electronics

MDPI

Article

From Claims to Stance: Zero-Shot Detection with Pragmatic-Aware Multi-Agent Reasoning

Zhiyu Xie [1,†], Fuqiang Niu [2,†], Genan Dai [1] and Bowen Zhang [1,*]

[1] School of Artificial Intelligence, Shenzhen Technology University, Shenzhen 518118, China
[2] School of Cyber Science and Technology, University of Science and Technology of China, Hefei 230026, China
* Correspondence: zhangbowen@sztu.edu.cn
† These authors contributed equally to this work.

Abstract

Stance detection aims to identify whether a text expresses a favorable, opposing, or neutral attitude toward a given target and has become increasingly important for analyzing public discourse on social media. Existing approaches, ranging from supervised neural models to prompt-based large language models (LLMs), face two persistent challenges: the scarcity of annotated stance data across diverse targets and the difficulty of generalizing to unseen targets under pragmatic and rhetorical variation. To address these issues, we propose PAMR (Pragmatic-Aware Multi-Agent Reasoning), a zero-shot stance detection framework that decomposes stance inference into structured reasoning steps. PAMR orchestrates three LLM-driven agents—a linguistic parser that extracts pragmatic markers and canonicalizes claims, an NLI-based estimator that produces calibrated stance probabilities through consensus voting, and a counterfactual and view-switching auditor that probes robustness under controlled rewrites. A stability-aware fusion integrates these signals, conservatively abstaining when evidence is uncertain or inconsistent. Experiments on SemEval-2016 and COVID-19-Stance show that PAMR achieves macro-F1 scores of 71.9% and 73.0%, surpassing strong zero-shot baselines (FOLAR and LogiMDF) by +2.0% and +3.1%. Ablation results confirm that pragmatic cues and counterfactual reasoning substantially enhance robustness and interpretability, underscoring the value of explicit reasoning and pragmatic awareness for reliable zero-shot stance detection on social media.

Keywords: zero-shot stance detection; multi-agent framework; large language model

check for updates

Academic Editor: Arkaitz Zubiaga

Received: 15 September 2025
Revised: 24 October 2025
Accepted: 28 October 2025
Published: 31 October 2025

Citation: Xie, Z.; Niu, F.; Dai, G.; Zhang, B. From Claims to Stance: Zero-Shot Detection with Pragmatic-Aware Multi-Agent Reasoning. *Electronics* **2025**, *14*, 4298. https://doi.org/10.3390/electronics14214298

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Social media has become a central arena for public discourse, where individuals express their opinions on controversial issues such as politics, social movements, healthcare, and environmental policies. Understanding public stance toward such topics is crucial for applications including opinion mining, misinformation detection, crisis monitoring, and policy-making support. Stance detection—the task of classifying a text as favorable, against, or neutral with respect to a target—has therefore emerged as a fundamental problem in natural language processing (NLP) and social computing [1]. Early studies treated stance detection as a supervised classification task on social media platforms such as Twitter, focusing on political debates and event-specific targets [2,3]. Subsequent research leveraged neural architectures and pre-trained language models to improve contextual understanding [4,5], yet most methods still depend on large annotated datasets and struggle to generalize across unseen targets. These limitations have motivated recent exploration

into cross-target and zero-shot stance detection paradigms [6,7], which aim to infer stance for new topics without task-specific training data. In practice, however, stance on social media is rarely expressed in a straightforward manner. Instead, it is often conveyed implicitly through pragmatic cues such as sarcasm, negation, rhetorical questions, or figurative language [8]. These linguistic devices can obscure or even invert the intended polarity, making it difficult for systems to distinguish stance from surface sentiment [9]. This suggests that effective stance detection should not treat the input merely as a text–target pair, but rather as a reasoning task that accounts for the author's underlying claim and its pragmatic framing.

Traditional stance detection methods initially focused on supervised neural architectures, including recurrent networks, attention mechanisms, and graph-based models, which rely on large quantities of annotated stance data [3,4,10,11]. With the rise of pre-trained language models (PLMs), fine-tuning strategies achieved stronger performance by leveraging contextual embeddings [5,12]. However, these methods still face two fundamental challenges. First, annotated stance data remain scarce and unevenly distributed across targets, making supervised training impractical for new domains. Second, generalization to unseen targets remains difficult, even with powerful PLMs, as stance expressions often vary in subtle ways across topics [13].

To alleviate these issues, recent research has turned to cross-target [6] and zero-shot stance detection (ZSSD) [7]. Cross-target methods attempt to transfer stance knowledge from source domains to unseen targets using contrastive learning, knowledge injection, or graph reasoning. Meanwhile, zero-shot approaches leverage PLMs and large language models (LLMs) directly via prompting or inference reformulation [14]. Despite recent progress, two persistent gaps remain. First, many approaches conflate stance with sentiment, especially in emotionally charged or figurative language. Second, existing models lack explicit mechanisms to handle pragmatic phenomena—such as sarcasm, hedging, and negation—or to assess whether stance predictions are stable under slight linguistic variations. These challenges suggest that monolithic architectures may struggle to disentangle the complex and often interacting cues that shape stance expression.

These observations suggest several concrete desiderata for zero-shot stance detection systems: (1) extract explicit, target-linked canonical claims rather than relying solely on surface expressions; (2) remain robust to pragmatic confounds such as sarcasm and negation; (3) ensure prediction stability through counterfactual and perspective-based probing; (4) adopt calibrated decision strategies that abstain to neutral when evidence is thin or unstable.

Inspired by recent advances in multi-agent reasoning—particularly division of labor, coordination, and self-verification capabilities [15,16], we introduce **PAMR**—Pragmatic-Aware Multi-Agent Reasoning—a zero-shot stance detection framework that decomposes the task into interpretable subtasks and re-assembles their signals through stability-aware decision making. PAMR orchestrates three LLM-driven agents: (1) a Linguistic Parser that distills the input into a canonical claim while extracting pragmatic markers; (2) an NLI-based Estimator that produces a calibrated distribution {favor, against, neutral}; and (3) a Counterfactual and View-Switching module that probes robustness by re-evaluating stance under meaning-preserving rewrites (e.g., removing sarcasm, switching voice). A lightweight stability-aware fusion integrates these signals, conservatively assigning neutral when confidence is low, predictions are unstable, or top classes are tied. PAMR requires no task-specific fine-tuning, yields auditable intermediate outputs (claims, pragmatic markers, robustness flips), and explicitly mitigates common failure modes such as sarcasm-induced polarity errors.

