Induce, Align, Predict: Zero-Shot Stance Detection via Cognitive Inductive Reasoning

Bowen Zhang[1], Jun Ma[1], Fuqiang Niu[2], Li Dong[1], Jinzhou Cao[1], Genan Dai[1*]

[1]School of Artificial Intelligence, Shenzhen Technology University
[2]School of Cyber Science and Technology, University of Science and Technology of China
zhang_bo_wen@foxmail.com, 2300411011@email.szu.edu.cn, nfq729@gmail.com, dongli@sztu.edu.cn, caojinzhou@sztu.edu.cn, daigenan@sztu.edu.cn

Abstract

Zero-shot stance detection (ZSSD) seeks to determine the stance of text toward previously unseen targets, a task critical for analyzing dynamic and polarized online discourse with limited labeled data. While large language models (LLMs) offer zero-shot capabilities, prompting-based approaches often fall short in handling complex reasoning and lack robust generalization to novel targets. Meanwhile, LLM-enhanced methods still require substantial labeled data and struggle to move beyond instance-level patterns, limiting their interpretability and adaptability. Inspired by cognitive science, we propose the Cognitive Inductive Reasoning Framework (CIRF), a schema-driven method that bridges linguistic inputs and abstract reasoning via automatic induction and application of cognitive reasoning schemas. CIRF abstracts first-order logic patterns from raw text into multi-relational schema graphs in an unsupervised manner, and leverages a schema-enhanced graph kernel model to align input structures with schema templates for robust, interpretable zero-shot inference. Extensive experiments on SemEval-2016, VAST, and COVID-19-Stance benchmarks demonstrate that CIRF not only establishes new state-of-the-art results, but also achieves comparable performance with just 30% of the labeled data, demonstrating its strong generalization and efficiency in low-resource settings.

Introduction

Zero-shot stance detection (ZSSD) aims to identify the stance of text towards targets unseen during training, a challenge increasingly important for analyzing polarized social media discourse as new topics rapidly emerge and labeled data remain scarce (Allaway and Mckeown 2020; Liang et al. 2022a). Recent advances in large language models (LLMs) offer new opportunities for ZSSD, as their zero-shot prompting and strong contextual understanding enable deeper semantic reasoning and generalization to novel targets (Binz and Schulz 2023). However, LLMs applied via prompting strategies (Li et al. 2023a; Zhang et al. 2023; Lan et al. 2024; Zhao et al. 2024; Weinzierl and Harabagiu 2024) typically exhibit suboptimal performance on ZSSD. In contrast, LLM-enhanced methods (LEM) combine the rich knowledge encoded in LLMs with the task-specific adaptation capabilities of supervised neural models, and have recently become a prominent research direction (Zhang et al. 2024; Dai et al. 2025; Zhang et al. 2025b). Nevertheless, these approaches still face two challenges: substantial reliance on labeled data and limited generalization beyond surface-level lexical cues (Wang et al. 2020).

On the other hand, cognitive science suggests that humans reason not only by memorizing specific examples, but also by abstracting generalizable reasoning schemas—structured templates that capture the logical relationships underlying diverse arguments (e.g., causality, value judgments, conditional inference) (Halpern 1990; Barwise 1977; Tenenbaum et al. 2011). For instance, arguments such as “increases health risks” and “reduces economic stability” both instantiate a schema of negative consequence, leading to *opposition*, regardless of surface wording or topical context. Such schemas disentangle reasoning logic from lexical or contextual specifics, enabling robust generalization to novel targets (Cocchi et al. 2013; Ragni and Knauff 2013). This cognitive insight motivates us to integrate schema-level reasoning into stance detection, aiming to overcome the data dependency and limited generalization of current instance-based approaches.

Reasoning schemas have shown strong potential for improving generalization and interpretability in domains such as question answering and causal inference (Peng et al. 2024; Li et al. 2024; Cheng et al. 2024; Su et al. 2025; Tao et al. 2025). However, their application to stance detection—especially in the zero-shot setting—remains largely unexplored. This is mainly due to two core challenges: (1) the lack of effective methods for modeling and extracting stance-specific reasoning schemas that can generalize across diverse targets; and (2) the absence of mechanisms for aligning input instances with abstract schemas to enable schema-guided inference. Existing neural and LLM-based approaches predominantly rely on instance-level lexical or contextual patterns, rarely inducing or leveraging reusable schema-level reasoning. This limits both their generalization and interpretability in ZSSD.

