portuguese text, based on cross-target classification.
- LLM-based methods that require no labelled training data at all, which achieve SOTA results for the task and outperform even in-domain classification.
- A novel method based on off-task social media polarisation data, which outperforms standard zero-shot alternatives with minimal computational costs if compared to LLM-based methods.

The rest of this paper is structured as follows. Section 2 reviews a number of recent studies in zero-shot stance detection. Section 3 describes the corpus to be used as train and test data for our models. Section 4 describes the adaptation of existing stance detection models for Portuguese, and novel approaches based on off-task labelling and prompt engineering. Section 5 presents our results, and Section 6 draws a number of final remarks and suggestions of future work.

2 Related work

Stance detection is the computational task of estimating whether a piece of text t expressing information about a target object x conveys a stance either in favour or against x, in some cases considering also a third (e.g., 'not applicable') alternative. When a set of labelled examples of stance towards x is available, stance detection models may be built in standard in-domain fashion, that is, using some of the available labelled data for training. This corresponds to the early approach to stance detection found in Mohammad *et al*. [2016] and others. In more recent zero-shot stance detection, by contrast, existing methods may be distinguished from each other according to the precise nature of the data used for training purposes, if any.

Table 1 summarises a number of studies of this kind alongside details regarding the training strategies taken into account, and underlying computational methods. To this end, the following training strategies are considered.

- Out-of-domain (ood) labelling: the model is trained from a set of labelled stances towards a target other than the intended test target. This strategy corresponds to the standard cross-target approach found in much of the existing work in the field [Allaway and McKeown, 2020; Liu *et al*., 2021; Zhao *et al*., 2022; Wen and Hauptmann, 2023].
- Distant (dist) labelling: the model is trained from a set of weakly labelled stances and additional (non-stance) information available from external sources.
- No (none) labelling: the model does not use any training data.

From Table 1, we notice that existing work in zero-shot stance detection has been largely based on two prominent datasets: the SemEval-2016 stance corpus [Mohammad *et al*., 2016], and the more recent VAST corpus [Allaway and McKeown, 2020], both of which devoted to the English language.

The SemEval-2016 corpus is a collection of 4870 manually labelled tweets covering six targets, and it was mainly intended as a resource for standard in-domain stance classification. Subsequently, the corpus has been applied also to limited out-of-domain stance detection in cross-target fashion (e.g., using known stances towards the feminist movement as training data to predict stance towards abort legislation, etc.) as in Xu *et al*. [2018].

VAST Allaway and McKeown [2020], on the other hand, has been specifically designed for zero- and few-shot stance detection. The corpus is a collection of manually labelled comments posted on the New York Times website about a wide range of issues, comprising 23,525 stances towards 5,634 targets. The comparatively large number of targets implies that many (if not most) are actually under represented, and motivates the use of so-called generalised topic (or target) representations to overcome the lack of data in zero-shot scenarios. However, we notice that the target/instance ratio in VAST also makes the corpus most likely unsuitable for in-domain stance detection, and in fact none of the zero-shot studies under discussion have actually reported VAST results in in-domain fashion.

Regarding the training strategy under consideration, nearly all studies make use of out-of-domain (ood) labelled data, that is, models are built from a set of labelled stances towards unseen targets. This setting, sometimes called cross-target stance classification [Zhao *et al*., 2022], is by far the most common zero-shot strategy among the selected studies, the only exceptions being the use of distant supervision [Xu *et al*., 2022] and prompt-based learning [Zhang *et al*., 2023] as discussed below.

Table 1. Related work in zero-shot stance detection and training strategies.

Ref.	Corpus	Training	Method
Allaway and McKeown [2020]	VAST	ood	topic-grouped network
Allaway *et al*. [2021]	SemEval	ood	topic-adversarial network
Liu *et al*. [2021]	VAST	ood	graph convolution network
Luo *et al*. [2022]	VAST	ood	common sense knowledge graph encoding
Liang *et al*. [2022]	SemEval, VAST, others	ood	graph contrastive learning
Xu *et al*. [2022]	SemEval, VAST, others	dist	text entailment, GPT-3, BERT
Pavan and Paraboni [2022]	UstanceBR-r1	ood	BERT ADDA
Zhao *et al*. [2022]	SemEval, VAST, others	ood	BERT contrastive learning
Chunling *et al*. [2023]	SemEval, others	ood	graph+BERT ADDA
Zhang *et al*. [2023]	SemEval, others	none	ChatGPT
Wen and Hauptmann [2023]	VAST	ood	Conditional generation

Central to any zero-shot approach is the question of how to mitigate the lack of training samples. To this effect, existing studies resort to a wide range of computational methods, with a certain prevalence of graph representations (e.g., to convey external knowledge such as common sense) [Liu *et al*., 2021; Liang *et al*., 2022; Chunling *et al*., 2023], and adversarial discriminative domain adaptation (ADDA) to learn from out-of-domain data [Pavan and Paraboni, 2022; Zhao *et al*., 2022; Chunling *et al*., 2023]. Individual details regarding each of the selected studies are discussed as follows.

In addition to introducing the VAST corpus, which has become a standard reference for zero-shot stance detection in the English language, the work in Allaway and McKeown [2020] presents a novel model called TGA Net. The model consists of a contextual conditional encoding layer followed by topic- (or target-) grouped attention using generalised topic representations, and a feed-forward neural network. This architecture implicitly captures relationships between targets, and it is particularly suitable for scenarios in which a large number of targets is available as in the case of the VAST dataset.

Following the initial results of TGA Net, the work in Allaway *et al*. [2021] presents a new model for zero-shot stance detection on Twitter that makes use of adversarial learning to generalise across targets. The model, called TOpic-ADversarial Network (TOAD), consists of the domain-transfer architecture from Zhang *et al*. [2017] coupled with the stance model in Augenstein *et al*. [2016] and additional topic- (or target-) specific attention layer intended to compute topic-invariant representations that generalise to other (i.e., unseen) targets.

