applied sciences

MDPI

Article

Exploring Multi-Agent Debate for Zero-Shot Stance Detection: A Novel Approach

Junxia Ma [1], Changjiang Wang [1], Lu Rong [2,*], Bo Wang [1] and Yaoli Xu [1]

[1] College of Software Engineering, Zhengzhou University of Light Industry, Zhengzhou 450000, China; jxma@zzuli.edu.cn (J.M.); changjiang001121@163.com (C.W.); wangb@zzuli.edu.cn (B.W.); yaolixu@zzuli.edu.cn (Y.X.)

[2] School of Education, Tianjin University, Tianjin 300072, China

* Correspondence: ronglu@tju.edu.cn

Abstract: Zero-shot stance detection aims to identify the stance expressed in social media text aimed at specific targets without relying on annotated data. However, due to insufficient contextual information and the inherent ambiguity of language, this task faces numerous challenges in low-resource scenarios. This work proposes a novel zero-shot stance detection method based on multi-agent debate (ZSMD) to address the aforementioned challenges. Specifically, we construct two debater agents representing the supporting and opposing stances. A knowledge enhancement module supplements the original tweet and target with relevant background knowledge, providing richer contextual support for argument generation. Subsequently, the two agents engage in debate over a predetermined number of rounds, employing rebuttal strategies such as factual verification, logical analysis, and sentiment analysis. If no consensus is reached within the specified rounds, a referee agent synthesizes the debate process and original input information to deliver the final stance determination. We evaluate ZSMD on two benchmark datasets, SemEval-2016 Task 6 and P-Stance, and compare it against strong zero-shot baselines such as MB-Cal and COLA. The experimental results show that ZSMD not only achieves higher accuracy than these baselines, but also provides deeper insights into subtle differences in opinion expression, highlighting the potential of structured argumentation in low-resource settings.

Keywords: zero-shot stance detection; multi-agent debate; knowledge augment; ZSMD

check for updates

Academic Editor: Jose María Alvarez Rodríguez

Received: 22 March 2025
Revised: 15 April 2025
Accepted: 20 April 2025
Published: 22 April 2025

Citation: Ma, J.; Wang, C.; Rong, L.; Wang, B.; Xu, Y. Exploring Multi-Agent Debate for Zero-Shot Stance Detection: A Novel Approach. *Appl. Sci.* **2025**, *15*, 4612. https://doi.org/10.3390/app15094612

1. Introduction

With the rapid proliferation of social media and online platforms, stance detection has emerged as a core task within the field of natural language processing (NLP). Stance detection aims to automatically identify the attitude or stance toward a specific target expressed in text—such as social events, political figures, or policies—typically categorized into three classes: support, opposition, or neutrality. As the speed and complexity of information dissemination continue to escalate, efficiently analyzing and understanding public attitude trends within vast datasets has become a critical research topic across multiple disciplines, including social sciences, political science, and business analytics. Particularly in social media contexts, text is often characterized by its brevity, informality, emotional richness, and inherent biases, which significantly heighten the technical challenges of stance detection [1–4]. Consequently, developing efficient methods capable of accurately capturing textual stances holds not only academic value but also profound practical significance for applications such as public opinion monitoring and market analysis.

Traditional stance detection methods predominantly rely on supervised learning, utilizing large volumes of annotated data to train models for identifying stances toward specific targets. However, the process of collecting high-quality annotated data is both time-consuming and costly, a challenge that becomes even more pronounced when addressing domain-specific or newly emerging targets [4]. To overcome this bottleneck, zero-shot stance detection (ZSSD) [5,6] has been developed. The primary objective of ZSSD is to determine the stance of text aimed at unknown targets without requiring annotated data specific to the target task [5]. The significance of this approach lies in its ability to break free from the dependency on extensive annotated datasets inherent in traditional supervised learning, substantially reducing data preparation costs and enhancing model generalization across emerging topics or cross-domain tasks. For instance, during sudden events (e.g., pandemics or political crises), ZSSD can swiftly adapt to new targets, enabling researchers or decision-makers to promptly capture public attitudes without awaiting the completion of data annotation processes [7]. Consequently, ZSSD offers a novel research paradigm for stance detection in theoretical terms and demonstrates broad practical potential in applications such as real-time public opinion analysis and cross-lingual stance identification.

Although ZSSD holds significant potential in addressing data scarcity, its practical application is still hindered by numerous challenges that limit its accuracy and robustness. Firstly, the absence of annotated data specific to the target task makes it difficult for models to learn semantic features directly related to particular targets, resulting in unstable performance on unknown targets [8,9]. For instance, traditional deep learning models rely on target-specific training samples; when confronted with unseen events or entities, they often fail to accurately comprehend the deeper connections between the text and the target. Secondly, the diversity and complexity of social media text further exacerbates the difficulties of ZSSD. This text frequently encompasses informal expressions, metaphors, sarcasm, and even multilingual mixtures, increasing the likelihood of models misjudging the author′s true stance [10,11]. For example, a sarcastic comment may superficially appear to support a viewpoint while its actual stance is one of opposition—a subtle distinction that existing methods often struggle to discern. Additionally, many current ZSSD approaches depend on transfer learning or prompt engineering to achieve zero-shot capabilities, yet these techniques remain inadequate in capturing the deep semantic relationships between text and target [12]. For instance, prompt-based methods leveraging pre-trained language models may be confined to surface-level semantics, lacking the ability to fully reason about the implicit attitudes underlying the text, which constrains model performance in complex contexts. These challenges underscore the need for ZSSD to incorporate stronger semantic understanding and reasoning mechanisms to address the dual shortcomings of data scarcity and contextual complexity [11,13].

To address these challenges, this work proposes a zero-shot stance detection method based on multi-agent debate (ZSMD). Our approach leverages the reasoning capabilities of multi-agent debate to enhance the model′s inference and judgment in complex contexts. By simulating viewpoint conflicts and argumentation processes among multiple agents, multi-agent debate effectively facilitates stance identification and classification. Within this framework, each agent represents a potential perspective or stance, and through their interactions and debates, they not only aid in distinguishing various stances but also uncover subtle, latent information embedded in the text, thereby improving detection accuracy. Compared to traditional stance detection methods, multi-agent debate offers a dynamic reasoning mechanism, enabling the model to better adapt to unknown targets and unstructured text.

Furthermore, existing stance detection methods often overlook the supplementation of target-related background knowledge, which is critical for accurately determining stance [14]. For instance, when assessing an individual's stance toward a specific policy, the absence of background information about the policy's content may prevent the model from correctly interpreting references or implicit meanings within the text. To address this, we propose a multi-task learning strategy that enhances the model's semantic understanding by integrating debate data and target-relevant background knowledge. During the multi-agent debate process, we dynamically incorporate external knowledge—such as event contexts and entity relationships—into the model, enabling it to reference more comprehensive contextual information during reasoning. This approach not only compensates for the shortcomings of traditional ZSSD in terms of data scarcity and contextual inadequacy but also improves the model's reasoning capabilities and cross-task robustness through multi-task training.

In summary, the primary contributions of this work are as follows:

- We propose a stance detection method based on multi-agent debate, integrating stance analysis and reasoning processes, which significantly improves model performance in complex contexts.
- By incorporating background knowledge and debate data, we enhance the model's semantic understanding, addressing performance bottlenecks in zero-shot stance detection.
- Experimental validation on two public datasets demonstrates that the method proposed in this paper significantly outperforms existing stance detection approaches, achieving state-of-the-art (SOTA) results.

2. Related Work

This section reviews three areas of research closely related to the present study: Stance detection, background knowledge-enhanced stance detection, and multi-agent debate. These studies provide the theoretical foundation and technical support for the multi-agent debate framework proposed in this paper.

2.1. Zero-Shot Stance Detection

Zero-shot stance detection (ZSSD) is an emerging task in natural language processing aimed at identifying stances toward unknown targets by leveraging stance information learned from known targets [15]. In contrast to traditional stance detection methods, ZSSD enhances model generality, making it particularly valuable in scenarios with scarce data or rapidly emerging new targets, such as rumor detection and public opinion analysis during sudden events.