To address these challenges, we propose the Cognitive Inductive Reasoning Framework (CIRF), a schema-driven approach for ZSSD that bridges linguistic input and abstract reasoning via automatic schema induction and schema-

*Corresponding author.

guided inference. Specifically, CIRF comprises two key components: (1) Unsupervised Schema Induction (USI), which leverages LLMs to abstract structured reasoning patterns from raw text by converting predicates into first-order logic (FOL) expressions and clustering them into a multi-relational schema graph, capturing stance-relevant logical relations independent of specific targets or vocabulary; and (2) Schema-Enhanced Graph Kernel Model (SEGKM), which represents each input as an FOL graph, maps predicate nodes to corresponding schema nodes, and employs a learnable graph kernel to align input structures with schema templates for stance prediction. Unlike standard graph neural networks, which often struggle to capture reusable high-order reasoning motifs, our kernel-based approach enables direct and interpretable alignment between input reasoning structures and abstract schema templates. This explicit structural and semantic matching is essential for robust zero-shot generalization and interpretability in stance detection, and cannot be easily achieved by conventional neural architectures. This framework enables CIRF to generalize effectively to novel and diverse targets by combining the interpretability of symbolic schemas with the adaptability of neural inference, setting a new paradigm for schema-driven zero-shot stance detection.

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

- We propose the CIRF, a first schema-driven approach for ZSSD that bridges linguistic input and abstract reasoning via automatic induction and application of cognitive reasoning schemas.
- We introduce a unified framework combining unsupervised schema induction—leveraging LLMs to abstract FOL patterns into multi-relational schema graphs—and SEGKM for effective schema-guided stance inference and unseen target generalization.
- Extensive experiments on the SemEval-2016, VAST, and COVID-19-Stance benchmarks demonstrate that CIRF outperforms state-of-the-art ZSSD baselines and achieves competitive results with 70% fewer labeled examples than LLM-enhanced methods.

Related Work

ZSSD Methods. ZSSD has attracted increasing attention due to its importance in identifying stances toward previously unseen targets (Liang et al. 2022a). Early approaches, such as JointCL (Liang et al. 2022b) and TarBK (Zhu et al. 2022), rely heavily on supervised learning with large annotated datasets, limiting their generalization to novel targets. Recent advances in LLMs have introduced new paradigms for ZSSD, including zero-shot prompting LLMs (Zhang et al. 2023; Lan et al. 2024) and LLM-enhanced fine-tuned models (Li et al. 2023a; Zhang et al. 2024; Dai et al. 2025; Zhang et al. 2025b). However, zero-shot prompting LLMs often underperform due to their lack of task-specific adaptation, while LLM-enhanced fine-tuned models still require extensive instance-level supervision. These limitations highlight the need for frameworks that can generalize reasoning patterns to unseen targets without relying on large amounts of labeled data, thereby motivating schema-driven approaches.

First-Order Logic for Neural Reasoning. FOL provides a structured and interpretable foundation for encoding logical relations such as causality, implication, and conditionality, and has been widely adopted to enhance consistency and transparency in neural reasoning (Hu et al. 2016b; Huang et al. 2022). Recent methods integrate FOL constraints into neural architectures via posterior regularization (Hu et al. 2016a; Zhang et al. 2022) or joint FOL-neural embeddings, aiming to unify symbolic rigor with statistical flexibility. In stance detection, recent studies (Dai et al. 2025; Zhang et al. 2025b) prompt LLMs to generate FOL-based reasoning chains, achieving notable improvements over conventional models. However, such FOL-based techniques typically rely on instance-specific rules, limiting their ability to induce domain-agnostic abstractions or generalize logic across unseen topics—a critical bottleneck in zero-shot scenarios. Moreover, manually crafted or instance-specific FOL rules often struggle to capture high-level reasoning schemas that generalize across domains or topics, an aspect critical for ZSSD.

Schema Induction Methods. Schema induction has been widely explored in event-centric scenarios (Edwards and Ji 2023; Huang et al. 2016; Shen et al. 2021), where both bottom-up concept linking and top-down clustering are employed to capture relational patterns. The advent of LLMs has further enabled schema construction through generative and summarization capabilities (Li et al. 2023b; Tang et al. 2023; Dror, Wang, and Roth 2023; Shi et al. 2024; Rong et al. 2025; Finch, Josyula, and Choi 2025; Lee et al. 2025). However, most existing approaches focus on multi-sentence or event-level analysis, making them ill-suited for sentence-level stance detection, which requires inferring implicit and diverse semantic relations from less structured input. Among recent efforts, SenticNet8 (Cambria et al. 2024) and LogiMDF (Zhang et al. 2025b) are most relevant, aiming to abstract lexical items or induce logical rules via LLMs. Yet, these methods fundamentally operate at the word or predicate level and largely overlook the relational structures that govern inter-concept dependencies—such as causality or contradiction—that are vital for robust reasoning. As a result, they often yield semantically shallow or fragmented schema representations, limiting generalization and structural expressivity. In contrast, our CIRF treats entire FOL reasoning chains as the unit of schema induction. This allows us to abstract and transfer not only individual semantic items but also the relational patterns binding them, enabling more expressive and transferable schemas for zero-shot stance detection.