The work in Liu *et al*. [2021] introduces CKE-NET, a model enriched with commonsense knowledge that is intended to exploit relational knowledge at both structural and semantic level. The model uses BERT [Devlin *et al*., 2019] to encode documents and targets. As a knowledge graph base, CKE-NET uses ConceptNet [Speer *et al*., 2017], a semantic network for the English language that conveying millions of instances of 34 types of relations. Commonsense relational knowledge is represented as triples $R = (u; r; v)$, where u is the head concept, r is the relation, and v is the tail concept, and it is embedded in the model with the aid of the Graph Convolution Network CompGCN [Vashishth *et al*., 2020].

As in the case of CKE-NET [Liu *et al*., 2021], the work in Luo *et al*. [2022] proposes to improve knowledge transfer with the aid of commonsense knowledge, and sentiment information. The proposed model, called BS-RGCN, once again uses ConceptNet [Speer *et al*., 2017] as the knowledge graph base, and RGCN encoding [Schlichtkrull *et al*., 2018] to compute concept latent feature representations. The model also includes sentiment information computed by performing BERT sentiment masking.

The work in Liang *et al*. [2022] proposes a framework for zero-shot stance detection called JointCL (joint contrastive learning), which comprises four main components: (a) stance contrastive learning based on the supervised signal of stance labels; (b) prototypes generation, which uses a clustering method to derive prototypes of the training data; (c) prototypical graph contrastive learning between known and unseen targets; and (d) stance classification.

The work in Xu *et al*. [2022] argues that existing systems for zero-shot stance detection are optimised on a particular dataset from a single domain, and hence do not perform well on other datasets, and that evaluation is usually based on a limited number of unseen targets, in which some of these are assumed to be richly annotated. Based on the observation that real-world applications are unlikely to meet these conditions, the issue of stance detection is investigated in an open-world scenario with neither domain constraints nor target-specific annotations. The work combines indirect supervision provided by textual entailment datasets and weak supervision obtained from pre-trained language models.

The work in Pavan and Paraboni [2022] introduced a preliminary version of the UstanceBR corpus [Pereira *et al*., 2023] to be taken as the basis of the present work, and presents cross-domain stance classification method for the Portuguese language based on out-of-domain labelled data, which is the standard problem definition adopted by most of the zero-shot studies under discussion. The model combines an existing domain adaptation method based on BERT with adversarial learning and knowledge distillation that has been shown to be successful in the related tasks of cross-domain sentiment analysis [Ryu and Lee, 2022] and cross-domain author profiling [Delmondes Neto and Paraboni, 2021].

The work in Zhao *et al*. [2022] attempts to mitigate the lack of knowledge in zero-shot settings by making use of a data augmentation method. More specifically, input instances are augmented with masked target words, and fed to an unsupervised contrastive learning module to capture features that may transfer across targets. In order to fit a given target, raw texts are encoded as target-specific features, and enhanced features for predicting previously unseen targets are computed with the aid of an attention mechanism combining both syntactic and target information.

The work in Chunling *et al*. [2023] introduces ANEK, an adversarial network enhanced with external knowledge. ANEK is analogous to previous CKE-NET [Liu *et al*., 2021] and BS-RGCN [Luo *et al*., 2022] in that it also utilises ConceptNet [Speer *et al*., 2017] as its auxiliary knowledge base, alongside sentiment knowledge. The adversarial learning component is based on pre-trained models intended to acquire transferable knowledge from the source targets, which is expected to generalise over unseen targets.

Unlike previous work in the field, the work in Zhang *et al*. [2023] presents a more extreme instance of zero-shot stance detection in which no training data is required at all. Instead, the model prompts the ChatGPT conversation agent to label stances directly, that is, without any previous examples. Results based on the SemEval-2016 stance corpus [Mohammad *et al*., 2016], and p-stance [Li *et al*., 2021] are shown to outperform a number of early stance detection alternatives such as Bicond [Augenstein *et al*., 2016] and BERT baseline systems. The work also points out the explanation abilities of this approach, and suggests that prompt-based methods may represent the SOTA for this particular task. Unlike most zero-shot studies based on VAST (which is unsuitable for in-domain classification, as discussed above), this is the only approach that actually compares - and indeed outperforms - in-domain stance classification based on the SemEval-2016 stance corpus.

Finally, the work in Wen and Hauptmann [2023] implements zero-shot stance detection as a conditional generation framework, and formulates the task as denoising from partially-filled templates. In this approach, a template contains two sentences in the form 'The target is <target>. The stance is $<$ *stance*$>$'. The target placeholder is filled in by the application, and the stance placeholder is to be filled in by the stance detection method as a response. The model uses BART [Lewis *et al*., 2020], an encoder-decoder language model pretrained with denoising objectives. The response (i.e., the value of the stance placeholder) is generated by the decoder component, which is trained by minimising the log-likelihood over the whole generated sequence. The final stance label response is computed with additional post-processing that attempts to determine the polarity word from the generated output.

3 Data

The present work is based on the *UstanceBR* r2 corpus [Pereira *et al*., 2023], a collection of 46.8 thousand tweets in the Portuguese language that have been manually labelled with binary (for/against) stance information towards six targets divided into three politically-charged pairs (Brazilian presidents, Covid-related measures, and local institutions.) The corpus has been previously assessed in standard in-domain settings, and will be presently considered in a zero-shot scenario to be discussed in Section 4.

Descriptive statistics of the data used in the present work are summarised in Table 2. Further details about the UstanceBR r2 corpus, including information on how the data have been collected, annotated and organised, are described in Pereira *et al*. [2023].

Table 2. Train and test data descriptive statistics.

	Train		Test	
	Against	For	Against	For
Lula	4,514	3,806	100	100
Bolsonaro	5,565	3,849	100	100
Hydrox.	3,978	4,017	100	100
Sinovac	4,058	3,915	100	100
Globo TV	3,341	2,672	100	100
Church	3,539	3,598	100	100

The training data to be used in the present work corresponds to the pre-defined train portion of the original corpus, although only used by our supervised approaches (i.e., using in-domain or out-of-domain training data) as discussed in the next sections. The test data, kept the same across all models, comprises a random selection (100 instances per class) of the actual test portion of the corpus. This simplification was motivated by the need to reduce the costs involved in prompting paid LLM services that are required by some of the approaches under discussion.