Early studies on ZSSD primarily relied on transfer learning techniques. Researchers trained models on known source targets and transferred the acquired knowledge to unknown targets. Allaway and McKeown [5] were the first to explicitly propose the ZSSD task, designing a method based on target-invariant representation. This approach learns text features independent of specific targets, enabling stance prediction for unseen targets. With the development of large-scale pre-trained language models (PLMs), recent approaches have transformed ZSSD into a prompt-based or reasoning-driven task. For example, Schick et al. [16] proposed Pattern-Exploiting Training (PET), which reformulates classification as a masked language modeling task guided by textual patterns. While effective, these methods often depend heavily on prompt quality and lack robustness in complex scenarios.

To further improve reasoning and generalization in zero-shot stance detection, several representative approaches have been proposed in recent years. Wang et al. [17] introduced the ANEK framework, which leverages adversarial training to extract transferable knowledge from seen targets, while incorporating sentiment information and commonsense knowledge to enhance contextual understanding. Taranukhin et al. [18] proposed Stance

Reasoner, a model that integrates pre-trained language models with chain-of-thought reasoning to generate intermediate steps for interpretable stance prediction. Zhang et al. [19] developed a commonsense-based adversarial framework, which uses external commonsense graphs and a feature separation adversarial network to capture both target-invariant and target-specific features. Additionally, Wen et al. [20] reformulated stance detection as a conditional generation task, combining target prediction and unlikelihood training to improve the semantic alignment between input text, targets, and stance labels.

These methods demonstrate that ZSSD is an active and evolving field, where knowledge incorporation, prompt design, and reasoning strategies play increasingly important roles. Our work builds on this line of research by introducing a structured multi-agent debate framework that integrates external knowledge while enabling interpretable, dynamic argumentation for robust zero-shot stance detection.

2.2. Background Knowledge-Enhanced Stance Detection

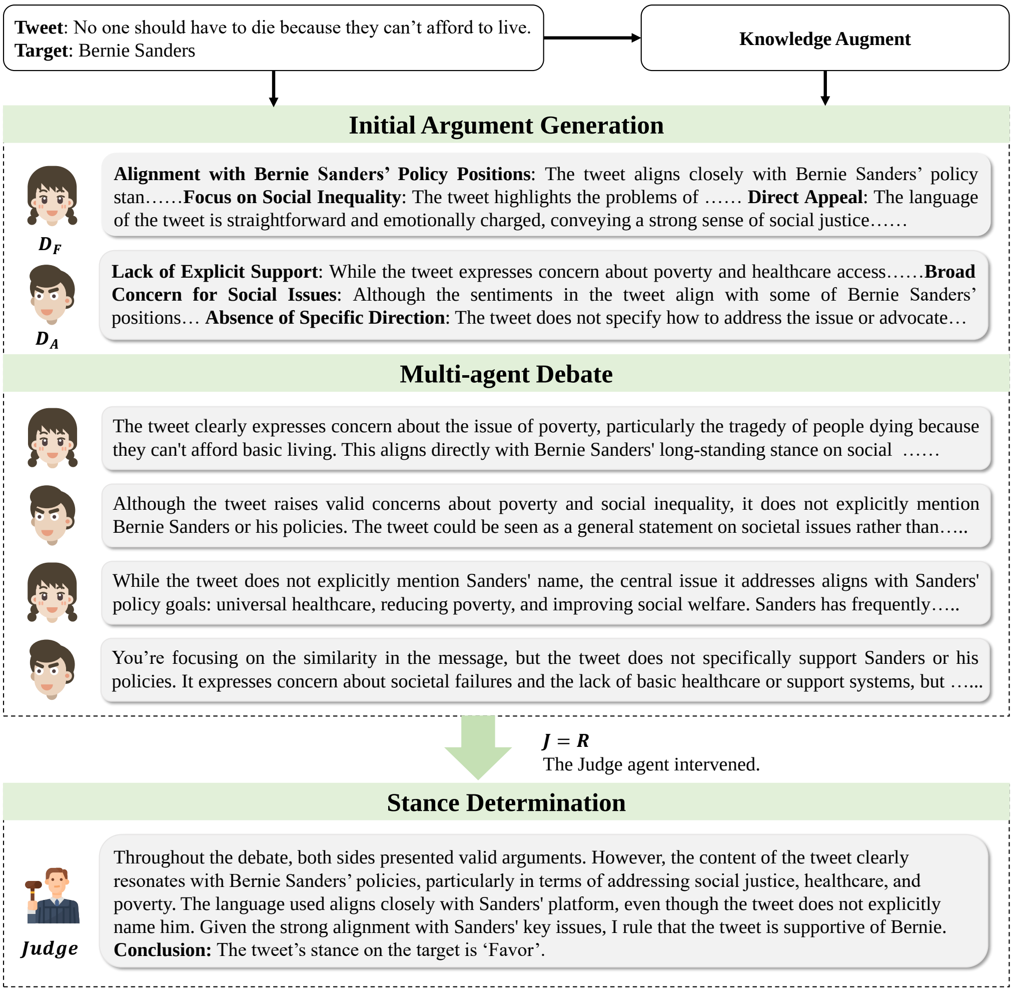
Background knowledge-enhanced stance detection improves a model's ability to understand the relationship between text and a target by incorporating external knowledge sources. Stance detection often relies on contextual information, yet the text itself may omit critical background details, such as policy content or event timelines. The introduction of background knowledge addresses this deficiency [21,22]. Research indicates that background knowledge plays a significant role in the following aspects. Disambiguation: Assists the model in interpreting ambiguous terms or references within the text. For instance, when analyzing a person's comment on a climate change policy, the absence of policy background might lead the model to misjudge the stance. Contextual clues: Provides additional target-related information to enhance semantic understanding. For example, entity relationships in a knowledge graph can uncover deeper connections between the text and the target. Methods leveraging knowledge enhancement to improve stance detection performance have garnered widespread attention from researchers. For instance, Yan et al. [14] aligned target-related background knowledge from diverse sources, collaboratively utilizing multi-source knowledge to enhance the model's understanding and detection capabilities regarding the target. Wang et al. [15] proposed a meta-contrastive learning approach with a data augmentation framework, employing a generative model to produce target-critical phrases to enrich the original text. Similarly, Ding et al. [23] introduced an encoder–decoder data augmentation framework.

2.3. Multi-Agent Debate

Multi-agent debate (MAD) is fundamentally an interactive reasoning and decision-making framework in which multiple agents assume distinct roles and engage in "debate" or "collaboration" to guide the model toward eliminating errors and arriving at more robust conclusions. Initially, this approach was predominantly applied in game theory and automated dialogue systems. However, in the context of large language models, researchers have discovered that leveraging mutual questioning, inquiry, and response among multiple agents can effectively mitigate hallucination phenomena while enhancing the diversity and interpretability of reasoning [24–26]. In recent years, MAD has increasingly been incorporated into various natural language processing tasks. Park et al. [27] proposed a framework that employs multi-agent debate for hate speech detection. Wang et al. [28] introduced a Multi-Agent Debate with Knowledge-Enhanced framework (MADKE) to address the issue of cognitive islands among individual agents. Du et al. [29] suggested a method where multiple agents independently analyze and propose solutions for a given task, engaging in critical debate and scrutiny of each other's proposals to enhance the model's factual accuracy and reasoning capabilities. Similarly, Smit et al. [30] utilized a multi-agent debate approach to improve performance in medical question answering. While these methods demonstrate the growing versatility of multi-agent debate in various NLP tasks, to the best of our knowledge, no prior work has explicitly applied a multi-agent debate framework to zero-shot stance detection. Our work is the first to explore this integration, combining MAD with knowledge enhancement for interpretable stance reasoning in zero-shot scenarios.