Method

Our CIRF tackles ZSSD through a two-stage pipeline that combines schema abstraction with graph-based neural inference. Specifically, as shown in Figure 1, CIRF consists of: (1) an USI module, which leverages large-scale unlabeled data and LLMs to automatically induce a library of abstract, multi-relational reasoning schemas; and (2) a SEGKM, which parses each input argument into a FOL graph, aligns it with the induced schemas, and predicts stance via learnable graph kernel matching.

Task Definition. Let $X = (x_i, q_i)_{i=1}^N$ represent the labeled data collection, where x refers to the input text and q corresponds to the source target. N denotes the total number of instances in X. Each sentence-target pair $(x, q) \in X$ is assigned a stance label y (e.g., *favor*, *against*, or *none*). ZSSD aims to train a model from source known targets and predict the stance polarity towards unseen targets.

USI: Unsupervised Schema Induction

To bridge the gap between specific instance-level reasoning and general, transferable inference patterns, we design an unsupervised schema induction pipeline that extracts abstract cognitive schemas from LLM-generated logical reasoning. Unlike prior approaches that directly cluster raw logical forms, our method involves a multi-stage reasoning analysis to ensure the resulting schemas are both meaningful and generalizable. Our pipeline consists of four main stages as follows.

FOL Reasoning Generation. For each sentence-target pair, we first prompt an LLM to generate a reasoning chain in the form of FOL.

FOL Interpretation and Abstraction. We further prompt the LLM to analyze the internal structure and inference logic of each generated FOL chain, and to produce alternative but logically equivalent FOL expressions that exhibit diverse syntactic or structural forms. By collecting these logically similar yet structurally varied formulations, we are able to expose the underlying reasoning strategy beyond surface realization. Subsequently, we prompt the LLM to summarize the shared reasoning pattern among these variants into a generalized FOL template, which captures the essential logic in a reusable and domain-independent manner. For example, consider the following case: $\forall x$, (is robot(x) $\rightarrow$ (helps humans(x) $\rightarrow$ (must be safe(x) $\wedge$ requires testing(x)))) $\rightarrow$ recommend regulation. After interpretation and abstraction, it will be generalized as: $\forall x$, ((is target(x) $\wedge$ meets condition(x)) $\rightarrow$ entails consequence(x)) $\rightarrow$ policy action.

Schema Clustering and Hierarchical Abstraction. We cluster the abstracted FOL templates into logical categories (schemas) by prompting the LLM to group them according to both semantic and inference pattern similarity. For each resulting schema, if a logic form does not fit any existing category, the LLM creates a new schema with a unique ID, name, and description. To capture the shared reasoning structure within each cluster, we prompt the LLM to synthesize a representative logic chain using both the cluster's description and its FOL members. For large clusters, we adopt a hierarchical strategy: FOLs are first grouped into fine-grained sub-clusters, each summarized into an intermediate logic chain, and these summaries are finally merged to form the schema-level template.

Schema Graph Construction. Finally, we represent the induced schemas as nodes in a multi-relational graph, with edges indicating logical relations such as causality, contrast, or implication. This multi-stage abstraction process distills diverse reasoning instances into a compact, interpretable, and transferable set of schema-level templates, enabling CIRF to robustly generalize reasoning patterns to previously unseen targets.

SEGKM: Schema-Enhanced Graph Kernel Model

To leverage both instance-specific logical reasoning and abstract schema knowledge for zero-shot stance detection, we propose the SEGKM, as illustrated in Figure 1. SEGKM is designed to integrate transferable, concept-level cognitive schemas directly into the reasoning process of each test instance. These schemas are induced from LLM rationales via the USI module. This enables the model to systematically align local reasoning structures with global, generalizable patterns, thus supporting interpretable and robust inference across unseen stance targets. While standard GNNs primarily rely on local message passing, SEGKM explicitly matches input reasoning structures to abstract schema templates, facilitating structure-level generalization and interpretability. Specifically, schema-driven reasoning in SEGKM proceeds as follows:

FOL Graph Construction. Given an input sentence-target pair (x, q), we first generate step-by-step reasoning rationales using LLM prompting (as in USI), and convert them into a FOL graph $G^f = (V^f, E^f)$, where nodes represent predicates and edges encode logical relations extracted from the input, such as implication, conjunction, or negation, corresponding to common FOL constructs.