4 Stance detection models

As a means to provide reference results for Portuguese zero-shot stance detection based on the *UstanceBR* r2 corpus described in the previous section, in what follows we will consider a range of novel and existing approaches to the task. These include a number of influential models that follow the definition of zero-shot that became standard in the field - that is, training the model from a set of labelled stances towards unseen targets only - and more knowledge-poor alternatives that require only off-task training data, or which rely upon more recent prompt engineering methods. In doing so, we would like to illustrate (i) how the zero-shot strategies compare to in-domain classification, that is, the optimal scenario in which a corpus of labelled stances towards the intended target happens to be available for use as training data, and (ii) which zero-shot strategy may be regarded as a benchmark for the present dataset.

To investigate question (i) we will compare results obtained by different zero-shot strategies to in-domain classification, which is presently intended to represent the possible upper limit for the task. The in-domain approach is described in Section 4.1.

To investigate question (ii), we will consider a number of zero-shot strategies that adapt existing work devoted to the English language to the Portuguese setting provided by the *UstanceBR* r2 corpus, and which differ from each other in the nature of the training method and data (if any) to be taken into account. More specifically, our zero-shot models are divided into three categories: first, as in the more traditional zero-shot definition, Section 4.2 describes a number of influential models that are trained from a set of labelled stances towards unseen targets, that is, using out-of-domain training data. Next, section 4.3 relaxes this requirement by

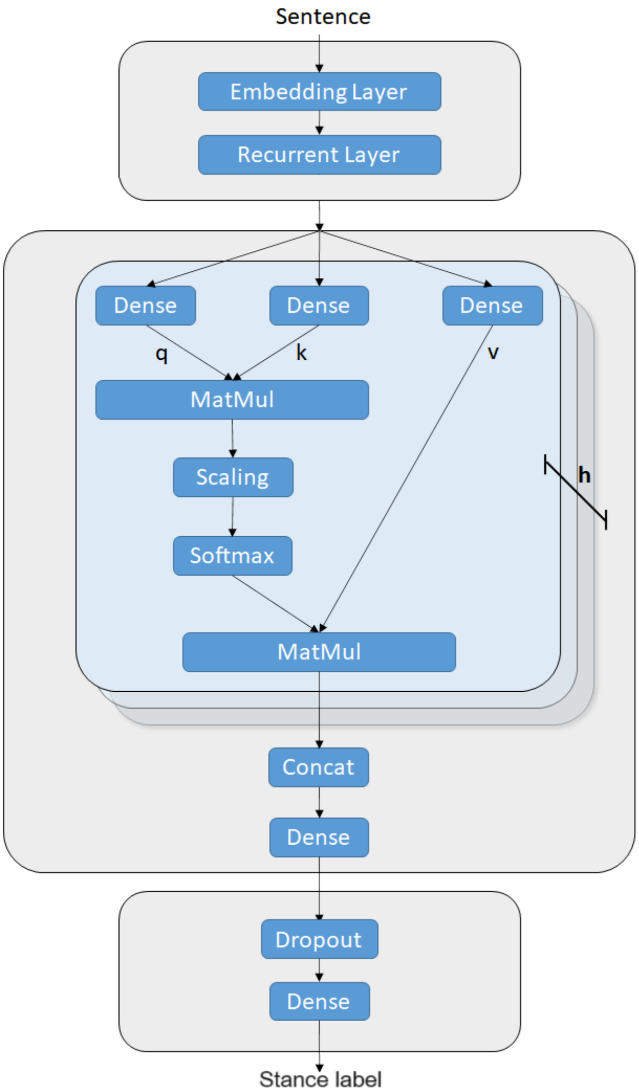

Figure 1. BertAttn architecture, adapted from Pavan *et al.* [2020].

introducing a novel zero-shot approach that makes use of off-task (that is, non-stance) labelled data instead. Finally, Section 4.4 describes a number of zero-shot models that do not require any training data at all, resorting instead to LLM prompting.

The models under consideration are summarised in Table 3, and discussed individually in the next sections. Models marked as '*' have been adapted from previous work to the Portuguese language and updated with a BERT layer instead of their original static embedding representations.

4.1 In-domain stance detection

In-domain settings - i.e., when a corpus of labelled instances towards the intended target is available for use as training data - are arguably the optimal scenario for stance detection, to the point where one may even ask whether less knowledge-intensive (e.g., zero-shot) methods may outperform in-domain classification at all. In fact, as discussed in the previous section, the comparison with in-domain stance detection is, in many cases, not even addressed in the evaluation of existing zero-shot models, and some datasets - most notably VAST [Allaway and McKeown, 2020] - arguably do not pay regard to in-domain stance detection.

As a means to contemplate a robust instance of in-domain classification in the present setting, we envisaged one such model, hereby called *BertAttn*, which has been previously shown to obtain high accuracy for the task and corpus at hand in in-domain settings [Pavan *et al.*, 2020]. The model consists of an adaptation of the bi-directional long short-term memory (BiLSTM) network with multi-head self-attention mechanism introduced in Pavan *et al.* [2020], presently using BERTabaporu [da Costa *et al.*, 2023] for text representation. This consists of a 128-256 embeddings layer followed by a recurrent layer with 16-128 LSTM units, attention model depth of 8 or 32, and 2 or 4 attention heads. The actual configuration for each task is determined by grid search. The model architecture is illustrated in Figure 1. For further details we refer to Pavan *et al.* [2020].

4.2 Zero-shot stance detection trained from out-of-domain labelled data

As discussed in Section 2, in much of the existing work in the field, zero-shot stance detection amounts to using out-of-domain (ood) labelled data for training purposes, that is, using data that has been labelled with stance information towards unseen targets only.