3. Method

This study proposes a zero-shot stance detection method based on multi-agent debate (ZSMD). By simulating the human debate process, we establish two debater agents, representing the supporting and opposing sides, while incorporating background knowledge relevant to the target task. Through a series of arguments and rebuttals, the framework infers the stance of an original tweet toward a specific target. Our approach comprises three key stages: the initialization stage, the debate stage, and the stance determination stage. An overview of the entire ZSMD framework, including an illustrative example, is shown in Figure 1. In the following sections, we provide a detailed description of each stage of the method.



Tweet: No one should have to die because they can’t afford to live.
Target: Bernie Sanders

Alignment with Bernie Sanders’ Policy Positions: The tweet aligns closely with Bernie Sanders’ policy stan……**Focus on Social Inequality**: The tweet highlights the problems of …… **Direct Appeal**: The language of the tweet is straightforward and emotionally charged, conveying a strong sense of social justice……

Lack of Explicit Support: While the tweet expresses concern about poverty and healthcare access……**Broad Concern for Social Issues**: Although the sentiments in the tweet align with some of Bernie Sanders’ positions… **Absence of Specific Direction**: The tweet does not specify how to address the issue or advocate…

The tweet clearly expresses concern about the issue of poverty, particularly the tragedy of people dying because they can't afford basic living. This aligns directly with Bernie Sanders' long-standing stance on social ……

Although the tweet raises valid concerns about poverty and social inequality, it does not explicitly mention Bernie Sanders or his policies. The tweet could be seen as a general statement on societal issues rather than…..

While the tweet does not explicitly mention Sanders' name, the central issue it addresses aligns with Sanders' policy goals: universal healthcare, reducing poverty, and improving social welfare. Sanders has frequently…..

You’re focusing on the similarity in the message, but the tweet does not specifically support Sanders or his policies. It expresses concern about societal failures and the lack of basic healthcare or support systems, but …...

Throughout the debate, both sides presented valid arguments. However, the content of the tweet clearly resonates with Bernie Sanders’ policies, particularly in terms of addressing social justice, healthcare, and poverty. The language used aligns closely with Sanders' platform, even though the tweet does not explicitly name him. Given the strong alignment with Sanders' key issues, I rule that the tweet is supportive of Bernie.
Conclusion: The tweet’s stance on the target is ‘Favor’.

Figure 1. The framework of our ZSMD. Based on the provided tweet, target, and supplemented background knowledge, both debater agents first present their respective viewpoints. In the subsequent stage, they engage in a multi-round debate to challenge and refine each other’s arguments. If, upon reaching the maximum number of debate rounds (denoted as $J = R$), neither agent adjusts its stance to align with the other, a judge agent intervenes to make the final stance determination based on the entire debate process.

3.1. Initialization Stage

In the initialization stage, we first perform data augmentation, after which each debater agent generates its respective initial arguments to prepare for the debate stage.

3.1.1. Stance Separation

First, we establish two debater agents, D_F and D_A, with clearly defined stances. The debater agent D_F asserts that the tweet expresses a supportive stance toward the target, while the debater agent D_A contends that the tweet conveys an opposing stance toward the target. The following is an example of a prompt for the debater agent D_F.

> Prompt 1: Stance Separation
>
> You are a debate expert, and you believe that the following tweet expresses a supportive stance toward the target. You will now engage in a debate with another expert who believes the tweet is against the target.

3.1.2. Knowledge Augmentation

To address the issue of insufficient domain-specific knowledge, we employ GPT-4o as an external knowledge base to perform targeted data augmentation for each instance. The overall procedure consists of three steps:

(1) Retrieval via prompting

For each tweet–target pair (x_i, t_i), we send a structured prompt (prompt 2) to GPT-4o, asking it to return concise, relevant background facts—such as key events, policies, or historical actions—related to the target and mentioned themes.

(2) Filtering and validation

The raw GPT-4o output is then passed through a lightweight filter to remove generic or irrelevant responses. Specifically, we perform the following:

- Discard any empty or “no relevant information” replies.
- Remove boilerplate phrases (e.g., “As an AI language model…”).
- Enforce a maximum length of 50 tokens to keep K_i focused and prevent hallucination.

(3) Integration into the debate pipeline

The filtered background knowledge K_i is then concatenated with the original tweet and target and fed into each debater agent. Concretely, the augmentation process is formalized as

$$K_i = Filter(GPT4oPrompt(x_i, t_i)) \tag{1}$$

where $GPT4oPrompt(\cdot)$ denotes the call to GPT-4o using prompt 2, and Filter$(\cdot)$ applies the above filtering rules. By explicitly retrieving, filtering, and integrating only the most pertinent background facts, we ensure that each debater agent benefits from concise, accurate, and task-relevant context, thereby improving argument quality and reasoning depth.

The following is an example of a prompt for knowledge augmentation.

> Prompt 2: Knowledge Augment (P-Stance)
>
> Please provide concise and relevant background information based on the tweet and target. Focus on specific facts such as key events, policies, beliefs, or actions associated with the target that directly relate to the issues or themes mentioned in the tweet. Avoid general descriptions. Limit the response to essential information that could help clarify the stance expressed in the tweet.

Tweet: “{tweet}”
Target: “{target}”

3.1.3. Initial Argument Generation

Following data augmentation, the debater agents—namely, the supporting agent D_F and the opposing agent D_A—generate preliminary arguments. The supporting agent posits that the tweet expresses a supportive stance toward the target, while the opposing agent asserts that the tweet conveys an opposing stance. The formulas for generating the initial arguments for each agent are as follows:

$$h_F^0 = LLM(p_{init}, x_i, t_i, K_i, Favor) \tag{2}$$

$$h_A^0 = LLM(p_{init}, x_i, t_i, K_i, Against) \tag{3}$$

Here , h_F^0 and h_A^0 denote the preliminary arguments generated by the supporting agent and the opposing agent, respectively, during the initialization stage. *LLM* refers to the large language model (e.g., DeepSeek-v3), p_{init} represents the prompt used to generate the initial arguments, and *Favor* and *Against* indicate the stance directives for the supporting and opposing sides, respectively. Through these steps, the supporting and opposing agents produce their respective preliminary arguments and are prepared to enter the debate stage. The following is an example of a prompt for the debater agent D_F.

Prompt 3: Initial Argument Generation (P-Stance)

Please generate your initial arguments. Your arguments should:
Align with the target’s known views, policies, or actions: Show how the content of the tweet supports or resonates with the target’s public stance, ideologies, or actions.
Highlight connections: Point out specific beliefs, proposals, or statements made by the target that directly correspond to the issues or concerns raised in the tweet.
Use evidence: Provide clear reasoning, examples, or quotes that demonstrate how the tweet reflects the target’s values and goals, reinforcing the idea that the tweet is indeed supportive of the target.
Focus on showing how the tweet positively connects with the target, ensuring your argument is strong and well supported.

Tweet: “{tweet}”
Target: “{target}”
Background knowledge: “{background_knowledge}”

3.2. Debate Stage

During the debate stage, the supporting and opposing agents exchange their viewpoints through multiple rounds of debate and formulate rebuttals based on each other’s arguments. The core of the debate hinges on the following key aspects.

3.2.1. Refutation Mechanism

The rebuttal mechanism forms the core of the debate, where agents employ distinct strategies to challenge each other’s arguments. As shown in Figure 2, this mechanism involves three core strategies: factual verification, logical analysis, and sentiment analysis.