Schema Knowledge as Graph Filters. To inject abstract reasoning motifs into node-level feature extraction, we initialize the graph kernel filters using local subgraphs of the induced cognitive schemas. Specifically, for each schema graph $G^{(j)}$, we extract a set of subgraph filters $H^{(j)} = \{H_1^{(j)}, \ldots, H_n^{(j)}\}$, where each filter $H_i^{(j)}$ is a subgraph centered on a schema node, retaining its neighborhood structure and relational context. Collecting filters from all schema graphs yields a filter pool $\mathcal{H} = \bigcup_j H^{(j)}$, where each group $H^{(j)}$ corresponds to a coherent, high-level reasoning schema. Compared to using the full schema graph, these subgraph filters efficiently capture fine-grained relationships while reducing computation, and enable the kernel network to match localized reasoning structures from input graphs to reusable schema patterns.

Kernel-Based Node Representation. To infuse schema-level reasoning patterns into the FOL-based input graph, we propose a kernel-based node representation method that integrates both structural and semantic similarity between the local subgraph around each node and pre-induced schema filters. Each node and edge embedding is initialized via a pretrained BERT encoder (Devlin et al. 2019), ensuring rich semantic information.

Let G_v^f denote the k-hop local subgraph centered at node $v \in V^f$, which serves as the local reasoning structure for v. To evaluate how well G_v^f aligns with a schema filter $H_i^{(j)}$, we define a semantic-aware deep kernel function inspired by the p-step random walk kernel (Feng et al. 2022), enhanced with edge features. Specifically, given the set of schema graphs $G^{(j)}$ and their subgraph filters $H_i^{(j)}$ constructed as

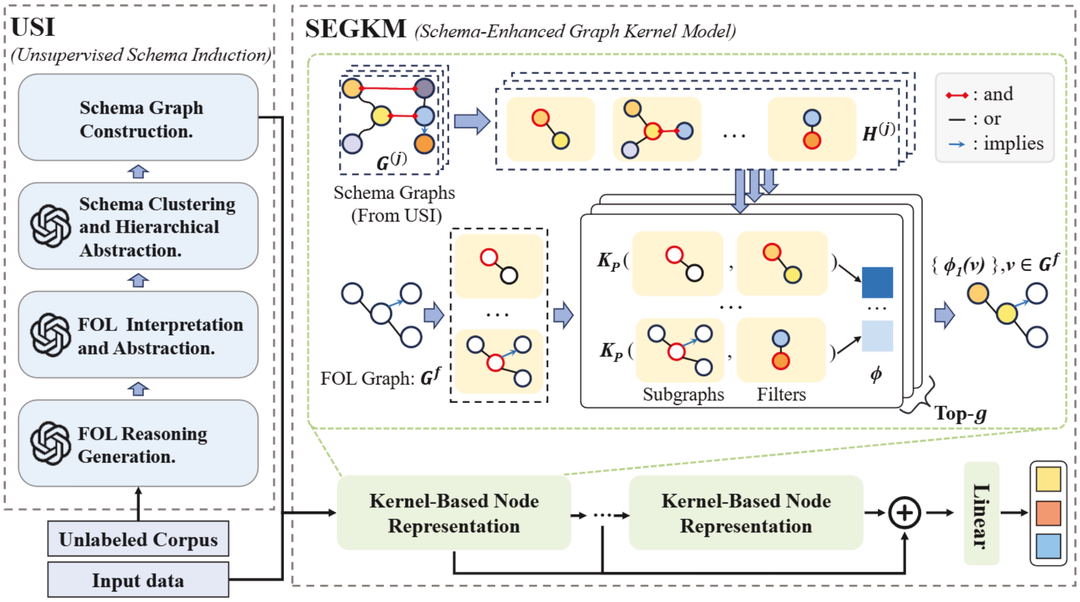

Figure 1: The framework of CIRF. The red-bordered circles represent the central nodes of subgraphs or filters.

described previously, inspired by Yin and Zhong (2024), we first encode edge semantics into node features via a relation-aware projection. For each node embedding x and its associated edge feature e, we define:

$$x' = \text{ReLU}\left(x + \text{Proj}(e)\right), \quad e \in \mathbb{R}^{|E|} \tag{1}$$

where $\text{Proj}(e)$ denotes a linear transformation of edge-type or relation embeddings. This mechanism ensures that both structural and relational patterns are captured during kernel evaluation. This operation is applied over all nodes in G_v^f and $H_i^{(j)}$, yielding semantic-aware node representations $\mathbf{X}'_{G_v^f}$ and $\mathbf{X}'_{H_i^{(j)}}$.