Zero-shot stance detection based on out-of-domain training data has been shown to obtain SOTA results, as documented in a wide range of studies devoted to the English language (e.g., Allaway and McKeown [2020]; Allaway *et al.* [2021]; Liu *et al.* [2021]; Luo *et al.* [2022]; Liang *et al.* [2022]; Zhao *et al.* [2022]; Chunling *et al.* [2023]; Wen and Hauptmann [2023]). However, we notice that adapting some of these systems to Portuguese stance detection may not be straightforward as many rely on language-specific resources currently unavailable for Portuguese[1]. This includes, for instance, the use of ConceptNet [Speer *et al.*, 2017], which plays a central role in Liu *et al.* [2021], Luo *et al.* [2022] and Chunling *et al.* [2023], text entailment relations in Xu *et al.* [2022], and the assumption that a large number of topics is available for the purpose of topic generalisation in Allaway and McKeown [2020]; Allaway *et al.* [2021], and others.

Bearing these difficulties in mind, we selected five models that are not overly reliant on language-specific resources, and which are often regarded as strong baseline systems for zero-shot stance detection based on out-of-domain labelled data. The present models, hereby called *BertBiCond*, *BertCrossNet*, *BertAttn*, *BertJointAttn*, and *BertAAD*, are updated versions of existing work in the field in which the original embedding layer has been replaced by a Portuguese BERT representation trained on a 237-million tweet corpus in the Brazilian Portuguese language as described in da Costa *et al.* [2023]. These models are summarised as follows.

- *BertBiCond* adapts the BiCond model in Augenstein *et al.* [2016] by replacing the original static embedding layer (originally computed using word2vec [Mikolov *et al.*, 2013]) for Portuguese BERTabaporu [da Costa *et al.*, 2023]. As pointed out in Allaway and McKeown

[1]For additional discussion on the reproducibility of existing stance detection models, see de Sousa and Becker [2023].

Table 3. Portuguese stance detection models.

Training	Model	Method
in-domain	BertAttn [Pavan *et al*., 2020]	BERT + BiLSTM
out-of-domain	BertAttn [Pavan *et al*., 2020]	BERT + BiLSTM
	BertJointAttn	BERT + BiLSTM
	BertCrossNet [Xu *et al*., 2018]*	
	BertBiCond [Augenstein *et al*., 2016]*	
	BertAAD [Pavan and Paraboni, 2022]	BERT + GAN
off-task	GovBR	TF-IDF + logreg
none	LLaMa	prompt-based
	ChatGPT	prompt-based
	GPT-4	prompt-based

[2020], the original approach has been regarded as the SOTA for tweet stance classification in the English language.

- *BertCrossNet* adapts the CrossNet model described in Xu *et al*. [2018], in which the original GloVe embedding layer [Pennington *et al*., 2014] has been replaced once again for BERTabaporu [da Costa *et al*., 2023].

- *BertAttn* is the multi-head attention architecture in Pavan *et al*. [2020], already used in our namesake in-domain baseline described in the previous section, but presently trained on out-of-domain data, that is, in standard zero-shot fashion.

- *BertJointAttn* is a target-aware version of *BertAttn* that encodes the target as an embedding vector represented by an additional BiLSTM layer, and then taken as an input to the multi-head attention layer. The joint embedding representation (i.e., text and target BiLSTM layers) is inspired by the use of target embeddings in Xu *et al*. [2018], and it is intended to provide contextual information to help disambiguate between targets in cross-target settings.

- *BertAAD* is an update from the adversarial discriminative domain adaptation (ADDA) sentiment analysis method originally introduced in Ryu and Lee [2022], and which was adapted to cross-target stance classification in Pavan and Paraboni [2022] using a preliminary version of the present dataset. The current model consists of a BERT model that has been fine-tuned for stance classification, and subsequently combined with ADDA [Tzeng *et al*., 2017] and knowledge distillation [Hinton *et al*., 2015].

4.3 Zero-shot stance detection trained from off-task labelled data

As an alternative to using out-of-domain labelled stances for training as discussed in the previous section, in what follows we introduce a novel approach to the task that considers the use of off-task information instead, that is, without any reliance on a stance corpus at all. This approach is motivated by the assumption that social media polarised discourse may correlate with a number of politically-charged targets (including those modelled in our present dataset). For instance, supporters of a particular left/right political party may hold reasonably consistent stance towards certain polarised subjects such as abortion legislation, drug control, etc. Thus, under this assumption, labelled data related to some other (i.e., non-stance) task may be used as a substitute for a labelled stance corpus, therefore implementing zero-shot stance classification in a distantly supervised fashion.

As a means to illustrate the use of off-task information, we consider a stance detection model called *GovBR*. The model is named after the corpus in da Silva and Paraboni [2023], a large database of Twitter timelines (i.e., collections of social media text publications) produced by both opponents and supporters of the former Brazilian government. The *GovBR* corpus was originally intended as train data for the task of author profiling for political orientation, that is, deciding whether a social media user is a supporter or an opponent to the government regardless of the actual topics under discussion. The corpus is labelled at the user (or timeline) level only, with information regarding the left/right political orientation of each individual determined by their use of carefully selected hashtags (e.g., '#EleNão' or '#NotHim' as a popular hashtag against the former president.) Most timeline publications are nevertheless largely unrelated to politics, comprising 13.5 million tweets written by 5452 unique users.

As training data to the *GovBR* model, we used a number of keywords to select tweets that are more likely to mention each of the targets in our stance detection corpus. These are summarised in Table 4 alongside the number of instances for each target.

By taking this off-task (that is, author profiling) corpus

Table 4. Number of tweets written by government supporters and opponents using selected keywords.

Target	Keywords	Government support	Government opposition
Lula	lula	96,387	122,397
Bolsonaro	bolsonaro, bozo (derogatory)	334,141	284,089
Hydrox.	cloroquina	24,445	18,881
Sinovac	coronavac,sinovac,vacina,vachina	10,388	7,616
Globo TV	globo	67,991	40,632
Church	igreja (church)	7,796	12,160

as training data for stance detection, *GovBR* aims to use political orientation information of Twitter users as a proxy to stance towards a particular target. However, this should not be confused with the use of out-of-domain training data in standard zero-shot stance detection discussed in the previous section. In particular, we notice that the standard (i.e., cross-target) approach uses as training data a set of labelled *stances* towards unseen targets, whereas the present approach does not take *any kind of stance data* at all (be it towards the intended target or towards any other.)