The main contributions of this paper are summarized as follows:

1. We propose PAMR, a pragmatic-aware multi-agent framework for zero-shot stance detection that factors inference into claim normalization, probabilistic NLI, and robustness probing, enabling interpretable and modular reasoning.
2. We introduce a counterfactual and view-switching probe that quantifies stability of stance under meaning-preserving edits and perspective shifts; this signal directly informs a stability-aware fusion rule that curbs over-confident polarity errors and disentangles stance from sentiment.
3. We conduct extensive experiments on benchmark datasets, demonstrating that PAMR matches or surpasses strong zero-shot baselines. Ablations reveal the contributions of pragmatic cues and robustness probing, while additional analyses show that PAMR produces interpretable intermediate artifacts that enable fine-grained audits.

Overall, this study bridges computational modeling and social media discourse analysis by explicitly integrating pragmatic reasoning into stance inference. We aim to provide a framework that not only improves zero-shot prediction accuracy but also deepens interpretability for real-world applications such as misinformation monitoring and public opinion tracking. The remainder of this paper is organized as follows. Section 2 introduces related work. Section 3 describes the proposed method. Section 4 presents the experimental setup. Section 5 reports the experimental results and analysis. Section 6 discusses key findings and limitations. Section 7 concludes the paper.

2. Related Works

2.1. Traditional Stance Detection Methods

Stance detection, closely related to sentiment analysis, argument mining, and fact verification, has been studied extensively in natural language processing. Early research primarily relied on supervised learning with handcrafted features and classical classifiers. With the development of deep learning, neural architectures such as CNNs [17,18] and LSTMs [10] became prevalent, enabling models to capture contextual and sequential dependencies in text. Graph neural networks were later introduced to encode relations among posts, users, and targets, further enriching stance representations. The emergence of PLMs significantly advanced the field. Fine-tuning strategies reformulated stance detection as a text–target classification problem by concatenating target and input sequences [19]. To reduce computational cost, lightweight adaptation methods froze most PLM parameters while tuning small modules [14]. Beyond fine-tuning, prompt-based techniques framed stance detection as masked language modeling. By filling in templates such as "The attitude toward <Target> is [MASK]," PLMs could predict stance more effectively. Recent studies improved this paradigm by designing adaptive prompts tailored to different targets [20]. In addition to general PLMs, social-media-oriented variants such as BERTweet [21] and CT-BERT [22] have been developed to better capture the linguistic characteristics of Twitter and COVID-19 discourse. While these models enhance representation learning on noisy text, they remain limited in addressing pragmatic phenomena such as sarcasm or negation and lack interpretability or stability validation mechanisms. Multi-modal variants have also been proposed, combining textual and visual cues through prompt-based mechanisms [23]. Despite these innovations, most traditional approaches remain reliant on annotated supervision and show limited robustness when confronted with pragmatic language use.

2.2. Zero-Shot Stance Detection

To overcome the dependence on labeled data, zero-shot stance detection has been widely explored. In this setting, models must transfer knowledge from source targets to

entirely new ones without direct supervision. Early work employed contrastive learning to align stance features across domains, as in JointCL [24], while other approaches leveraged external knowledge bases to bridge semantic gaps between known and unseen targets, such as TarBK. Graph-based methods further advanced the field by constructing heterogeneous or multi-view graphs that capture transferable stance signals among tweets, claims, and targets. Recently, researchers have investigated reasoning-based enhancements to improve zero-shot generalization. LogiMDF [25] integrates first-order logic constraints into multi-decision fusion, ensuring consistency across LLM outputs while leveraging hypergraph propagation. These methods highlight a shift from purely data-driven transfer to structured reasoning, which improves both robustness and explainability. Nevertheless, pragmatic challenges such as sarcasm and rhetorical framing remain largely unresolved in ZSSD.

2.3. LLM-Based Stance Detection Methods

The remarkable zero-shot and few-shot capabilities of large language models have reshaped stance detection research. One line of work directly treats LLMs as stance predictors: with carefully designed prompts, LLMs can classify stance without additional training [14,26]. Strategies include direct zero-shot prompting, chain-of-thought reasoning, and prompt designs enriched with background knowledge. While flexible, such direct use often leads to unstable outputs and inconsistent predictions. Another line of research employs LLMs as knowledge providers that augment smaller, trainable models. For example, domain-relevant background information can be elicited from LLMs and injected into stance classifiers [27], combining the interpretive capacity of LLMs with the adaptability of supervised models. More recent efforts emphasize structured reasoning with interpretable intermediate steps. Enhanced CoT methods [28–30] decompose stance prediction into step-wise inferences (e.g., from factual entailment to subjective alignment), but rarely define rule-based constraints or auditable outputs. Some systems, such as FOLAR, integrate symbolic representations like sentiment trajectories or discourse role tags, while LogiMDF incorporates logical rules over hypergraphs for traceable reasoning. However, the internal reasoning paths often remain implicit or non-verifiable.

2.4. Summary and Research Gap

Across these paradigms, most prior studies emphasize semantic transfer or large-scale contextual modeling but seldom incorporate explicit pragmatic reasoning or stability verification. For instance, social-media-oriented models such as BERTweet, CT-BERT, and RoBERTa-Twitter capture linguistic variations in tweets but remain limited in handling pragmatic phenomena like sarcasm or negation. Reasoning-based approaches (e.g., LogiMDF, CIRF) introduce logical or cognitive structures, yet their reasoning traces are often implicit and lack auditable intermediate outputs. Overall, existing frameworks can be contrasted along three dimensions—(1) whether pragmatic cues are explicitly modeled, (2) whether prediction stability is verified, and (3) whether interpretable intermediate reasoning is provided. PAMR addresses these gaps by integrating all three: it explicitly encodes pragmatic markers, employs counterfactual probing for stability, and yields interpretable intermediate artifacts for transparent stance reasoning.

3. Methods

3.1. Task Definition

Given a short text x and a target t, the goal of zero-shot stance detection is to predict

$$y \in \{\texttt{favor}, \texttt{against}, \texttt{neutral}\}$$

without using any annotated stance data for t, i.e., $\mathcal{T}_{\text{train}} \cap \mathcal{T}_{\text{test}} = \emptyset$. Our approach decomposes this prediction into multi-agent reasoning steps with interpretable intermediate outputs and evaluates prediction stability via agent-level sampling.