Next, we compute a similarity matrix based on these refined node features:

$$S = \mathbf{X}'_{G_v^f} \cdot (\mathbf{X}'_{H_i^{(j)}})^\top \tag{2}$$

Let $\mathbf{s} = \text{vec}(S) \in \mathbb{R}^{mn}$ denote the flattened similarity vector between all node pairs in G_v^f and $H_i^{(j)}$. We then define the deep kernel response as:

$$\phi_{1,i}(v) = K_p(G_v^f, H_i^{(j)}) = \mathbf{s}^\top W A_\times^p \mathbf{s} \tag{3}$$

Here, $A_\times^p$ denotes a learnable p-step transition matrix over the product graph of G_v^f and $H_i^{(j)}$, which captures higher-order structural alignment, and W is a learnable weight matrix that parametrizes the kernel response. This explicit product graph kernel formulation enables direct modeling of complex structural correspondences between input and schema. In contrast, standard GNNs mainly rely on local message passing and often struggle to capture reusable, high-level reasoning motifs. Note that while edge semantics are not directly injected into $A_\times^p$, they implicitly influence kernel computation through the edge-aware node features used in S.

Rather than selecting individual schema filters, we group filters by their originating schema graph and perform selection at the graph level. This grouping further reinforces the transferability and interpretability of schema-guided reasoning.

To compute the overall alignment between the input graph G^f and a schema graph $G^{(j)}$, we aggregate the average kernel responses of all constituent filters $H_i^{(j)}$ across every node $v \in V^f$:

$$S^{(j)} = \sum_{v \in V^f} \left(\frac{1}{|H^{(j)}|} \sum_{H_i^{(j)} \in H^{(j)}} \phi_{1,i}(v) \right) \tag{4}$$

It ensures that the schema score reflects both local compatibility at each node and global coverage across the graph, capturing the degree to which the reasoning pattern embedded in $G^{(j)}$ manifests throughout the input structure.

We then select the top-g schema graphs with the highest scores $S^{(j)}$ as the candidate schema set $\mathcal{G}^*$, which will provide the filter set for node representation. The filters associated with these selected schema graphs are aggregated to form the candidate reasoning motifs used for downstream node representation learning.

Finally, we construct the node representation $\phi_1(v)$ by concatenating the responses $\phi_{1,i}(v)$ for all filters $H_i^{(j)}$ belonging to the selected top-g schema graphs:

$$\phi_1(v) = \text{Concat}\left(\phi_{1,i}(v) \,\middle|\, H_i^{(j)} \in \bigcup_{G^{(j)} \in \mathcal{G}^*} H^{(j)} \right) \tag{5}$$

This yields a schema-aware node representation that reflects the most semantically and structurally relevant reasoning motifs. By explicitly aligning local reasoning patterns with high-level schemas through the graph kernel, our method supports interpretable and transferable inference, which is particularly advantageous for ZSSD.

Hierarchical Kernel-Based Graph Representation. Real-world stance reasoning often requires multi-hop logic and the flexible composition of multiple schema patterns. To support this, we stack multiple layers of schema-driven kernel feature extraction. At each layer l, node representations are recursively updated by matching their expanded local subgraphs to the schema filters, allowing deeper layers to aggregate schema knowledge from increasingly broader reasoning contexts. This design enables SEGKM to capture both fine-grained local logic and higher-order, composite reasoning motifs. The final graph representation is constructed by concatenating the aggregated node features from all layers:

$$\Phi(G^f) = \text{Concat}\left(\sum_{v \in G^f} \phi_l(v) \mid l = 0, 1, ..., L\right) \tag{6}$$

The final graph representation is used for stance prediction, and the model is trained end-to-end using cross-entropy loss.