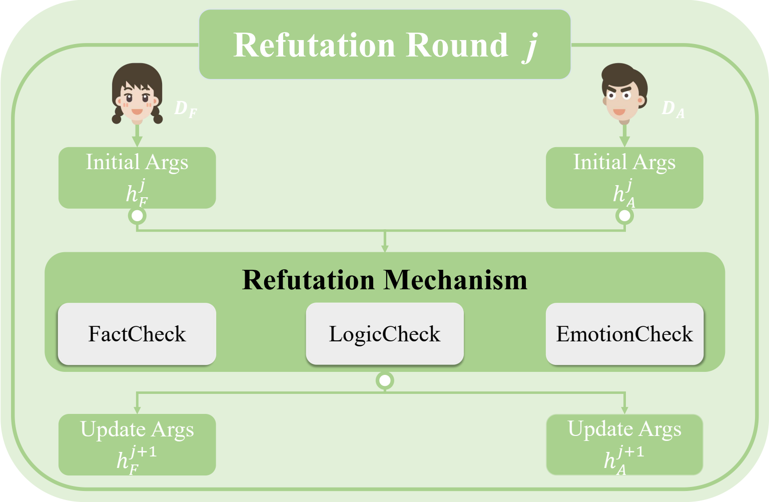


Figure 2. Refutation mechanism in multi-agent debate framework. This diagram illustrates how, in each debate round j, the two agents exchange and update their arguments via three refutation strategies—FactCheck (Equation (4)), LogicCheck (Equation (5)), and EmotionCheck (Equation (6))—and terminate the debate either upon reaching consensus or upon completing the maximum number of rounds.

Factual verification: The agent examines whether the facts presented in the opponent's arguments are accurate, investigating the presence of any factual errors. The specific formula is as follows:

$$FactCheck(h_F^j, h_A^j) = LLM(FactValidation, h_F^j, h_A^j) \tag{4}$$

Logical analysis: The agent evaluates whether the opponent's reasoning is logically sound, identifying any logical flaws or unreasonable inferences. The formula is as follows:

$$LogicCheck(h_F^j, h_A^j) = LLM(LogicValidation, h_F^j, h_A^j) \tag{5}$$

Emotional analysis: The agent assesses the emotional tendencies within the opponent's arguments, determining whether emotionally charged reasoning is present and whether it impacts the objectivity of the inference. The formula is as follows:

$$EmotionCheck(h_F^j, h_A^j) = LLM(EmotionValidation, h_F^j, h_A^j) \tag{6}$$

The following is an example of a prompt for the debater agent D_F.

Prompt 4: Debate Mechanism (P-Stance)

As you present your arguments and engage in rebuttal, please keep the following strategies in mind for challenging your opponent's stance.
Factual verification: Pay attention to any factual claims made by your opponent. Check if the facts they present are accurate or whether there are any errors. If you find a factual discrepancy or incorrect statement, correct it with verified facts that support your stance.
Logical analysis: Analyze the logic behind your opponent's arguments. Look for logical fallacies, contradictions, or weak reasoning that could undermine their stance. Point out any flaws in their reasoning and present a more coherent, logically sound argument that supports your stance.

Emotional analysis: Evaluate the emotional tone of your opponent's argument. If their stance relies heavily on emotional appeal or biased sentiment, highlight this. Show how their emotional reaction may cloud their judgment and distract from a logical analysis of the tweet's true intent.
As you debate, use these strategies to challenge your opponent's arguments effectively, ensuring that your stance—that the tweet is supportive of the target—is well defended. Be respectful in your tone, back up your claims with facts, and ensure your arguments are logically sound.

Against arguments: {con_args}

3.2.2. Agent Update

The debate does not conclude in a single round but progresses iteratively through multiple rounds. In each round of debate, the supporting and opposing agents update their arguments—and potentially their stances—based on the rebuttals provided by their counterparts. The update formulas for the agents after each debate round are as follows:

$$h_F^j = LLM(p_{debate}, \{h_A^0, h_A^1, \ldots, h_A^{j-1}\}, \{h_F^0, h_F^1, \ldots, h_F^{j-1}\}) \tag{7}$$

$$h_A^j = LLM(p_{debate}, \{h_F^0, h_F^1, \ldots, h_F^{j-1}\}, \{h_A^0, h_A^1, \ldots, h_A^{j-1}\}) \tag{8}$$

Here, p_{debate} represents the debate prompt, and h_F^j and h_A^j represent the updated arguments of the opposing debater agent and the supporting debater agent in the j-th round, respectively.

To prevent the debate from continuing indefinitely, we establish a maximum number of debate rounds, denoted as R. If, upon reaching the maximum number of rounds, the supporting and opposing agents have not reached a consensus, the system transitions to the adjudication stage, where a referee agent makes the final stance determination.

$$\text{If } j = R, \text{then procced to Judge} \tag{9}$$

The following is an example of a prompt for the Judge agent.

Prompt 5: Stance Determination (P-Stance)

Please maintain an objective and neutral stance, comprehensively weighing all information based on the provided original tweet, target, supplementary background knowledge, and the arguments and rebuttals presented by both sides during the debate (covering factual verification, logical analysis, and sentiment analysis). Identify the factual accuracy, logical rigor, and emotional expression biases in the arguments, and accordingly provide a clear stance determination result (support or opposition), while concisely explaining the key decision basis and points of contention.

Tweet: "{tweet}"
Target: "{target}"
Background knowledge: "{background_knowledge}"
Debate history: "{debate_history}"

The detailed process is presented in Algorithm 1. Additionally, Table A1 in Appendix A provides a complete example to illustrate this workflow.

Algorithm 1 ZSMD

```
Require: Agents $D_F$, $D_A$, $J$, Rounds $R$, $C = \{x_i, t_i\}_{i=1}^{n}$
Ensure: Final prediction $\hat{y}$
procedure
    Stage1:
    //Knowledge Augment
    for $(x_i, t_i)$ in $C$ do
        $K_i \leftarrow (x_i, t_i)$ by LLM
    end for
    //Initial Argument Generation
    $h_F^0 \leftarrow$ Initialize $D_F$ with $(p_{init}, x_i, t_i, K_i)$ by LLM
    $h_A^0 \leftarrow$ Initialize $D_A$ with $(p_{init}, x_i, t_i, K_i)$ by LLM
    Stage2:
    for $j$ = 1 to $R$ do
        $Q_F^j \leftarrow$ FactCheck, LogicCheck, EmotionCheck($h_A^{j-1}$) by LLM
        $Q_A^j \leftarrow$ FactCheck, LogicCheck, EmotionCheck($h_F^{j-1}$) by LLM
        $D_F \leftarrow$ Update the status of $D_F$ based on $Q_F^j$ and $h_A^{j-1}$
        $D_A \leftarrow$ Update the status of $D_A$ based on $Q_A^j$ and $h_F^{j-1}$
        $h_F^j \leftarrow$ Update by $D_F$
        $h_A^j \leftarrow$ Update by $D_A$
    end for
    if $h_F$ = $h_A$ then
        return $h_F$
    else
        return $J's$ determination
    end if
end procedure
```

4. Experiment

To validate the effectiveness of our proposed approach, we conducted experiments on two publicly available stance detection datasets. In this section, we sequentially introduce the datasets used in the experiments, the experimental setup, the baseline models for comparison, and an analysis of our experimental results.

4.1. Dataset and Evaluation Metrics

(1) SemEval-2016. SemEval-2016 Task 6 is a widely utilized benchmark dataset for stance detection in social media text. The objective of this dataset is to identify the stance expressed in a given tweet toward a specific target. The dataset comprises 4870 tweets, covering five distinct targets: Hillary Clinton (HC), feminist movement (FM), legalization of abortion (LA), atheism (A), and climate change is a real concern (CC). Each tweet is annotated with one of three stances: Favor, against, or neutral. This task is formulated as a multi-class classification problem, aiming to determine whether the attitude expressed in a tweet toward its target is supportive, opposing, or neutral. The dataset is designed to provide a diverse range of stance expressions, facilitating the training of stance detection models, particularly in the context of social media, where text is typically brief, informal, and may contain sarcasm or ambiguous phrasing.

(2) P-Stance. The P-Stance dataset is primarily designed for stance detection concerning political targets, specifically focusing on social media text related to the 2020 U.S. presidential election. This dataset comprises 21,574 tweets and covers three political figures: Donald Trump (DT), Joe Biden (JB), and Bernie Sanders (BS). It is predominantly structured as a binary classification task, with the objective of predicting whether a tweet expresses support for or opposition to these political figures. Compared to other stance detection

datasets, P-Stance presents unique challenges due to the more subtle nature of expressions in the political domain. Tweets often feature indirect references to political figures, necessitating nuanced contextual analysis to discern stances. This may involve identifying oblique references or broader political discourse associated with these individuals. Consequently, the P-Stance dataset is particularly valuable for training stance detection models capable of handling implicit expressions and subtle sentiments.