3.2. PAMR Overview

As shown in Figure 1, PAMR (*Pragmatic-Aware Multi-Agent Reasoning*) decomposes zero-shot stance detection into four interpretable stages, each implemented as an explicit agent that transforms inputs into structured intermediate representations. These agents interact in a modular pipeline to progressively refine stance decisions.

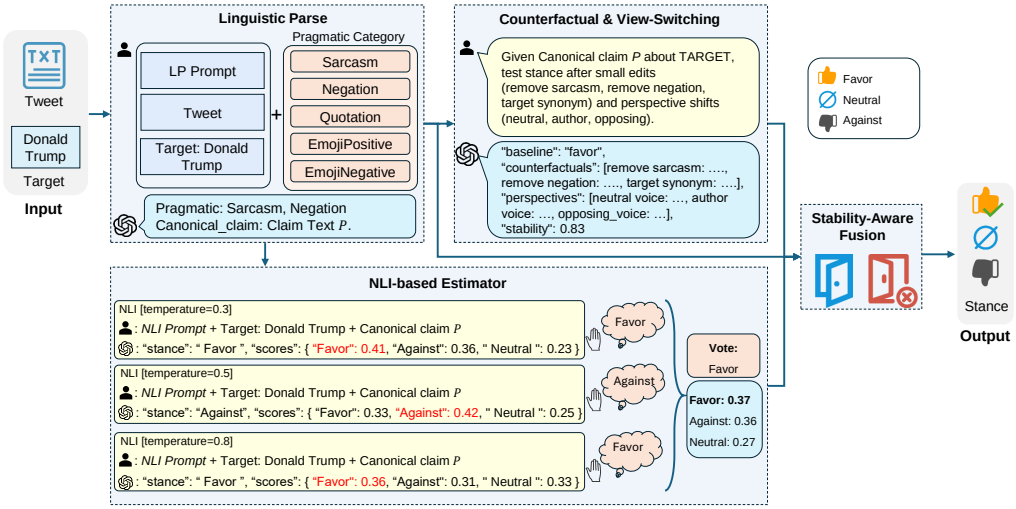


Figure 1. PAMR Framework overall.

(1) **Linguistic Parse Agent**: takes the raw social media post x and target t as input, and outputs a *target-linked canonical claim* along with pragmatic markers such as sarcasm, hedging, or negation. This ensures that subsequent reasoning is grounded in normalized propositions instead of noisy surface forms.

(2) **NLI Estimation Agent**: Reformulates stance classification as a natural language inference task. Given the canonical claim and target, it runs multiple inference passes using diverse prompts to produce both a probability distribution over stance labels and a consensus vote count, mitigating randomness.

(3) **Counterfactual and View-Switching Agent**: Evaluates the stability of the predicted stance by applying minimal, meaning-preserving perturbations (e.g., changing tone, perspective). It outputs a scalar *stability score* reflecting how robust the original decision is under linguistic variation.

(4) **Stability-Aware Fusion Agent**: Receives all signals—canonical claim, pragmatic tags, NLI probabilities and consensus, stability score—and integrates them under a conservative policy. It abstains to "neutral" if confidence is low or predictions are unstable, ensuring robustness.

Each agent has a well-defined role, input–output interface, and interpretable output, and the full pipeline corresponds directly to the annotated modules in Figure 1. In particular, the Linguistic Parse Agent is responsible for the canonicalization and pragmatic enrichment mentioned in the figure.

3.3. Linguistic Parse

The Linguistic Parse Agent aims to normalize noisy social media text into a target-linked canonical claim, while detecting pragmatic markers (e.g., sarcasm, negation, quota-

tion, emoji) that may invert or obscure stance polarity. Given a tweet x and target t, the LLM generates a structured output containing pragmatic tags and a canonical claim c.

Canonical Claim Extraction: The model distills the author's underlying proposition regarding the target into a clear canonical claim, ensuring that the subsequent stance reasoning operates on an explicit statement rather than noisy surface text.

Pragmatic Tagging: The parser identifies pragmatic markers such as sarcasm, negation, quotation, and emoji usage, which are later incorporated in fusion to calibrate stance decisions.

Input Data: Tweet x, Target t
Prompt: Identify pragmatic markers (sarcasm, negation, quotation, emoji) and rewrite the tweet into a canonical claim about the target.
Expected Output: Pragmatic tags and a canonical claim c.

3.4. NLI-Based Estimator

The NLI-based Estimator stage reframes stance detection as a natural language inference (NLI) problem. The canonical claim c and target t are paired as premise–hypothesis inputs to the LLM. To mitigate stochasticity and improve reliability, the model performs n independent inference runs under different decoding parameters.

Multi-run Voting: Each run outputs a stance label $\{favor, against, neutral\}$ and a probability distribution $\mathbf{p}$. A majority vote and averaged probability vector $\bar{\mathbf{p}}$ are then aggregated for stability.

Input Data: Canonical claim c, Target t
Prompt: Given c and t, determine whether the author's stance is favor, against, or neutral. Output both the label and probability scores.
Expected Output: $\{stance, scores\}$ per run, aggregated into majority label $y^\star$, vote ratio r, and averaged scores $\bar{\mathbf{p}}$.

3.5. Counterfactual and View-Switching

The Counterfactual and View-Switching (CVS) stage evaluates robustness of stance predictions by rephrasing the canonical claim under minimal edits and perspective shifts. This stage probes whether stance remains consistent across semantically faithful variations.

Counterfactual Edits: The claim is modified by (i) removing sarcasm, (ii) removing negation, and (iii) replacing the target with a synonym. Perspective Shifts: The claim is paraphrased into three perspectives: neutral voice, author voice, and opposing voice. Stability Score: The proportion of paraphrases consistent with the baseline stance is reported as the stability score S. By definition, the stability score $S \in [0, 1]$ quantifies the proportion of paraphrases whose predicted stance remains consistent with the baseline. Higher values of S therefore indicate greater prediction stability under controlled linguistic rewrites, and thus higher confidence in the stance.

Input Data: Canonical claim c, Target t
Prompt: Given c, re-evaluate stance after small edits (remove sarcasm, remove negation, target synonym) and perspective shifts (neutral, author, opposing).
Expected Output: JSON with baseline stance, stance for each edit/perspective, and stability score S.