Experimental Setups

Experimental Data. We evaluate CIRF on three widely used stance detection benchmarks: **SEM16** (SemEval-2016 Task 6; (Mohammad et al. 2016)) consists of tweets annotated for stance towards three targets: *Hillary Clinton* (HC), *Feminist Movement* (FM), and *Legalization of Abortion* (LA). **VAST** (Allaway and Mckeown 2020) contains New York Times opinion texts covering diverse topics, with 4,003 training, 383 development, and 600 test examples. Topic phrases are refined by human annotators to ensure quality. **COVID-19** (COVID-19-Stance; (Glandt et al. 2021)) includes data on four COVID-19-related targets: *Wearing a Face Mask* (WA), *Keeping Schools Closed* (SC), *Anthony S. Fauci, M.D.* (AF), and *Stay at Home Orders* (SH). All datasets follow standard splits and label settings from previous work.

Evaluation Metrics. For the SEM16 and COVID-19 datasets, which include three classes ("FAVOR", "AGAINST", and "NONE"), we follow Liang et al. (2022a) and report macro-F1 (F_{avg}) computed over the "FAVOR" and "AGAINST" categories. For the VAST dataset, which comprises three categories ("Pro", "Con", and "Neutral"), we follow Li, Zhao, and Caragea (2023) and report the macro-F1 across all classes, along with individual macro-F1 scores for the "Pro" and "Con" categories.

Implementation Details. We use GPT-3.5 as the default LLM for FOL elicitation and summary generation, following prior baselines. To evaluate robustness, we also test with GPT-4o and DeepSeek-v3 . Dataset splits follow Li et al. (2023a); Lan et al. (2024): for SEM16 and COVID-19, one target is held out for testing; for VAST, we use the official zero-shot setup. USI uses temperature 0 for LLM queries. SEGKM adopts a two-layer graph kernel, and applies a fully connected ReLU layer. Training (on a 40GB A100 GPU) uses AdamW (batch size 32, learning rate 5×10^{-4}), early stopping (patience 10), up to 20 epochs, and selects the checkpoint with the lowest validation loss (validation every 0.2 epoch). We report the average result over three independent runs.

Baseline Methods. To evaluate the performance of existing stance detection models on our dataset, we employed the following models: (1) Non-LLM methods: CrossNet (Du et al. 2017), JointCL (Liang et al. 2022b), TarBK (Zhu et al. 2022), and PT-HCL (Liang et al. 2022a). (2) Zero-shot prompting LLMs: GPT-3.5 (Zhang et al. 2023), COLA (Lan et al. 2024), and KASD (Li et al. 2023a). (3) LLM-enhanced fine-tuned models: KAI (Zhang et al. 2024), LCDA (Zhang et al. 2025a), FOLAR (Dai et al. 2025), and LogiMDF (Zhang et al. 2025b).

Experimental Results

Main ZSSD Experimental Results. The main ZSSD results are summarized in Table 1. CIRF consistently outperforms all baseline models across all three datasets, demonstrating the effectiveness of our approach in challenging zero-shot scenarios. For example, CIRF achieves an average F1 score of 76.2 on SEM16 and 80.9 on VAST, representing a 1.9-point improvement over FOLAR and a 0.6-point gain over LCDA, respectively, and surpassing LogiMDF, the most competitive method on the COVID-19 dataset, by 3.7 points. Statistical significance tests ($p < 0.05$) confirm that CIRF's improvements over FOLAR and LogiMDF are significant.

A closer examination reveals several trends: First, traditional non-LLM methods perform substantially worse on ZSSD, while LLM-based models yield notable gains, emphasizing the crucial role of LLMs' reasoning capabilities. Furthermore, fine-tuned LLM-based models (KAI, LogiMDF, FOLAR, and CIRF) generally surpass direct prompting strategies (GPT-3.5, COLA, and KASD), suggesting that combining annotated data with the knowledge encoded in LLMs is more effective.

Breaking down by dataset, CIRF's advantage is most pronounced on VAST, which features greater topic diversity and fine-grained targets, showcasing its robustness in real-world scenarios. On SEM16 and COVID-19, where the targets are more constrained, all LLM-based models perform more closely, but CIRF still consistently ranks first.

We also compare different knowledge representations. Models using FOL knowledge elicited from LLMs (CIRF and FOLAR) generally outperform those using natural language (KAI), indicating that FOL provides a more compact and effective abstraction for reasoning transfer. Notably, CIRF further outperforms KAI, the strongest prior LLM-based method, highlighting the added value of our cognitive framework enhancement.

These results collectively demonstrate that CIRF not only sets a new state-of-the-art for ZSSD but also provides insights into the importance of structured LLM-elicited knowledge and cognitive schema design for robust zero-shot stance reasoning. Overall, this validates our hypothesis that explicit schema-guided abstraction enables better generalization to unseen targets and diverse domains.