To ensure fairness in evaluation, we adopt the micro-average F1 score as our primary evaluation metric, as recommended by the creators of the datasets. The specific calculation method for this evaluation metric is as follows:

$$F_{\mathrm{avg}} = \frac{F_{favor} + F_{against}}{2} \tag{10}$$

The calculation methods for F_{favor} and $F_{against}$ are as follows:

$$F_{favor} = \frac{2P_{favor}R_{favor}}{P_{favor} + R_{favor}} \tag{11}$$

$$F_{against} = \frac{2P_{against}R_{against}}{P_{against} + R_{against}} \tag{12}$$

where P and R represent precision and recall, respectively.

4.2. Compared Methods

(1) JointCL [31]. This model designs a new joint contrastive learning framework that enhances the model's ability to detect stances on unknown targets by constructing positive and negative samples for contrastive learning.

(2) KASD [32]. This model enhances stance detection by integrating two types of background knowledge: episodic knowledge (retrieved and filtered from Wikipedia via GPT-3.5 Turbo prompts) and discourse knowledge (extracted from hashtags and references). These knowledge sources are injected into the input to provide richer context for stance inference.

(3) KAI [33]. This work proposes a zero-shot stance detection framework that leverages target-independent transferable knowledge derived from large language models (LLM-KE). This knowledge is integrated into the prediction process using a bidirectional knowledge-guided neural production system.

(4) MB-Cal [34]. This model designs a calibration network to mitigate potential biases of the large language model (LLM) in stance detection. Additionally, it constructs counterfactual-augmented data to address the challenges of effectively learning bias representations and the difficulties associated with debiased generalization.

(5) KEL-CA [35]. This model incorporates external knowledge through a dual injection mechanism, combining information from a Wikipedia-based local knowledge base with a chain-of-thought reasoning process. Additionally, it employs contrastive learning and adversarial domain transfer to enhance feature generalization. We include KEL-CA as a strong and up-to-date baseline due to its effective use of external knowledge and its competitive performance.

(6) COLA [36]. The model proposes a collaborative role-injection framework involving multiple large language models (LLMs), wherein the LLMs are assigned distinct roles. This framework consists of three stages: the multidimensional text analysis stage, the reasoning-enhanced debating stage, and the stance conclusion stage. The framework does not require additional labeled data, relying solely on interactions with pre-trained LLMs, which renders it highly practical and accessible.

4.3. Implementation Details

MDSD is conducted under a zero-shot setting, requiring no training dataset. We utilized GPT-3.5 Turbo and DeepSeek-v3 to perform our experiments. To ensure the stability of the debate process, we set the temperature to 0.5 during the experiments. In our experiments, we primarily focused on a dual-agent debate scenario, thus configuring two debater agents and one referee agent, with the maximum number of debate rounds set to two. For the knowledge enhancement module, we employed GPT-4o as the external knowledge base.

In the SemEval-2016 Task 6 dataset, stance labels include favor, against, and none. To address this, after supplementing background knowledge, we introduce an expert agent tasked with determining whether the stance of a tweet toward the target is none. If the expert agent classifies the tweet's stance as none, the corresponding label is directly assigned; otherwise, the tweet proceeds to the subsequent multi-agent debate process for further determination of its specific stance. The specific examples are shown in Table 1.

Table 1. A case study on instances with a true stance of 'none' in the SemEval-2016 Task 6 dataset. We established a 'none' expert agent, whose task is to first determine, prior to the debate, whether the stance of a tweet toward the target is none.

Tweet: I do not know how well a socialist is gonna play with southern democrats. #BernieSanders Target: Hillary Clinton Background Knowledge: Bernie Sanders is a self-proclaimed democratic socialist, and his policies often emphasize wealth redistribution, universal healthcare, and tuition-free public college. Southern Democrats have historically been more moderate or conservative compared to the broader Democratic Party. The term "Southern Democrats" traditionally refers to Democratic voters in the U.S. South, where political preferences can differ from those in more liberal regions. Hillary Clinton, as a key figure in the Democratic Party, has positioned herself as a more centrist candidate compared to Sanders.
Reasoning: The tweet focuses on Bernie Sanders and his potential reception among Southern Democrats. It does not make any explicit statement supporting or opposing Hillary Clinton. Therefore, the stance toward the target should be classified as 'none'.

4.4. Results

In Table 2, we present the zero-shot stance detection performance of ZSMD on the SemEval-2016 Task 6 and P-Stance benchmark datasets, compared to various baseline models. The results demonstrate that ZSMD, whether paired with GPT-3.5 Turbo or DeepSeek-v3, consistently outperforms all baseline models on both the Sem16 and P-Stance datasets, highlighting its robustness and applicability in zero-shot stance detection tasks.

Specifically, using DeepSeek-v3 as an example, ZSMD achieves absolute average score improvements of 4.64% and 3.29% over the best baseline, MBCal-ChatGPT, on the Sem16 and P-Stance datasets, respectively. Additionally, at the sub-target level (e.g., different topics in Sem16 and candidates in P-Stance) ZSMD exhibits strong consistency and generalization, indicating its effectiveness in handling diverse targets. Furthermore, when combined with DeepSeek-v3, ZSMD achieves optimal performance, validating the successful integration of multi-agent debate with zero-shot detection strategies and demonstrating that the knowledge enhancement mechanism further improves reasoning and judgment accuracy atop large-scale models.

Table 2. Comparison of ZSMD and baselines in zero-shot stance detection task. Bold and underlined values refer to the best and second-best performance. For clarity, the target abbreviations used in the table are Hillary Clinton (HC), feminist movement (FM), legalization of abortion (LA), atheism (A), and climate change is a real concern (CC).

	Sem16 (%)						**P-Stance (%)**			
	HC	**FM**	**LA**	**A**	**CC**	**Avg**	**Biden**	**Sanders**	**Trump**	**Avg**
Baseline Models										
JointCL	54.80	53.80	49.50	54.50	39.70	50.46	-	-	-	-
KASD-LLaMA2	77.70	65.57	57.07	39.55	50.72	50.98	75.28	74.09	69.27	72.88
KASD-ChatGPT	80.32	70.41	62.71	63.95	55.83	66.64	83.60	79.66	84.31	82.52
KAI	76.40	73.70	69.40	-	-	-	85.70	80.50	75.90	80.70
MBCal-LLaMA2	77.19	74.71	72.49	58.29	67.71	70.08	84.04	81.22	77.57	80.94
MBCal-ChatGPT	80.26	75.76	68.77	66.54	**71.00**	72.47	85.14	81.05	85.08	83.76
KEL-CA	81.70	72.30	**73.30**	68.30	65.70	72.26	-	-	-	-
COLA	81.70	63.40	71.00	70.80	67.50	70.88	84.00	79.70	86.60	83.43
Our Models										
GPT3.5+ZSMD	82.20	73.82	72.71	70.94	68.43	73.62	85.91	81.10	85.26	84.09
DeepSeek-v3+ZSMD	**86.32**	**76.43**	73.08	**72.88**	70.42	**75.83**	**87.67**	**83.71**	**88.19**	**86.52**

In summary, the experimental results clearly affirm ZSMD's significant advantages and scalability in stance detection tasks, laying a solid foundation for future research in more complex scenarios.

5. Discussion

5.1. Ablation Study

We conducted ablation experiments using DeepSeek-v3 as the baseline model on both the SemEval-2016 Task 6 and P-Stance datasets to investigate the impact of each module in the ZSMD framework, while evaluating the performance of ZSMD when individual modules are removed. The results are presented in Table 3.

Table 3. Experimental results of ablation study.