3.6. Stability-Aware Fusion

Thresholds such as $\tau_{\text{unstable}} = 0.40$ control model behavior; for example, if $S < \tau_{\text{unstable}}$, the model abstains by predicting "neutral" to avoid over-confident errors. Finally, we

aggregate signals with a gate-by-gate cascade that mirrors Figure 1. Let $\bar{\mathbf{p}}$ be the averaged scores, $y^\star/r$ the consensus, S the stability, and Π pragmatic tags. Algorithm 1 lists the fusion procedure.

Filtering Policy

(1) Validity: If c is empty or $\bar{\mathbf{p}}$ missing, return `neutral`. (2) Stability: If $S < \tau_{\text{unstable}}$, return `neutral`. (3) Pragmatics-aware confidence: Let $y = \arg\max \bar{\mathbf{p}}$ and $p_{\max} = \max \bar{\mathbf{p}}$; if `Sarcasm` or `Negation` $\in \Pi$, subtract a penalty λ_{prag} from $p_{\max}$. (4) Consensus override under low confidence: If adjusted $p_{\max} < \tau_{\text{prob}}$ and $r \geq \tau_{\text{cons}}$, output $y^\star$; else `neutral`. (5) Prefer consensus: If $r \geq \tau_{\text{cons}}$, output $y^\star$. (6) Small-margin flip: When $y =$ `favor` and lead $= p_{\text{favor}} - \max(p_{\text{against}}, p_{\text{neutral}}) \leq \tau_{\text{flip}}$, flip to `against` if `Sarcasm` or `Negation` $\in \Pi$. (7) Tie-to-neutral: If the top-2 gap $< \epsilon$, output `neutral`. (8) Fallback: Otherwise output y.

Algorithm 1 Stability-Aware Fusion.

Require: Π, $\bar{\mathbf{p}}$, $y^\star$, r, S; thresholds $\tau_{\text{unstable}}, \tau_{\text{prob}}, \tau_{\text{cons}}, \tau_{\text{flip}}, \epsilon$; penalty λ_{prag}

if c empty or $\bar{\mathbf{p}}$ missing **then return** `neutral`
if $S < \tau_{\text{unstable}}$ **then return** `neutral`
$y \leftarrow \arg\max_i p_i$, $p_{\max} \leftarrow \max_i p_i$ ▷ $\bar{\mathbf{p}} = (p_{\text{favor}}, p_{\text{against}}, p_{\text{neutral}})$
if `Sarcasm`$\in \Pi$ **or** `Negation`$\in \Pi$ **then**
    $\tilde{p}_{\max} \leftarrow \max(0,\ p_{\max} - \lambda_{\text{prag}})$
else
    $\tilde{p}_{\max} \leftarrow p_{\max}$
end if
if $\tilde{p}_{\max} < \tau_{\text{prob}}$ **then**
    if $r \geq \tau_{\text{cons}}$ **then**
        return $y^\star$
    else
        return `neutral`
    end if
end if
if $r \geq \tau_{\text{cons}}$ **then**
    return $y^\star$
end if
if $y =$ `favor` **then**
    lead $\leftarrow p_{\text{favor}} - \max(p_{\text{against}}, p_{\text{neutral}})$
    if lead $\leq \tau_{\text{flip}}$ **and** (`Sarcasm`$\in \Pi$ **or** `Negation`$\in \Pi$) **then**
        return `against`
    end if
end if
Let $p_{(1)} \geq p_{(2)}$ be the top-2 of $\bar{\mathbf{p}}$
if $|p_{(1)} - p_{(2)}| < \epsilon$ **then**
    return `neutral`
end if
return y

4. Experiments

4.1. Experimental Data

To evaluate the effectiveness of our approach, we run thorough experiments on two datasets: SemEval-2016 Task 6 (SEM16) [2] and COVID-19-Stance (COVID19) [31].

- SemEval-2016 (SEM16) [2]: The SEM16 dataset contains 4870 tweets, each targeting various subjects and annotated with one of three stance labels: "favor", "against", or "neutral". SEM16 provides six targets (Donald Trump (DT), Hillary Clinton (HC), Feminist Movement (FM), Legalization of Abortion (LA), Atheism (AT), and Climate Change (CC)). Following [25] for zero-shot evaluation, we exclude Atheism (AT) and

Climate Change (CC) and use only the official test split for fair comparison with prior zero-shot settings. Per-target class counts for the four retained targets are reported in Table 1.

- COVID-19-Stance (COVID-19) [31]: We also use the COVID-19 stance dataset, which assesses public attitudes toward pandemic-related policies and figures across four targets: *Wearing a Face Mask* (WA), *Keeping Schools Closed* (SC), *Anthony S. Fauci, M.D.* (AF), and *Stay at Home Orders* (SH). Each tweet is labeled with *Favor*/*Against*/*Neutral*. As with SEM16, we report results using the test split only; class distributions are shown in Table 1.

Table 1. Statistics of the datasets used in our experiments.

Dataset	Target	Favor	Against	Neutral	Total
SEM16	DT	148	299	260	707
	HC	163	565	251	979
	FM	268	511	170	949
	LA	167	544	222	933
COVID-19	AF	492	610	762	1864
	SH	615	250	325	1190
	WA	693	190	668	1551
	SC	400	782	346	1528

Both datasets were chosen as they are widely adopted benchmarks for stance detection and provide diverse, topic-specific challenges. SEM16 offers classic political and social targets that test generalization across ideological domains, while COVID-19-Stance introduces a contemporary and highly pragmatic context involving public health discourse. We acknowledge that both datasets are relatively small in size (a few thousand tweets each), which may limit the statistical power of comparisons. To mitigate this, all evaluations are conducted in a strict zero-shot setting using the official test splits only, ensuring comparability with prior studies. Furthermore, our stability-aware fusion mechanism and counterfactual probing explicitly reduce the sensitivity of results to sampling variance, partially alleviating dataset-size constraints.