In addition to that, we notice that many texts containing a particular keyword (e.g., '*church*') are factual, that is, they do not convey any particular stance towards that target (e.g., '*That church square is poorly lit*' does not convey a stance towards the church as an institution). In other words, the selected text simply inherits the (supporter/opponent) political orientation label assigned to its author, but this label does not represent a stance towards the target 'church'. Moreover, some messages may even be off-target , as in, e.g., '*I believe the globe (globo, in Portuguese) is flat*', which does not refer to the Globo TV network at all.

Although these examples may in principle suggest that using GovBR author profiling data may be unhelpful for zero-shot stance detection, we assume that the sheer number of unlabelled training samples available (in some cases several times larger than the UstanceBR corpus training dataset), and at a very low cost, may overcome some of these difficulties, and that the effect of political alignment will still manifest some level of positive or negative correlation with the target. Thus, for instance, the use of the keyword '*church*' by supporters of the former (conservative) government may be more likely to include (unlabelled) stances in favour of the church than against it, whereas the use of the keyword '*vaccination*' by the same individuals may be more likely to include stances in favour of Hydroxichloroquine (as an alternative treatment for Covid-19) than a stance against it.

Using as training data only the selected portions of the GovBR corpus that are deemed relevant for each target, we envisaged a simple method based on logistic regression classifier over Tf-Idf counts. This uses balanced class weights, tol=0.0001, L2 penalty, a maximum of 150 iterations and lbfgs solver.

4.4 Zero-shot stance detection without training data

The rise of large language models (LLMs) has presented a wide range of novel opportunities for zero-shot stance detection in which the lack of task-specific training data is mitigated by querying the much larger knowledge base encoded by the LLM through prompt-based methods. More importantly, initial results for the *SemEval* stance corpus using ChatGPT in Zhang *et al*. [2023] suggest that prompt-based stance detection may outperform existing zero-shot methods trained on out-of-domain data, possibly representing the latest SOTA in the field.

Based on these observations, we presently consider three prompt-based methods for zero-shot stance detection in Portuguese. These models, hereby called *ChatGPT*, *LLaMa*, and *GPT-4* after their underlying language models, are described individually as follows.

As a means to assess whether a typical end user may perform stance detection simply by asking direct questions to a LLM-powered chatbot, we envisaged a baseline method - hereby called *ChatGPT* - that takes a prompt-based approach not unlike Zhang *et al*. [2023]. In this approach, the namesake chatbot tool receives instructions to evaluate a stance as either for or against a particular target.

We used a (Portuguese) prompt in the form 'Does the phrase <*tweet*> reflect positively or negatively upon <*target*>?', in which both tweet and targets are taken from the test instances in the UstanceBR corpus. Prompts were submitted to ChatGPT March 2023 version, and responses that were not positive or negative were assigned a random positive/negative label. These included both responses that were deemed neutral or ambivalent, as well as those that somewhat evaded the question.

As an alternative to the direct query method, we also consider two more elaborate strategies based on LLM-prompting for stance detection. These strategies - hereby called *LLaMa* and *GPT-4* - differ from each other mainly in their underlying language model, but they do use slightly modified prompt versions obtained after a number of trials (presently omitted for brevity.) This was necessary due to variations in the way each model interprets the input instruction, and in the way responses are produced.

The *LLaMa* approach uses Alpaca[2], a fine-tuned version of the LLaMa language model described in Touvron *et al*. [2023]. The actual prompt used for inference and evaluation is as follows.

[2]`https://crfm.stanford.edu/2023/03/13/alpaca.html`

Below is an instruction that describes a task, paired with an input that provides further context. Write a response that appropriately completes the request.

Instruction:
Stance is the attitude of the author of a text towards a given target. Read the following text and give a score between 0 and 10 where 0 means that the text is totally against of the target and 10 means the text is totally in favour of the target, in the format 'x/10'

Input:
Text: {*text*}
Target: {*target*}

Response:

When using the prompt above, the tokens "*{text}*" and "*{target}*" are replaced by the corresponding values that are required for inference. As a means to enforce greater consistency, model temperature is set to zero, and the prompt specifies that the response should be provided as a numeric value between 0 and 10. However, as it is often the case with LLMs, there is still considerable variability in the response contents and format. For that reason, a post-processing function was created using the regular expressions "([0-9]+)(/[0-9]+)*" to search for the response pattern. This is then converted to a decimal number to be taken as the prediction score.

Similarly, our *GPT-4* approach uses GPT-4-0314[3] as the underlying LLM, and a slightly modified prompt as follows:

Introduction
You are an expert professional assistant with over 30 years of experience in the job. You specialised in reading and identifying stances in texts in relation to a given target. Below is a text and a target of interest.

Main Task
Read the text below and identify what is the stance of the text with relation to the specified target. The rest of this prompt contains further instructions on this task.

Stance
Generate a score that represents the stance such that:
1. Stance is the attitude of the author of a text towards a given target
2. In this scenario, a stance can be in favour or against the target.
3. Your task is to identify the stance of the author of the text in relation to the specified target.
4. Your response should be a score between 0 and 10 where 0 means that the text is totally against the target and 10 means the text is totally in favour of the target, in the format 'x/10'

<|**endofprompt**|>

Target:
$Target

Text:
$Text

Possible answer:

At inference time, the tokens "*$Target*" and "*$Text*" are replaced by the respective values. The generated output is post-processed using the same regular-expression based function used in the above *LLaMa* approach.

[3] `https://openai.com/gpt-4`

5 Results

Table 5 summarises F1 score test results obtained by the ten models under discussion across the six targets available from the *UstanceBR* corpus. The best zero-shot results for each target (that is, disregarding the in-domain approach results on the top row of the table) are highlighted.