	Sem16 (%)					**P-Stance (%)**		
	HC	**FM**	**LA**	**A**	**CC**	**Biden**	**Sanders**	**Trump**
DeepSeek-v3+ZSMD	86.32	76.43	73.08	72.88	70.42	87.67	83.71	88.19
w/o KA	83.16	74.85	72.10	71.47	68.55	85.36	82.02	86.40
w/o DM	82.47	72.63	70.40	70.28	67.91	82.06	76.90	86.15
w/o Debate	80.76	71.08	68.92	67.17	65.36	80.88	71.62	69.36

The variants of ZSMD are as follows:

- "w/o KA": The knowledge augmentation module is removed, meaning no external knowledge is supplemented for the tweets or targets. The debater agents rely solely on the original tweet content to formulate arguments and engage in debate. This configuration assesses the impact of background knowledge on stance detection performance.
- "w/o DM": The debate mechanism is eliminated, and only the initially generated arguments are used as the basis for the final stance determination, without undergoing multiple rounds of argumentation or viewpoint updates. This setup evaluates the role of the debate mechanism in refining viewpoints and enhancing reasoning capabilities.
- "w/o Debate": The entire debate process is completely removed, and the stance determination is made directly by the referee agent based solely on the initial arguments of the agents, without any iterative debate. This tests the contribution of the full debate process to the ultimate stance classification task.

In Figure 3, we systematically evaluate the impact of the number of debate rounds on the experimental results using DeepSeek-v3. The results show that when the number of debate rounds does not exceed two, model performance improves significantly with additional rounds, suggesting that agents effectively extract and leverage information from the tweet and background knowledge during the initial interactions. However, when the number of rounds exceeds two, performance on both datasets begins to exhibit slight fluctuations and eventually stabilizes at a level comparable to that of two rounds. This phenomenon may indicate that, after two thorough rounds of debate, the available information in the tweet and background knowledge has been fully exploited, with additional rounds yielding diminishing marginal returns on performance improvement.

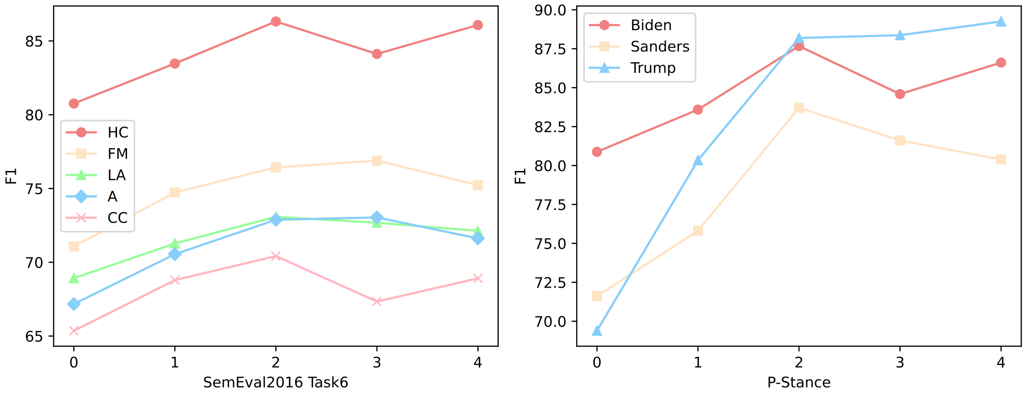


Figure 3. Comparison of different numbers of debate rounds in our proposed ZSMD approach using DeepSeek-v3.

5.2. Error Analysis

Based on the experimental results, we conducted a detailed error analysis of the ZSMD framework. Figure 4 summarizes the distribution of classification errors made by DeepSeek-v3 on the SemEval-2016 Task 6 dataset, categorized by underlying causes. We categorized the primary sources of errors into five types, as follows:

(1) Insufficient or biased background knowledge supplementation: Due to the reliance on GPT-4o for data augmentation, the supplemented background knowledge may not always be entirely accurate or sufficiently comprehensive. This can lead to agents debating based on inadequate information, thereby affecting the final stance determination.

(2) Complex sentence structures and implicit stance expressions: Certain tweets express stances in subtle ways, such as through sarcasm, double negatives, or metaphors, making it challenging for agents to accurately identify the stance.

(3) Interference from language style and social media noise: Tweets may contain spelling errors, slang, abbreviations, or internet memes, which hinder the agents' ability to precisely interpret the true meaning of the text.

(4) Ambiguity in stance judgment (subjectivity): Some tweets may lack a clear stance, or their interpretation may vary depending on the reader, introducing subjectivity that complicates stance detection.

(5) Insufficient logical reasoning capability: During reasoning, the model may struggle to form a coherent and robust chain, leading to weak arguments. Supporting and opposing agents might list facts or opinions independently without logical progression, limiting the referee agent's ability to determine an accurate stance.

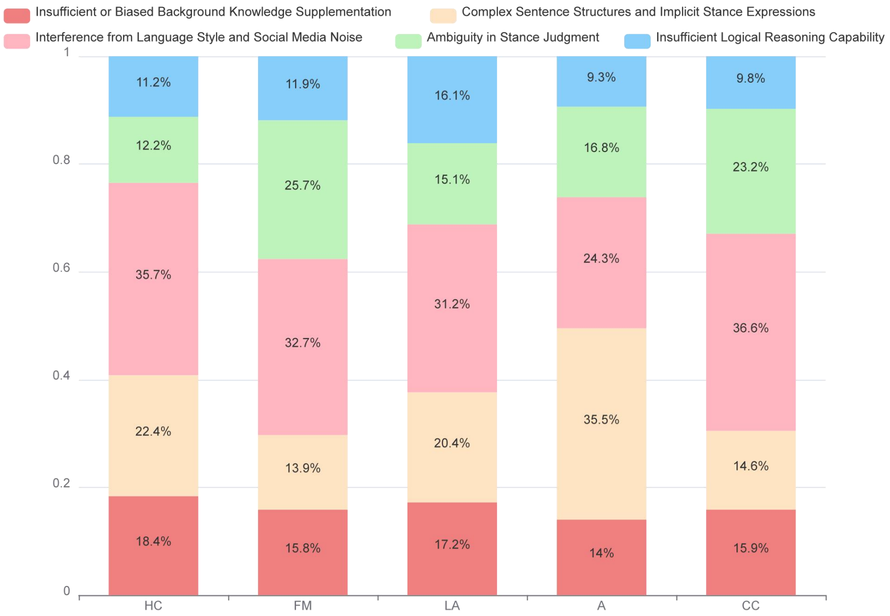


Figure 4. Error analysis of DeepSeek-v3 on SemEval-2016.

To further illustrate the identified error types, we present two representative misclassified examples in Table 4. These error categories highlight the challenges faced by the ZSMD framework and provide insights into areas for potential improvement.

Table 4. Representative error cases in zero-shot stance detection on SemEval-2016 Task 6. This table presents two typical examples of stance misclassification, each corresponding to a distinct source of error: (1) implicit stance expressions, which obscure the true intent of the text, and (2) insufficient logical reasoning by the model during argument generation.

Tweet and Background Knowledge	Target	Stance	Predict	Reason for the Error
Tweet: Where is the childcare program @joanburton which you said would be in place? Background knowledge: The tweet mentions Joan Burton, an Irish Labour Party politician who promised childcare support for single-parent families. "#7istooyoung" is a campaign hashtag opposing the removal of single mothers with children over 7 years old from the welfare system. The tweet focuses on social support policies for single parents but does not directly address the topic of abortion legalization.	Legalization of Abortion	Against	None	Implicit stance expressions
Tweet: @Deb_Hitchens @JudgeBambi. There's no capacity for discourse when you assume ppl are your enemy. Hate. Misery. Paranoia. A waste. Background knowledge: The feminist movement focuses on gender equality, emphasizing women's rights across various domains. Some critics argue that the positions of certain feminists are overly extreme, overlooking the possibility of building constructive dialogue with others, and instead fostering hostility and unnecessary confrontation.	Feminist Movement	None	Against	Insufficient logical reasoning capability

6. Conclusions

In this paper, we propose a novel stance detection method based on multi-agent debate, termed ZSMD (zero-shot stance detection based on multi-agent debate). This approach introduces two debater agents representing opposing stances (supporting and opposing), which generate initial arguments under the enhancement of background knowledge. Through multiple rounds of debate, these agents iteratively refine their viewpoints, culminating in a final stance determination by a referee agent based on the debate process. The experimental results demonstrate that ZSMD achieves superior performance on the classic stance detection datasets SemEval-2016 and P-Stance, while also exhibiting strong generalization capabilities.