4.2. Evaluation Metrics

We adopt the Macro-F1 score F_{avg} computed as the average of the F1 scores of the *Favor* and *Against* categories (the *Neutral* class is excluded from the average), following standard practice in stance detection [2,7,26]. Macro-F1 equally weights each class and is less affected by label imbalance, which is critical for social media datasets where "neutral" or "against" instances often dominate. In addition, we report per-target Macro-F1 averages to ensure comparability with prior zero-shot studies.Let P_y and R_y denote precision and recall for class $y \in \{\text{favor}, \text{against}\}$ The per-class F1 and the macro average are defined as

$$F_{\text{favor}} = \frac{2 \cdot P_{\text{favor}} \cdot R_{\text{favor}}}{P_{\text{favor}} + R_{\text{favor}}}, \tag{1}$$

$$F_{\text{against}} = \frac{2 \cdot P_{\text{against}} \cdot R_{\text{against}}}{P_{\text{against}} + R_{\text{against}}}, \tag{2}$$

$$F_{\text{avg}} = \frac{F_{\text{favor}} + F_{\text{against}}}{2}. \tag{3}$$

4.3. Baseline Methods

To ensure a comprehensive evaluation, we compare PAMR with a broad range of existing stance detection approaches, which can be grouped into traditional DNNs, fine-tuning-based methods, and LLM-based frameworks.

(1) TraditionalDNNs.

- **BiLSTM and Bicond [32]**: These two approaches utilize separate BiLSTM encoders, where one captures sentence-level semantics and the other encodes the given target, thereby enabling the model to jointly represent stance-related information.
- **CrossNet [33]**: This model leverages BiLSTM architectures to encode both the input text and its corresponding target, while introducing a target-specific attention mechanism before classification, which enhances the model's ability to generalize across unseen targets.
- **TPDG [34]**: This method automatically identifies stance-bearing words and distinguishes target-dependent from target-independent terms, adjusting them adaptively to better capture the relationship between text and target.
- **TOAD [35]**: To improve generalization in zero-shot scenarios, TOAD adopts an adversarial learning strategy that allows the model to resist overfitting to specific targets while transferring stance knowledge.

(2) Fine-tuning methods.

- **TGA-Net [36]**: This approach establishes associations between training and evaluation topics in an unsupervised way, using BERT as the encoder and fully connected layers for classification, thereby linking topic domains without annotated supervision.
- **Bert-Joint [37]**: It combines bidirectional encoder representations from transformers that have been pre-trained on large-scale unlabeled corpora, producing dense contextual embeddings for both tokens and full sentences.
- **Bert-GCN [38]**: This method enhances stance detection with common-sense knowledge by integrating both structural and semantic graph relations, which makes it more effective in generalizing to zero and few-shot target scenarios.
- **JointCL [24]**: It unifies stance-oriented contrastive learning with target-aware prototypical graph contrastive learning, allowing the model to transfer stance-relevant features learned from seen topics to unseen targets.
- **TarBK [39]**: By leveraging Wikipedia-derived background knowledge, TarBK reduces the semantic gap between training and evaluation targets, thereby improving the reasoning capability of stance classifiers.
- **PT-HCL [7]**: This contrastive learning approach utilizes both semantic and sentiment features to improve cross-domain stance transferability, enabling robust generalization beyond source data. Its combination of semantics and sentiment proves critical for disambiguating stance from emotion.
- **KEPrompt [40]**: This method proposes an automatic verbalizer to generate label words dynamically, while simultaneously injecting external background knowledge to guide stance recognition. It reduces reliance on manually designed verbalizers and improves flexibility.

(3) LLM-based methods.

- **COLA [14]**: This approach employs a three-stage framework where different LLM roles are orchestrated for multidimensional text understanding and reasoning, resulting in state-of-the-art zero-shot stance performance. It demonstrates the effectiveness of task decomposition within LLM pipelines.

- **Ts-CoT [41]**: It introduces a chain-of-thought prompting mechanism for stance detection with LLMs, upgrading the base model to GPT-3.5 in order to take advantage of improved reasoning capacity. The CoT design encourages step-by-step reasoning rather than direct label prediction.
- **EDDA [27]**: This method exploits LLMs to automatically generate rationales and substitute stance-bearing expressions, thereby increasing semantic relevance and expression diversity for stance detection. By focusing on rationales, EDDA improves both performance and interpretability.
- **FOLAR [42]**: A reasoning framework that augments stance detection with factual knowledge and chain-of-thought logical reasoning, aiming to improve interpretability and robustness in zero-shot settings.
- **LogiMDF [25]**: A logic-augmented multi-decision fusion framework that extracts first-order logic rules from multiple LLMs, constructs a logical fusion schema, and employs a multi-view hypergraph neural network to integrate diverse reasoning processes for consistent and accurate stance detection.

4.4. Implementation Details

We implement PAMR using `GPT-3.5-turbo` accessed via the OpenAI API. For Linguistic Parse and CVS, the maximum generation length is set to 256 and 512 tokens, respectively, with temperature 0.3. For NLI-based estimation, we run the model three times with temperatures $\{0.3, 0.5, 0.8\}$ and aggregate results by majority vote and averaged probabilities. Fusion thresholds are fixed across datasets: $\tau_{\text{unstable}} = 0.40$, $\tau_{\text{prob}} = 0.45$, $\tau_{\text{cons}} = 0.60$, $\tau_{\text{flip}} = 0.15$, $\epsilon = 0.02$, and pragmatic penalty $\lambda_{\text{prag}} = 0.05$. All evaluations are conducted in a strict zero-shot setting using the official test splits, and performance is reported in terms of macro-F1 over `favor` and `against`.

5. Overall Performance

5.1. Analysis of Main Results

Table 2 reports the results on the SEM16 dataset. Early neural baselines such as BiLSTM, Bicond, and CrossNet achieve relatively low performance, with average scores around 34–38. With the introduction of pre-trained models, methods like JointCL, PT-HCL, and NPS4SD raise the average performance to around 52–55. More recent LLM-based approaches, such as COLA, FOLAR, and LogiMDF, achieve competitive results on specific targets (e.g., COLA and FOLAR both exceed 81 on HC). However, their performance tends to fluctuate across different targets. In contrast, PAMR achieves strong and consistent results across all targets, with best scores on FM (75.9) and LA (71.8), and a balanced overall average of 71.9, surpassing all baselines. This demonstrates the effectiveness of pragmatic-aware reasoning and stability fusion in achieving robust zero-shot stance detection. Here, Avg denotes the arithmetic mean over all targets.

Table 3 shows the results on the COVID-19 dataset. Traditional baselines (CrossNet, BERT, TPDG) perform poorly, with average scores between 40 and 45. Stronger models like TOAD and JointCL improve performance, while LogiMDF achieves the best results on AF (70.4) and WA (75.4), and FOLAR performs strongly on WA (73.1). Nevertheless, PAMR consistently achieves competitive or best performance on nearly all targets, obtaining the top scores on SC (72.0), SH (72.2), and WA (78.8). Its overall average reaches 73.0, outperforming all comparison methods. This indicates that PAMR is better able to handle pragmatic confounds such as sarcasm and negation in COVID-19 tweets, yielding more stable and reliable stance predictions. Again, Avg is the mean over all targets in the dataset.