Results in Table 5 show that *GPT-4* is the overall best strategy, in most cases surpassing even in-domain classification. More generally, both *GPT-4* and the other prompt-based models *LLaMa* and *ChatGPT* outperform the standard out-of-domain alternatives that have dominated the field. This outcome confirms the recent advances in LLMs and prompt engineering for NLP tasks and, particularly in the case of the present *GPT-4* model, sets a considerably high benchmark for stance detection.

Leaving aside the comparison between the standard out-of-domain approach versus prompt engineering, a remarkable outcome is the results obtained by the use of off-task politically aligned information represented by distant labelled data in *GovBR*, which fares considerably above all the out-of-domain options. This model, which relies on a simplistic linear method and bag-of-words text representation, is second only to the prompt-based models based on large (and costly) language models, an outcome that suggests opportunities for further improvement.

Finally, the observation that some zero-shot methods outperform even in-domain classification may seem in principle counter-intuitive. Upon close inspection, however, we notice that the corpus includes a certain number of annotation errors. These errors, which are perhaps intrinsic to human annotated data, are obviously a source of noise for methods that rely on training data, including both in- and out-of-domain models. Prompt-based methods, by contrast, do not use the annotated corpus at all, and are therefore immune from this kind of noise.

6 Final remarks

This article presented a number of Portuguese stance detection models adapted from existing work in zero-shot with SOTA results for the English language, and novel methods based on social media politically aligned data and LLM prompt engineering. The models under evaluation cover different uses of labelled data, ranging from standard out-of-domain approaches to the use of off-task information and no labelled stance data at all.

All models were tested on a large corpus of social media data in Portuguese, whose results present evidence of the superiority of recent prompt-based models over the alternatives. This is particularly compelling in the case of *GTP-4* prompting, which outperforms even in-domain classification.

Perhaps more surprisingly, however, the use of off-task politically aligned discourse information in *GovBR* turned out to produce competitive results if compared to the more robust LLM-base strategies, and for a fraction of their computational costs.

As future work, we intend to improve the present *GovBR* approach by using more sophisticated learning methods and

Table 5. F1 results across training strategies, models and test targets.

Training	Model	Lula	Bolsonaro	Hydrox.	Sinovac	Globo TV	Church	Overall
in-domain	BertAttn	0.87	0.82	0.95	0.79	0.93	0.86	0.87
out-of-domain	BertAttn	0.58	0.54	0.51	0.49	0.59	0.29	0.50
	BertJointAttn	0.50	0.51	0.41	0.59	0.49	0.33	0.47
	BertCrossNet	0.36	0.33	0.38	0.42	0.43	0.44	0.39
	BertBiCond	0.47	0.42	0.60	0.44	0.41	0.52	0.47
	BertAAD	0.55	0.38	0.53	0.43	0.46	0.39	0.45
off-task	GovBR	0.82	0.65	0.76	0.72	0.65	0.56	0.69
none	LLaMa	0.74	0.52	0.69	0.45	0.70	0.34	0.57
	ChatGPT	0.84	0.69	**0.82**	0.76	0.81	0.77	0.78
	GPT-4	**0.93**	**0.87**	0.81	**0.88**	**0.93**	**0.92**	**0.89**

text representations, and investigate whether the off-task approach may obtain results on a par with LLM-based strategies. Moreover, we intend to add further models to the current set, and consider using a translated version of the current corpus, which would allows us to exploit a wider range of resources only available for the English language while still presenting results that could be compared against the present experiments.

Finally, yet another avenue of research is the evaluation of *GovBR* in other datasets and domains, possibly extending the use of off-task politically aligned information to languages other than Portuguese as a means to assess its generalisation capabilities.

Declarations

Funding

This research was partially funded by the São Paulo Research Foundation (FAPESP grant #2021/08213-0). The authors are also grateful to the anonymous reviewers for their valuable feedback regarding this work.

Authors' Contributions

MCP is the main contributor, who developed and tested the models under discussion. IP contributed to the conception of this study and helped writing this manuscript. Both authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The main dataset used during the current study, called UstanceBR corpus, is available from `https://drive.google.com/drive/folders/1qThfcIeOHjwVbsDVgkot-AqgnXoJdBOK`.

Code for the models under discussion is available from `https://github.com/mcpavan/pt_zero_shot_stance`.

References

Aldayel, A. and Magdy, W. (2021). Stance detection on social media: State of the art and trends. *Information Processing & Management*, 58(4):102597. DOI: 10.1016/j.ipm.2021.102597.

Allaway, E. and McKeown, K. R. (2020). Zero-shot stance detection: A dataset and model using generalized topic representations. In *EMNLP-2020 proceedings*, pages 8913–8931, Online. Assoc. for Computational Linguistics. DOI: 10.18653/v1/2020.emnlp-main.717.

Allaway, E. and McKeown, K. R. (2022). Zero-shot stance detection: Paradigms and challenges. *Frontiers in Artificial Intelligence*, 5:1070429. DOI: 10.3389/frai.2022.1070429.

Allaway, E., Srikanth, M., and McKeown, K. (2021). Adversarial learning for zero-shot stance detection on social media. In *ACL-2021 proceedings*, pages 4756–4767, Online. Assoc. for Computational Linguistics. DOI: 10.18653/v1/2021.naacl-main.379.

Alturayeif, N., Luqman, H., and Ahmed, M. (2023). A systematic review of machine learning techniques for stance detection and its applications. *Neural Comput & Applic*, 35:5113–5144. DOI: 10.1007/s00521-023-08285-7.

Augenstein, I., Rocktäschel, T., Vlachos, A., and Bontcheva, K. (2016). Stance detection with bidirectional conditional encoding. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 876–885, Austin, Texas. Association for Computational Linguistics. DOI: 10.18653/v1/D16-1084.