In future work, we aim to further enhance the reasoning and debating abilities of the agents to address more complex stance expressions. Additionally, we plan to explore more effective knowledge augmentation strategies, such as integrating large language models (LLMs) for adaptive knowledge retrieval, to improve the relevance and accuracy of background knowledge. Furthermore, we are interested in incorporating multimodal information (e.g., images or videos) to improve the model's understanding of rich social media content. However, we acknowledge that the lack of publicly available, well-annotated multimodal stance detection datasets, as well as the increased model complexity and training cost, present significant challenges. We consider addressing these limitations as important directions for future research.

Author Contributions: Conceptualization, C.W.; data curation, J.M. and C.W.; formal analysis, L.R. and Y.X.; funding acquisition, Y.X.; investigation, C.W. and B.W.; methodology, J.M., L.R. and Y.X.; project administration, B.W.; resources, J.M. and B.W.; software, J.M. and C.W.; validation, J.M.; writing—original draft, J.M. and C.W.; writing—review and editing, L.R. and Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: The research was supported by the National Natural Science Foundation of China under Grant No. 62102372, the Project of Science and Technology in Henan Province under Grant Nos. 252102211072 and 232102210078, the Doctor Scientific Research Fund of Zhengzhou University of Light Industry under Grant No. 2021BSJJ029, and the School-Level Young Backbone Teacher Training Program of Zhengzhou University of Light Industry.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The P-Stance dataset is available at https://github.com/chuchun8/PStance, accessed on 20 March 2025. The SemEval-2016 Task 6 dataset is available at https://alt.qcri.org/semeval2016/task6/, accessed on 20 March 2025.

Acknowledgments: We thank all anonymous reviewers for their constructive comments.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

In the appendix, we provide an example table that details the complete processing workflow for a single sample from the SemEval-2016 Task 6 dataset. This example covers the entire process, starting from the input of the original tweet and target, through background knowledge enhancement, the multi-agent debate process, the application of rebuttal mechanisms, agent updates, and culminating in the final stance determination. Through this example, we aim to visually illustrate the specific execution details of the proposed method, facilitating a clearer understanding for readers of how the ZSMD framework operates on real data.

Table A1. An example of the complete processing workflow for instances with a true stance of ‘against’ in the SemEval-2016 Task 6 dataset.

Tweet: If Feminism wants equal rights as men they must so advocate for equal prison time and losing custody rights, which they won’t do!
Target: Feminist Movement
Background knowledge: Feminism is a social and political movement advocating for gender equality, addressing issues such as workplace rights, reproductive rights, and gender-based violence. Some critics argue that feminists do not always advocate for equality in all areas, particularly in legal matters such as prison sentencing and child custody decisions. These critiques often stem from broader debates about whether feminism seeks true equality or prioritizes certain gender-based issues over others.
Reasoning: The tweet makes a claim about the feminist movement’s actions and intentions, implying a stance rather than neutrality. Since it is making a judgment rather than being neutral or purely informational, its stance toward the target is **not** classified as ‘None’.
Initial argument generation literal meaning of the tweet: While the tweet criticizes feminism for its focus on certain areas, it indirectly supports feminism. The tweet implies that feminism should address broader issues of legal and social equality, such as incarceration periods and custody rights, suggesting that it is, in fact, a call for feminism to pursue more comprehensive attention and deeper equality. Social justice issues expressed in the tweet: The tweet highlights issues like incarceration periods and custody rights, and although it critiques feminism’s selective focus, it actually underscores existing gender inequalities in society. Advocates of feminism should prioritize these concerns and push for reform. Thus, the tweet encourages feminism to extend its influence into a wider range of domains, beyond its traditional focal points. Background support: Background knowledge indicates that some critics argue the feminist movement has occasionally failed to address equality across all domains. The tweet echoes this critique, thereby indirectly supporting feminism’s potential to promote equality in additional areas.
Literal criticism in the tweet: The tweet explicitly criticizes feminism, asserting that if feminism demands equality, it should advocate for equal incarceration periods and custody rights. In essence, it expresses dissatisfaction with feminism’s incomplete focus on certain domains (e.g., legal and family issues). It suggests that feminism has not genuinely pursued equality across all areas, particularly in relation to men’s rights. Criticism of feminism’s selective focus: The tweet exposes a bias in feminism’s pursuit of rights, highlighting that the rights it champions often prioritize women’s interests while overlooking equal rights for men in certain contexts. Thus, the tweet effectively challenges feminism’s “selective” approach to equality, critiquing its failure to address all gender-related issues equitably. Background support: According to the background knowledge, critics argue that feminism has not succeeded in promoting comprehensive equality in some legal domains (e.g., incarceration and custody rights). The tweet reflects this perspective, questioning whether feminism truly seeks gender equality or merely focuses on women’s issues. Consequently, the tweet opposes feminism.
Multi-agent debate: Although the tweet criticizes feminism for not promoting equality across all domains, it effectively calls for feminism to broaden its influence and address additional gender equality issues. By mentioning incarceration periods and custody rights, the tweet highlights areas where feminism could expand its focus, essentially advocating for a more comprehensive pursuit of gender equality under the feminist framework. You overlook the core of the tweet’s criticism. The tweet does not support feminism; rather, it critiques its failure to be fully inclusive in addressing equality. It points out that feminism tends to neglect equal treatment for men in certain legal matters, thereby questioning its stance and suggesting that it does not promote equality across all domains, especially concerning men. The tweet is not entirely a criticism of feminism; it reveals areas where feminism might have overlooked equality demands. In essence, it serves as a call to advance gender equality, urging us to recognize that feminism can extend its scope to advocate for broader equality policies, including incarceration periods and custody rights. Thus, it actually supports a more comprehensive feminist stance. Your argument disregards the tweet’s fundamental critique of feminism. The tweet does not suggest that feminism should address more issues; instead, it highlights feminism’s limitations, particularly in neglecting men’s rights. Its core lies in criticizing feminism’s selective focus, explicitly asserting that feminism fails to pursue equality for all genders when advocating for equitable treatment.
Stance determination: During the debate, the supporting side argued that while the tweet criticizes feminism, it implicitly calls for feminism to address equality in more domains. Conversely, the opposing side contended that the tweet explicitly critiques feminism’s bias, questioning whether it genuinely promotes comprehensive gender equality. The supporting side’s argument fails to fully capture the critical nature of the tweet, whereas the opposing side clearly highlights the tweet’s expression of dissatisfaction with feminism’s incomplete focus in certain areas. Therefore, I determine that the tweet’s stance toward feminism is one of ’against’.

References

1. Mohammad, S.; Kiritchenko, S.; Sobhani, P.; Zhu, X.; Cherry, C. Semeval-2016 task 6: Detecting stance in tweets. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), San Diego, CA, USA, 16–17 June 2016; pp. 31–41.
2. Ma, J.; Wang, C.; Xing, H.; Zhao, D.; Zhang, Y. Chain of stance: Stance detection with large language models. In Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing, Hangzhou, China, 1–3 November 2024; Springer: Singapore, 2024; pp. 82–94.