Table 2. Results on the SEM16 dataset. F_{avg} is reported for each target. The Avg is the arithmetic mean over all listed targets. The best results are in bold. ‡ indicates the first-best result.

Method	HC	FM	LA	DT	Avg
BiLSTM	31.6	40.3	33.6	30.8	34.1
Bicond	32.7	40.6	34.4	30.5	34.6
CrossNet	38.3	41.7	38.5	35.6	38.5
TPDG	50.9	53.6	46.5	47.3	49.6
TOAD	51.2	54.1	46.2	49.5	50.3
TGA-Net	49.3	46.6	45.2	40.7	45.5
Bert-Joint	50.1	42.1	44.8	41.0	44.5
Bert-GCN	50.0	44.3	44.2	42.3	45.2
JointCL	54.4	54.0	50.0	50.5	52.2
TarBK	55.1	53.8	48.7	50.8	52.1
PT-HCL	54.5	54.6	50.9	50.1	52.5
KEPROMPT	57.0	53.6	53.0	41.8	51.3
NPS4SD	60.1	56.7	51.0	51.4	54.8
COLA	**81.7**	63.4	71.0	68.5	71.2
Ts-CoT$_{GPT}$	78.9	68.3	62.3	68.6	69.5
EDDA	77.4	69.7	62.7	**69.8**	69.9
FOLAR	81.9	71.2	69.9	–	–
LogiMDF	75.1	67.9	68.0	67.6	69.7
PAMR (Ours)	73.7	**75.9** ‡	**71.8** ‡	66.1	**71.9**

Table 3. Results on the COVID-19 dataset. We report per-target Macro-F1 and the overall Avg, which is the mean across all targets. The best numbers are in bold. ‡ indicates the first-best result.

Method	AF	SC	SH	WA	Avg
CrossNet	41.3	40.0	40.4	38.2	40.0
BERT	47.3	45.0	39.9	44.3	44.1
TPDG	46.0	51.5	37.2	48.0	45.7
TOAD	53.0	68.3	62.9	41.1	56.3
JointCL	57.6	49.3	43.5	63.1	53.4
Ts-CoT	69.2	43.5	66.5	57.8	59.3
COLA	65.7	46.6	53.5	73.9	59.9
FOLAR	69.5	67.2	65.4	73.1	68.8
LogiMDF	**70.4**	68.8	64.9	75.4	69.9
PAMR	68.6	**72.0** ‡	**72.2** ‡	**78.8** ‡	**73.0**

To further validate these improvements, we conduct paired t-tests between PAMR and the strongest baseline across all targets on both datasets. The results show that PAMR's performance gains are statistically significant ($p < 0.05$), confirming that the observed advantages are unlikely due to random variation.

5.2. Ablation Study

To better understand the contributions of different components in PAMR, we conduct ablation experiments by removing the Linguistic Parser (w/o LP) and the Counterfactual and View-Switching module (w/o CVS). The results are shown in Figure 2.

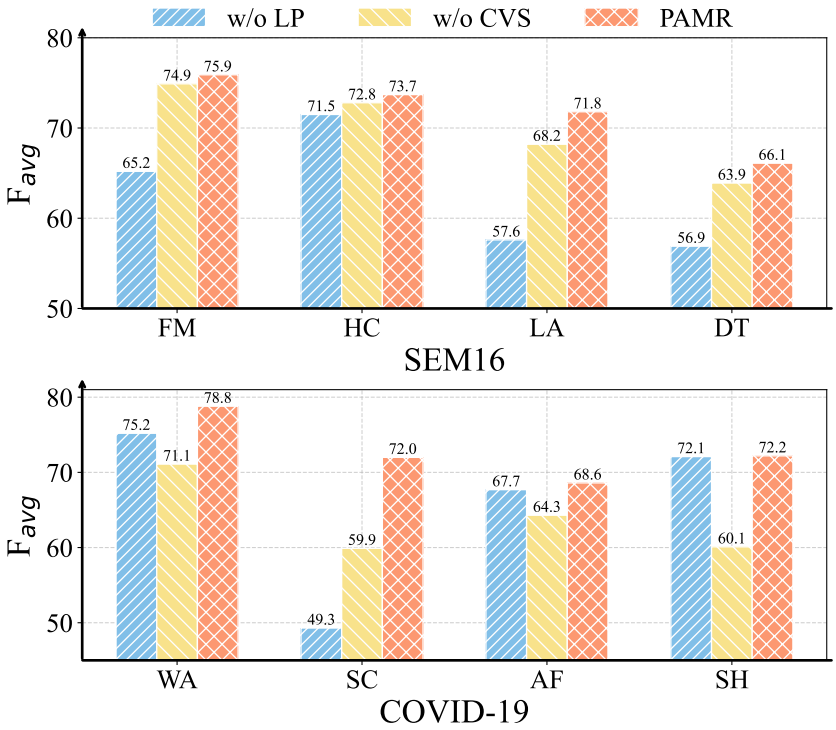


Figure 2. Ablation study of PAMR on SEM16 and COVID-19.

We observe that removing either component consistently degrades performance across both COVID-19 and SemEval-2016 datasets. Specifically, eliminating the Linguistic Parser leads to a substantial drop, e.g., from 72.0 to 49.3 on SC and from 66.1 to 56.9 on DT, confirming the importance of extracting canonical claims and pragmatic cues for stance reasoning. Similarly, removing the Counterfactual and View-Switching module also results in performance declines, particularly on WA (from 78.8 to 71.1) and FM (from 75.9 to 74.9), highlighting the necessity of stability probing to mitigate pragmatic confounds and prevent over-confident errors.

Overall, the full PAMR framework consistently outperforms its ablated variants, demonstrating that both pragmatic parsing and counterfactual stability probing are complementary and essential for robust zero-shot stance detection.

5.3. Case Study

To further illustrate the behavior of PAMR, we analyze representative examples from the evaluation set, as shown in Table 4.