Chunling, W., Yijia, Z., Xingyu, Y., Guantong, L., Fei, C., and Hongfei, L. (2023). Adversarial network with external knowledge for zero-shot stance detection. In *Proceedings of the 22nd Chinese National Conference on Computational Linguistics*, pages 824–835, Harbin, China. Chinese Information Processing Society of China. DOI: 10.1007/978-981-99-6207-$5_2$6.

Cignarella, A., Lai, M., Bosco, C., Patti, V., and Rosso, P. (2020). SardiStance @ EVALITA2020: Overview of the task on stance detection in italian tweets. In

CEURS Proceedings vol. 2765, pages 177–186, online. CEUR-WS.org. Available at: `https://ceur-ws.org/Vol-2765/paper159.pdf`.

da Costa, P. B., Pavan, M. C., dos Santos, W. R., da Silva, S. C., and Paraboni, I. (2023). BERTabaporu: assessing a genre-specific language model for Portuguese NLP. In *Recents Advances in Natural Language Processing (RANLP-2023)*, pages 217–223, Varna, Bulgaria. Available at: `https://aclanthology.org/2023.ranlp-1.24`.

da Silva, S. C., Ferreira, T. C., Ramos, R. M. S., and Paraboni, I. (2020). Data driven and psycholinguistics motivated approaches to hate speech detection. *Computación y Systemas*, 24(3):1179–1188. DOI: 10.13053/cys-24-3-3478.

da Silva, S. C. and Paraboni, I. (2023). Politically-oriented information inference from text. *Journal of Universal Computer Science*, 29(6):570–595. DOI: 10.3897/jucs.96652.

de Sousa, A. and Becker, K. (2023). SSSD: Leveraging pretrained models and semantic search for semi-supervised stance detection. In *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*, pages 264–273, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria. Available at: `https://aclanthology.org/2023.ranlp-1.30`.

Delmondes Neto, J. P. and Paraboni, I. (2021). Multi-source BERT stack ensemble for cross-domain author profiling. *Expert Systems*, pages –. DOI: 10.1111/exsy.12869.

Devlin, J., Chang, M., Lee, K., and Toutanova, K. (2019). BERT: pre-training of deep bidirectional transformers for language understanding. In *NAACL-2019 proceedings*, pages 4171–4186, Minneapolis, USA. Assoc. for Computational Linguistics. DOI: 10.18653/v1/N19-1423.

Flores, A. M., Pavan, M. C., and Paraboni, I. (2021). User profiling and satisfaction inference in public information access services. *Journal of Intelligent Information Systems*, pages –. DOI: 10.1007/s10844-021-00661-w.

Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network. In *NIPS Deep Learning and Representation Learning Workshop.*. DOI: 10.48550/arXiv.1503.02531.

Hosseinia, M., Dragut, E., and Mukherjee, A. (2020). Stance prediction for contemporary issues: Data and experiments. In *8th International Workshop on Natural Language Processing for Social Media*, pages 32–40, Online. Assoc. for Computational Linguistics. DOI: 10.18653/v1/2020.socialnlp-1.5.

Jaziriyan, M. M., Akbari, A., and Karbasi, H. (2021). ExaASC: A General Target-Based Stance Detection Corpus in Arabic Language. In *11th International Conference on Computer Engineering and Knowledge (ICCKE)*, pages 424–429, Mashhad, Iran. IEEE. DOI: 10.1109/ICCKE54056.2021.9721486.

Kochkina, E., Liakata, M., and Augenstein, I. (2017). Turing at SemEval-2017 task 8: Sequential approach to rumour stance classification with branch-LSTM. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 475–480, Vancouver, Canada. Association for Computational Linguistics. DOI: 10.18653/v1/S17-2083.

Kucuk, D. and Can, F. (2020). Stance detection: A survey. *ACM Computing Surveys*, 53(1):1–37. DOI: 10.1145/3369026.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., and Zettlemoyer, L. (2020). BART: Denoising sequence-to-sequence pretraining for natural language generation, translation, and comprehension. In Jurafsky, D., Chai, J., Schluter, N., and Tetreault, J., editors, *58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics. DOI: 10.48550/arXiv.1910.13461.

Li, Y., Sosea, T., Sawant, A., Nair, A. J., Inkpen, D., and Caragea, C. (2021). P-stance: A large dataset for stance detection in political domain. In *Findings of ACL-IJCNLP-2021*, pages 2355–2365, Online. Assoc. for Computational Linguistics. DOI: 10.18653/v1/2021.findings-acl.208.

Liang, B., Zhu, Q., Li, X., Yang, M., Gui, L., He, Y., and Xu, R. (2022). JointCL: A joint contrastive learning framework for zero-shot stance detection. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 81–91, Dublin, Ireland. Association for Computational Linguistics. DOI: 10.18653/v1/2022.acl-long.7.

Liu, R., Lin, Z., Tan, Y., and Wang, W. (2021). Enhancing zero-shot and few-shot stance detection with commonsense knowledge graph. In *Findings of ACL-IJCNLP-2021*, pages 3152–3157, Online. Assoc. for Computational Linguistics. DOI: 10.18653/v1/2021.findings-acl.278.

Luo, Y., Liu, Z., Shi, Y., Li, S. Z., and Zhang, Y. (2022). Exploiting sentiment and common sense for zero-shot stance detection. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 7112–7123, Gyeongju, Republic of Korea. International Committee on Computational Linguistics. DOI: 10.48550/arXiv.2208.08797.

Mikolov, T., Wen-tau, S., and Zweig, G. (2013). Linguistic regularities in continuous space word representations. In *Proc. of NAACL-HLT-2013*, pages 746–751, Atlanta, USA. Assoc. for Computational Linguistics. Available at: `https://aclanthology.org/N13-1090`.

Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X., and Cherry, C. (2016). SemEval-2016 task 6: Detecting stance in tweets. In *10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41, San Diego, California. Assoc. for Computational Linguistics. DOI: 10.18653/v1/S16-1003.

Oshikawa, R., Qian, J., and Wang, W. Y. (2020). A survey on natural language processing for fake news detection. In *12th Language Resources and Evaluation Conference*, pages 6086–6093, Marseille, France. European Language Resources Association. Available at: `https://aclanthology.org/2020.lrec-1.747`.