	Model	SEM16			VAST (100%)			VAST (10%)			COVID-19			
		HC	FM	LA	Pro	Con	All	Pro	Con	All	AF	SC	SH	WA
Glove	CrossNet	38.3	41.7	38.5	46.2	43.4	43.4	37.3	32.9	36.2	41.3	40.0	40.4	38.2
Bert	JointCL	54.4	54.0	50.0	64.9	63.2	71.2	53.8	57.1	65.5	57.6†	49.3†	43.5†	63.1†
Bert	TarBK	55.1	53.8	48.7	65.7	63.9	73.6	-	-	-	-	-	-	-
Bert	PT-HCL	54.5	54.6	50.9	61.7	63.5	71.6	-	-	-	-	-	-	-
GPT-3.5	GPT-3.5	78.9	68.3	62.3	63.8	56.8	65.1	63.8	56.8	65.1	69.2†	43.5†	66.5†	57.8†
GPT-3.5	COLA	81.7	63.4	71.0	-	-	73.0	-	-	73.0	65.7†	46.6†	53.5†	73.9†
GPT-3.5	KASD	80.3	70.4	62.7	-	-	67.0	-	-	-	-	-	-	-
GPT-3.5	KAI	76.4	73.7	69.4	66.7	73.0	76.3	63.5	74.0	75.2	-	-	-	-
GPT-3.5	LCDA	79.8	70.0	69.4	73.8	75.9	80.3	72.1	77.0	80.1	-	-	-	-
GPT-3.5	FOLAR†	**81.9**	71.2	69.9	71.2	75.8	77.2	70.1	75.5	76.7	69.5	67.2	65.4	73.1
GPT-3.5	LogiMDF†	75.1	67.9	68.0	-	-	76.7	-	-	76.6	70.4	68.8	64.9	75.4
GPT-3.5	**CIRF‡**	80.1	**74.7**	**73.9**	**74.1**	**78.5**	**80.9**	**73.4**	**78.4**	**80.7**	**74.1**	**70.3**	**68.8**	**81.0**
	DeepSeek-v3	80.7	78.3	77.7	75.0	74.3	75.9	75.0	74.3	75.9	76.3	74.3	77.1	87.6
	GPT-4o	81.6	79.8	77.6	77.6	78.7	80.0	77.6	78.7	80.0	82.7	78.8	76.9	89.2
	CIRF (DeepSeek-v3)	83.8	79.8	78.2	76.4	77.1	80.3	76.2	76.8	80.1	78.5	78.8	78.3	89.1
	CIRF (GPT-4o)	83.2	80.4	78.2	78.8	80.1	82.8	78.6	79.6	82.5	84.9	80.5	78.6	89.4

Table 1: Comparison of different models on the ZSSD task. The best scores are highlighted in bold, and the second-best scores are underlined. ‡ indicates that CIRF outperforms FOLAR, LCDA, and LogiMDF in macro-F1 across targets (paired t-test, $p < 0.05$). † denotes results reproduced using the official open-source code. The lower part of the table represents CIRF evaluated with more powerful LLM backbones to examine scalability.

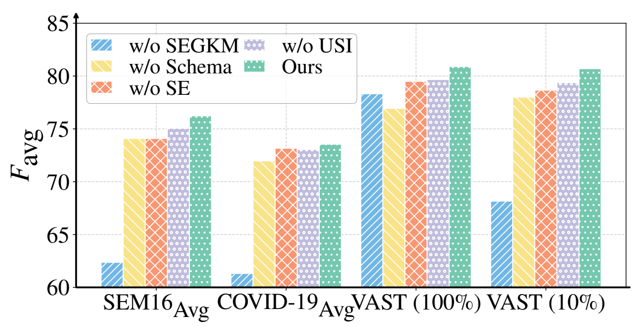

Figure 2: Ablation study results. Here, $_{Avg}$ denotes the average performance.

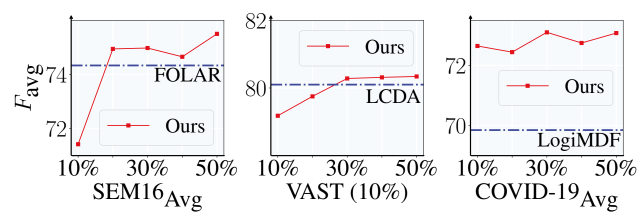

Figure 3: Performance comparison across different training scales. Dashed lines denote the strongest baseline method for each dataset.