Case 1. In this example, the canonical claim and the original tweet exhibit strong semantic alignment. The NLI-based estimation assigns a dominant probability to `favor` ($p = 0.65$), clearly surpassing both `against` and `neutral`. The CVS stability score reaches $S = 0.83$, well above the instability threshold, indicating robust consistency under counterfactual and perspective shifts. Moreover, majority voting yields complete unanimity (3/3 runs), reinforcing the reliability of the prediction. Since no pragmatic markers (e.g., sarcasm or negation) are present, the system directly outputs `favor`, which matches the gold annotation. This case shows that, under standard conditions, PAMR can capture entailment relations both stably and accurately.

Table 4. Case studies on **SemEval-2016** Trump target showing intermediate outputs.

Tweet	Claim	Pragmatic Tags	NLI Scores	Stability	Prediction
People are saying that kids will not have a safe place to go to if the schools are closed. Large gatherings are not safe with coronavirus. The coronavirus is not safe.	Keeping schools closed may leave kids without a safe place to go, but large gatherings are unsafe due to coronavirus.	[Quotation]	{favor: 0.65, against: 0.15, none: 0.20}	0.833	Favor
"2 years ago Hillary never answered whether she used private email. Liberal media passed on reporting." #equality	Questioning Hillary's email while excusing others reflects the sexist double standards women in politics face.	[Sarcasm, Quotation]	{favor: 0.50, against: 0.46, none: 0.04}	0.666	Against

Case 2. Here, the canonical claim and initial NLI probabilities suggest an approval stance (`favor`) toward the target, with the aggregated vote leaning in the same direction. However, the CVS stability score is only $S = 0.67$, indicating borderline robustness. Importantly, pragmatic markers such as `Sarcasm` are detected. Within the fusion stage, these cues trigger a polarity flip, shifting the final decision from `favor` to `against`, which aligns with the gold annotation. This case demonstrates PAMR's ability to exploit pragmatic signals to correct initial misclassifications in sarcastic contexts, thereby disentangling stance polarity from sentiment polarity.

6. Discussion

Our results highlight key insights into zero-shot stance detection. PAMR consistently outperforms baselines on both SEM16 and COVID-19 datasets, demonstrating that explicit modeling of pragmatics and stability improves generalization. Unlike prior models that conflate sentiment with stance, PAMR leverages pragmatic cues like sarcasm and negation to reduce polarity errors. Its modular design—with components for claim extraction, counterfactual probing, and fusion—enables interpretable outputs and fine-grained analysis. While PAMR currently relies on GPT-3.5, all reasoning steps are prompt-based and generalizable, making the framework compatible with future open-source or symbolic alternatives. Fusion strategies further enhance robustness under figurative or domain-specific inputs. Future work may explore lightweight replacements to improve accessibility and stability.

7. Conclusions

In this paper, we introduced PAMR, a pragmatic-aware multi-agent reasoning framework for zero-shot stance detection on social media. The framework enables explicit and interpretable reasoning by integrating linguistic, inferential, and counterfactual analyses without requiring target-specific training. Experiments on the SemEval-2016 and COVID-19-Stance datasets show that PAMR achieves stable improvements over strong zero-shot baselines, confirming the value of incorporating pragmatic cues and stability probing. While our results demonstrate promising robustness and interpretability, the model has been evaluated only on English social media datasets; its applicability to other domains or high-stakes contexts should be considered exploratory. Future work will extend

PAMR to multilingual and multi-modal settings, investigate domain adaptation under distribution shifts, and explore hybrid integration with symbolic reasoning to further enhance stability and transparency in zero-shot stance detection. Beyond these directions, PAMR offers a flexible foundation for future research on pragmatic reasoning and stance analysis. Its modular components can be adapted to evaluate or enhance other stance models, while its interpretable outputs provide useful tools for analyzing how pragmatic cues shape stance, supporting broader advancements in social media understandings.

Author Contributions: Conceptualization, Z.X. and B.Z.; Methodology, Z.X. and F.N.; Investigation, G.D.; Writing—original draft, Z.X. and F.N.; Writing—review and editing, B.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the Natural Science Foundation of Top Talent of SZTU (grant no. GDRC202320) and the Research Promotion Project of Key Construction Discipline in Guangdong Province (2022ZDJS112).

Data Availability Statement: No new data were created in this study. The datasets analyzed are publicly available benchmark datasets (SemEval-2016 and COVID-19), which can be obtained from their respective sources. Processed data and code are available from the corresponding author upon reasonable request.

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

References

1. Küçük, D.; Can, F. Stance detection: A survey. *ACM Comput. Surv. (CSUR)* **2020**, *53*, 12. [CrossRef]
2. 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.
3. Addawood, A.; Schneider, J.; Bashir, M. Stance classification of twitter debates: The encryption debate as a use case. In Proceedings of the 8th International Conference on Social Media & Society, Toronto, ON, Canada, 28–30 July 2017; pp. 1–10.
4. Sun, Q.; Wang, Z.; Zhu, Q.; Zhou, G. Stance detection with hierarchical attention network. In Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, NM, USA, 21–24 August 2018; pp. 2399–2409.
5. Li, Y.; Caragea, C. Target-Aware Data Augmentation for Stance Detection. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Online, 6–11 June 2021; pp. 1850–1860.
6. Xu, C.; Paris, C.; Nepal, S.; Sparks, R. Cross-target stance classification with self-attention networks. *arXiv* **2018**, arXiv:1805.06593.
7. Liang, B.; Chen, Z.; Gui, L.; He, Y.; Yang, M.; Xu, R. Zero-Shot Stance Detection via Contrastive Learning. In Proceedings of the ACM Web Conference 2022, Virtual, 25–29 April 2022; pp. 2738–2747.
8. Hong, G.N.S.Y.; Gauch, S. Sarcasm Detection as a Catalyst: Improving Stance Detection with Cross-Target Capabilities. *arXiv* **2025**, arXiv:2503.03787. [CrossRef]
9. Maynard, D.; Greenwood, M. Who cares about Sarcastic Tweets? Investigating the Impact of Sarcasm on Sentiment Analysis. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*; Calzolari, N., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., Piperidis, S., Eds.; European Language Resources Association (ELRA): Reykjavik, Iceland, 2014; pp. 4238–4243.
10. Du, J.; Xu, R.; He, Y.; Gui, L. Stance classification with target-specific neural attention networks. In Proceedings of the International Joint Conferences on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017.
11. Wei, P.; Lin, J.; Mao, W. Multi-target stance detection via a dynamic memory-augmented network. In Proceedings of the The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, Ann Arbor, MI, USA, 8–12 July 2018; pp. 1229–1232.
12. Li, Y.; Sosea, T.; Sawant, A.; Nair, A.J.; Inkpen, D.; Caragea, C. P-stance: A large dataset for stance detection in political domain. In Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Online Event, 1–6 August 2021; pp. 2355–2365.
13. Conforti, C.; Berndt, J.; Pilehvar, M.T.; Giannitsarou, C.; Toxvaerd, F.; Collier, N. Synthetic Examples Improve Cross-Target Generalization: A Study on Stance Detection on a Twitter corpus. In Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, WASSA@EACL 2021, Online, 19 April 2021; pp. 181–187.