Pavan, M. C., dos Santos, V. G., Lan, A. G. J., ao Trevisan Martins, J., dos Santos, W. R., Deutsch, C., da Costa, P. B., Hsieh, F. C., and Paraboni, I. (2023). Morality classification in natural language text. *IEEE trans-*

actions on Affective Computing, 14(1):857–863. DOI: 10.1109/TAFFC.2020.3034050.

Pavan, M. C., dos Santos, W. R., and Paraboni, I. (2020). Twitter Moral Stance Classification using Long Short-Term Memory Networks. In *BRACIS-2020 proceedings LNAI 12319*, pages 636–647. Springer. DOI: 10.1007/978-3-030-61377-8$_2$8.

Pavan, M. C. and Paraboni, I. (2022). Cross-target stance classification as domain adaptation. In Pichardo Lagunas, O., Martínez-Miranda, J., and Martínez Seis, B., editors, *Advances in Computational Intelligence - MICAI 2022 - Lecture Notes in Artificial Intelligence vol 13612*, pages 15–25, Cham. Springer Nature Switzerland. DOI: 10.1007/978-3-031-19493-1.

Pennington, J., Socher, R., and Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. In *EMNLP-2014 proceedings*, pages 1532–1543. DOI: 10.3115/v1/D14-1162.

Pereira, C., Pavan, M., Yoon, S., Ramos, R., Costa, P., Cavalheiro, L., and Paraboni, I. (2023). UstanceBR: a multimodal language resource for stance prediction. *arXiv:2312.06374*. DOI: 10.48550/arXiv.2312.06374.

Ryu, M. and Lee, K. (2022). Knowledge distillation for BERT unsupervised domain adaptation. *Knowledge Information Systems*, 64:3113–3128. DOI: 10.1007/s10115-022-01736-y.

Schlichtkrull, M., Kipf, T. N., Bloem, P., van den Berg, R., Titov, I., and Welling, M. (2018). Modeling relational data with graph convolutional networks. In *The Semantic Web*, pages 593–607, Cham. Springer International Publishing. DOI: 10.1007/978-3-319-93417-4$_3$8.

Schmidt, A. and Wiegand, M. (2017). A survey on hate speech detection using natural language processing. In *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, pages 1–10, Valencia, Spain. Association for Computational Linguistics. DOI: 10.18653/v1/W17-1101.

Speer, R., Chin, J., and Havasi, C. (2017). ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, pages 4444–4451. AAAI Press. DOI: 10.1609/aaai.v31i1.11164.

Taulé, M., Martí, M. A., Pardo, F. M. R., Rosso, P., Bosco, C., and Patti, V. (2017). Overview of the task on stance and gender detection in tweets on catalan independence at IberEval 2017. In *IberEval-2017 proceedings*, pages 157–177, Murcia, Spain. CEUR-WS.org. Available at: `https://api.semanticscholar.org/CorpusID:958864`.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., and Lample, G. (2023). Llama: Open and efficient foundation language models. *arXiv*. DOI: 10.48550/arXiv.2302.13971.

Tzeng, E., Hoffman, J., Saenko, K., and Darrell, T. (2017). Adversarial discriminative domain adaptation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2962–2971. DOI: 10.1109/CVPR.2017.316.

Vashishth, S., Sanyal, S., Nitin, V., and Talukdar, P. P. (2020). Composition-based multirelational graph convolutional networks. In *8th International Conference on Learning Representations (ICLR 2020)*, Addis Ababa, Ethiopia. DOI: 10.48550/arXiv.1911.03082.

Wen, H. and Hauptmann, A. (2023). Zero-shot and few-shot stance detection on varied topics via conditional generation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1491–1499, Toronto, Canada. Association for Computational Linguistics. DOI: 10.18653/v1/2023.acl-short.127.

Xu, C., Paris, C., Nepal, S., and Sparks, R. (2018). Cross-target stance classification with self-attention networks. In *ACL-2018 proceedings*, pages 778–783, Melbourne, Australia. Assoc. for Computational Linguistics. DOI: 10.18653/v1/P18-2123.

Xu, H., Vucetic, S., and Yin, W. (2022). OpenStance: Real-world zero-shot stance detection. In *Proceedings of the 26th Conference on Computational Natural Language Learning (CoNLL)*, pages 314–324, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics. DOI: 10.18653/v1/2022.conll-1.21.

Yang, R., Xie, W., Liu, C., and Yu, D. (2019). BLCU_NLP at SemEval-2019 task 7: An inference chain-based GPT model for rumour evaluation. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 1090–1096, Minneapolis, Minnesota, USA. Association for Computational Linguistics. DOI: 10.18653/v1/S19-2191.

Zarrella, G. and Marsh, A. (2016). MITRE at SemEval-2016 task 6: Transfer learning for stance detection. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 458–463, San Diego, California. Association for Computational Linguistics. DOI: 10.18653/v1/S16-1074.

Zhang, B., Ding, D., and Jing, L. (2023). How would stance detection techniques evolve after the launch of chatgpt?

Zhang, B., Yang, M., Li, X., Ye, Y., Xu, X., and Dai, K. (2020). Enhancing cross-target stance detection with transferable semantic-emotion knowledge. In *ACL-2020 proceedings*, pages 3188–3197, Online. Assoc. for Computational Linguistics. DOI: 10.18653/v1/2020.acl-main.291.

Zhang, Y., Barzilay, R., and Jaakkola, T. S. (2017). Aspect-augmented adversarial networks for domain adaptation. *Trans. Assoc. Comput. Linguistics*, 5:515–528. DOI: 10.48550/arXiv.1701.00188.

Zhao, X., Zou, J., Zhang, Z., Xie, F., Zhou, B., and Tian, L. (2022). Zero-shot stance detection based on cross-domain feature enhancement by contrastive learning. *arXiv:2210.03380*. DOI: 10.48550/arXiv.2210.03380.