3. Wang, X.; Wang, Y.; Cheng, S.; Li, P.; Liu, Y. Deem: Dynamic experienced expert modeling for stance detection. *arXiv* **2024**, arXiv:2402.15264.
4. Augenstein, I.; Rocktäschel, T.; Vlachos, A.; Bontcheva, K. Stance detection with bidirectional conditional encoding. *arXiv* **2016**, arXiv:1606.05464.
5. Allaway, E.; McKeown, K. Zero-shot stance detection: A dataset and model using generalized topic representations. *arXiv* **2020**, arXiv:2010.03640.
6. Zhao, X.; Zou, J.; Xie, F.; Wang, H.; Wu, H.; Zhou, B.; Tian, J. A unified framework for unseen target stance detection based on feature enhancement via graph contrastive learning. In Proceedings of the Annual Meeting of the Cognitive Science Society, Sydney, NSW, Australia, 26–29 July 2023; Volume 45.
7. Zhang, Z.; Xie, F.; Zhao, X.; Zhou, B.; Chen, J.; Tian, L. Multitask learning neural networks for pandemic prediction with public stance enhancement. In Proceedings of the 2022 IEEE 34th International Conference on Tools with Artificial Intelligence (ICTAI), Macao, China, 31 October–2 November 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1262–1270.
8. Fan, Q.; Lu, J.; Sun, Y.; Pi, Q.; Shang, S. Enhancing zero-shot stance detection via multi-task fine-tuning with debate data and knowledge augmentation. *Complex Intell. Syst.* **2025**, *11*, 151. [CrossRef]
9. Allaway, E.; Srikanth, M.; McKeown, K. Adversarial learning for zero-shot stance detection on social media. *arXiv* **2021**, arXiv:2105.06603.
10. Hardalov, M.; Arora, A.; Nakov, P.; Augenstein, I. Few-shot cross-lingual stance detection with sentiment-based pre-training. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 22 February–1 March 2022; Volume 36, pp. 10729–10737.
11. Zhao, X.; Zou, J.; Miao, J.; Tian, L.; Gao, L.; Zhou, B.; Pang, S. Zero-shot stance detection based on multi-perspective transferable feature fusion. *Inf. Fusion* **2024**, *108*, 102386. [CrossRef]
12. Zhang, H.; Zhang, X.; Huang, H.; Yu, L. Prompt-based meta-learning for few-shot text classification. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Abu Dhabi, United Arab Emirates, 7–11 December 2022; pp. 1342–1357.
13. Liu, R.; Lin, Z.; Tan, Y.; Wang, W. Enhancing zero-shot and few-shot stance detection with commonsense knowledge graph. In Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Online Event, 1–6 August 2021; pp. 3152–3157.
14. Yan, M.; Joey, T.Z.; Ivor, W.T. Collaborative knowledge infusion for low-resource stance detection. *Big Data Min. Anal.* **2024**, *7*, 682–698. [CrossRef]
15. Wang, C.; Zhang, Y.; Wang, S. A meta-contrastive learning with data augmentation framework for zero-shot stance detection. *Expert Syst. Appl.* **2024**, *250*, 123956. [CrossRef]
16. Schick, T.; Schütze, H. Exploiting cloze questions for few shot text classification and natural language inference. *arXiv* **2020**, arXiv:2001.07676.
17. Wang, C.; Zhang, Y.; Yu, X.; Liu, G.; Chen, F.; Lin, H. Adversarial network with external knowledge for zero-shot stance detection. In Proceedings of the China National Conference on Chinese Computational Linguistics, Harbin, China, 3–5 August 2023; Springer: Singapore, 2023; pp. 419–433.
18. Taranukhin, M.; Shwartz, V.; Milios, E. Stance reasoner: Zero-shot stance detection on social media with explicit reasoning. *arXiv* **2024**, arXiv:2403.14895.
19. Zhang, H.; Li, Y.; Zhu, T.; Li, C. Commonsense-based adversarial learning framework for zero-shot stance detection. *Neurocomputing* **2024**, *563*, 126943. [CrossRef]
20. Wen, H.; Hauptmann, A.G. 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), Toronto, ON, Canada, 9–14 July 2023; pp. 1491–1499.
21. Huang, H.; Zhang, B.; Li, Y.; Zhang, B.; Sun, Y.; Luo, C.; Peng, C. Knowledge-enhanced prompt-tuning for stance detection. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2023**, *22*, 1–20. [CrossRef]
22. He, Z.; Mokhberian, N.; Lerman, K. Infusing knowledge from wikipedia to enhance stance detection. *arXiv* **2022**, arXiv:2204.03839.
23. Ding, D.; Dong, L.; Huang, Z.; Xu, G.; Huang, X.; Liu, B.; Jing, L.; Zhang, B. Edda: A encoder-decoder data augmentation framework for zero-shot stance detection. *arXiv* **2024**, arXiv:2403.15715.
24. Zhang, M.; Gong, H.; Liu, Q.; Wu, S.; Wang, L. Breaking Event Rumor Detection via Stance-Separated Multi-Agent Debate. *arXiv* **2024**, arXiv:2412.04859.
25. Chan, C.M.; Chen, W.; Su, Y.; Yu, J.; Xue, W.; Zhang, S.; Fu, J.; Liu, Z. Chateval: Towards better llm-based evaluators through multi-agent debate. *arXiv* **2023**, arXiv:2308.07201.
26. Liang, T.; He, Z.; Jiao, W.; Wang, X.; Wang, Y.; Wang, R.; Yang, Y.; Shi, S.; Tu, Z. Encouraging divergent thinking in large language models through multi-agent debate. *arXiv* **2023**, arXiv:2305.19118.

27. Park, S.; Kim, J.; Jin, S.; Park, S.; Han, K. PREDICT: Multi-Agent-based Debate Simulation for Generalized Hate Speech Detection. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, Miami, FL, USA, 12–16 November 2024; pp. 20963–20987.
28. Wang, H.; Du, X.; Yu, W.; Chen, Q.; Zhu, K.; Chu, Z.; Yan, L.; Guan, Y. Learning to break: Knowledge-enhanced reasoning in multi-agent debate system. *Neurocomputing* **2025**, *618*, 129063. [CrossRef]
29. Li, Y.; Du, Y.; Zhang, J.; Hou, L.; Grabowski, P.; Li, Y.; Ie, E. Improving multi-agent debate with sparse communication topology. *arXiv* **2024**, arXiv:2406.11776.
30. Smit, A.P.; Duckworth, P.; Grinsztajn, N.; Tessera, K.; Barrett, T.D.; Pretorius, A. Are we going MAD? Benchmarking multi-agent debate between language models for medical Q&A. In Proceedings of the Deep Generative Models for Health Workshop NeurIPS 2023, New Orleans, LA, USA, 15 December 2023.
31. Liang, B.; Zhu, Q.; Li, X.; Yang, M.; Gui, L.; He, Y.; Xu, R. 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), Dublin, Ireland, 22–27 May 2022; Association for Computational Linguistics: Stroudsburg, PA, USA, 2022; Volume 1, pp. 81–91.
32. Li, A.; Liang, B.; Zhao, J.; Zhang, B.; Yang, M.; Xu, R. Stance detection on social media with background knowledge. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Singapore, 6–10 December 2023; pp. 15703–15717.
33. Zhang, B.; Ding, D.; Huang, Z.; Li, A.; Li, Y.; Zhang, B.; Huang, H. Knowledge-augmented interpretable network for zero-shot stance detection on social media. *IEEE Trans. Comput. Soc. Syst.* **2024**, 1–12. [CrossRef]
34. Li, A.; Zhao, J.; Liang, B.; Gui, L.; Wang, H.; Zeng, X.; Wong, K.F.; Xu, R. Mitigating biases of large language models in stance detection with calibration. *arXiv* **2024**, arXiv:2402.14296.
35. Ding, Y.; Lei, Y.; Wang, A.; Liu, X.; Zhu, T.; Li, Y. Adversarial contrastive representation training with external knowledge injection for zero-shot stance detection. *Neurocomputing* **2025**, *614*, 128849. [CrossRef]
36. Lan, X.; Gao, C.; Jin, D.; Li, Y. Stance detection with collaborative role-infused llm-based agents. In Proceedings of the International AAAI Conference on Web and Social Media, Buffalo, NY, USA, 3–6 June 2024; Volume 18, pp. 891–903.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.