14. 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.
15. Guo, T.; Chen, X.; Wang, Y.; Chang, R.; Pei, S.; Chawla, N.V.; Wiest, O.; Zhang, X. Large Language Model based Multi-Agents: A Survey of Progress and Challenges. *arXiv* **2024**, arXiv:2402.01680. [CrossRef]
16. Tran, K.T.; Dao, D.; Nguyen, M.D.; Pham, Q.V.; O'Sullivan, B.; Nguyen, H.D. Multi-Agent Collaboration Mechanisms: A Survey of LLMs. *arXiv* **2025**, arXiv:2501.06322. [CrossRef]
17. Zarrella, G.; Marsh, A. MITRE at SemEval-2016 Task 6: Transfer Learning for Stance Detection. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), San Diego, CA, USA, 16–17 June 2016; pp. 458–463.
18. Sobhani, P.; Inkpen, D.; Zhu, X. A dataset for multi-target stance detection. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, Valencia, Spain, 3–7 April 2017; pp. 551–557.
19. He, Z.; Mokhberian, N.; Lerman, K. Infusing Wikipedia Knowledge to Enhance Stance Detection. *arXiv* **2022**, arXiv:2204.03839.
20. Wang, S.; Pan, L. Target-Adaptive Consistency Enhanced Prompt-Tuning for Multi-Domain Stance Detection. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), Torino, Italia, 20–25 May 2024; pp. 15585–15594.
21. Nguyen, D.Q.; Vu, T.; Nguyen, A.T. BERTweet: A Pre-trained Language Model for English Tweets. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Online, 16–20 November 2020; pp. 9–14. [CrossRef]
22. Müller, M.; Salathé, M.; Kummervold, P.E. COVID-Twitter-BERT: A Natural Language Processing Model to Analyse COVID-19 Content on Twitter. *arXiv* **2020**, arXiv:2005.07503. [CrossRef] [PubMed]
23. Liang, B.; Li, A.; Zhao, J.; Gui, L.; Yang, M.; Yu, Y.; Wong, K.F.; Xu, R. Multi-modal Stance Detection: New Datasets and Model. *arXiv* **2024**, arXiv:2402.14298. [CrossRef]
24. 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; Volume 1, pp. 81–91.
25. Zhang, B.; Ma, J.; Fu, X.; Dai, G. Logic Augmented Multi-Decision Fusion Framework for Stance Detection on Social Media. *Inf. Fusion* **2025**, *122*, 103214. [CrossRef]
26. 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.
27. Ding, D.; Dong, L.; Huang, Z.; Xu, G.; Huang, X.; Liu, B.; Jing, L.; Zhang, B. EDDA: An Encoder-Decoder Data Augmentation Framework for Zero-Shot Stance Detection. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), Torino, Italia, 20–25 May 2024; pp. 5484–5494.
28. Fei, H.; Li, B.; Liu, Q.; Bing, L.; Li, F.; Chua, T.S. Reasoning Implicit Sentiment with Chain-of-Thought Prompting. *arXiv* **2023**, arXiv:2305.11255.
29. Ling, Z.; Fang, Y.; Li, X.; Huang, Z.; Lee, M.; Memisevic, R.; Su, H. Deductive Verification of Chain-of-Thought Reasoning. *arXiv* **2023**, arXiv:2306.03872.
30. Cai, Z.; Chang, B.; Han, W. Human-in-the-Loop through Chain-of-Thought. *arXiv* **2023**, arXiv:2306.07932.
31. Glandt, K.; Khanal, S.; Li, Y.; Caragea, D.; Caragea, C. Stance detection in COVID-19 tweets. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Long Papers), Online, 1–6 August 2021; Volume 1.
32. Augenstein, I.; Rocktaeschel, T.; Vlachos, A.; Bontcheva, K. Stance Detection with Bidirectional Conditional Encoding. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Sheffield, Austin, TX, USA, 1–5 November 2016.
33. Du, J.; Xu, R.; He, Y.; Gui, L. Stance Classification with Target-specific Neural Attention. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, Melbourne, Australia, 19–25 August 2017; pp. 3988–3994. [CrossRef]
34. Liang, B.; Fu, Y.; Gui, L.; Yang, M.; Du, J.; He, Y.; Xu, R. Target-Adaptive Graph for Cross-Target Stance Detection. In Proceedings of the WWW '21: The Web Conference 2021, Virtual Event/Ljubljana, Slovenia, 19–23 April 2021; pp. 3453–3464.
35. Allaway, E.; Srikanth, M.; Mckeown, K. Adversarial Learning for Zero-Shot Stance Detection on Social Media. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Online, 6–11 June 2021; pp. 4756–4767.
36. Allaway, E.; Mckeown, K. Zero-Shot Stance Detection: A Dataset and Model using Generalized Topic Representations. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 16–20 November 2020; pp. 8913–8931.

37. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.
38. 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.
39. Zhu, Q.; Liang, B.; Sun, J.; Du, J.; Zhou, L.; Xu, R. Enhancing Zero-Shot Stance Detection via Targeted Background Knowledge. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, 11–15 July 2022; pp. 2070–2075.
40. 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*, 159. [CrossRef]
41. Zhang, B.; Fu, X.; Ding, D.; Huang, H.; Li, Y.; Jing, L. Investigating Chain-of-thought with ChatGPT for Stance Detection on Social Media. *arXiv* **2023**, arXiv:2304.03087.
42. Dai, G.; Liao, J.; Zhao, S.; Fu, X.; Peng, X.; Huang, H.; Zhang, B. Large Language Model Enhanced Logic Tensor Network for Stance Detection. *Neural Netw.* **2025**, *183*, 106956. [CrossRef] [PubMed]

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.