To further investigate the scalability of CIRF, we evaluate it with stronger LLM backbones, including DeepSeek-v3 and GPT-4o. CIRF's performance rises notably with these advanced models: for example, switching from GPT-3.5 to GPT-4o increases the average F1 on VAST from 80.9 to 82.8, and on WA from 81.0 to 89.4. Similar gains are observed with DeepSeek-v3, confirming that CIRF can effectively capitalize on the enhanced reasoning capabilities of newer LLMs. These results highlight CIRF's scalability and suggest that combining schema-guided abstraction with continually improving LLMs is a promising direction for future zero-shot stance detection research.

Ablation Study. Figure 2 presents ablation results on SEM16, COVID-19, and VAST, quantifying the contribution of each key component in CIRF. Removing the cognitive schema (*w/o Schema*) or the SEGKM (*w/o SEGKM*) leads to the most substantial performance drops across all benchmarks, underscoring the central role of explicit schema-guided reasoning and relational logic encoding in robust zero-shot transfer. We also observe that eliminating edge semantics (*w/o SE*) or replacing LLM-based schema induction with simple clustering (*w/o USI*) consistently degrades results, though to a lesser extent, highlighting the importance of both rich relational structure and semantically grounded schema construction. Notably, the performance gaps become even wider under low-resource settings such as VAST (10%), emphasizing that each component is critical for generalization in challenging domains. Overall, these findings confirm that the synergy of schema abstraction, semantic relations, and LLM-driven structure underpins CIRF's effectiveness in zero-shot stance detection.

Low-Resource Performance. To assess data efficiency under limited supervision, we evaluate CIRF on VAST (10%), SEM16, and COVID-19 with varying proportions of labeled data. As shown in Figure 3, CIRF consistently and substantially outperforms strong LLM-based baselines (FOLAR, LCDA, LogiMDF) across all training scales. On

COVID-19, CIRF outperforms LogiMDF by 2.8 points with only 10% of supervision. Similarly, on SEM16, CIRF exceeds FOLAR by 0.6 points with just 20% of labeled data. On VAST (10%), CIRF steadily improves with more supervision and begins to outperform LCDA at 30%. The performance gap against most baselines remains stable or widens as supervision increases. These results demonstrate that the abstract cognitive schema, constructed from unlabeled data in other domains, enables CIRF to achieve strong cross-domain generalization with minimal supervision. Our findings highlight the power of explicit, domain-agnostic schema abstraction for robust and data-efficient stance detection in low-resource settings.

Case Study. We present an example illustrating how schema-guided graph reasoning enables correct stance prediction under mixed sentiment (Figure 4). While the text overall supports body cameras, it contains a hypothetical cost concern introduced by an adversative "BUT." Baseline models may be misled by this structure and overemphasize the cost issue, leading to misclassification. In contrast, our schema graph aligns key arguments with salient, domain-agnostic nodes such as *prevents harm*, *ensures collective responsibility*, and *mitigates risks*. This structured reasoning allows CIRF to robustly capture the author's supportive stance, even in nuanced or hedged cases.

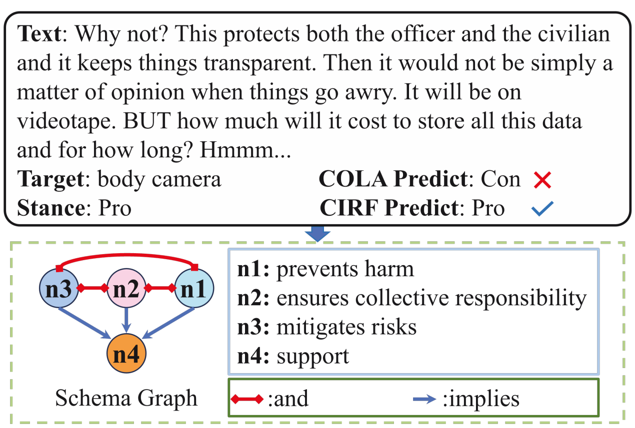

Figure 4: Case Study.

Effect of the Number of Induced Schemas. We investigate the effect of varying the number of schemas on CIRF's zero-shot stance detection performance. As illustrated in Figure 5, the model's performance remains highly stable across a wide range of schema counts, with only minor fluctuations (generally less than 1 point) observed for each dataset. This robustness indicates that CIRF is largely insensitive to the schema number: a moderate schema set is sufficient to capture key reasoning patterns, and further increasing the number yields little additional benefit or risk of instability. These results support our hypothesis that stance reasoning can be well-abstracted by a structured, logic-inspired schema set, and demonstrate that CIRF is practical and easy to tune in real-world scenarios.

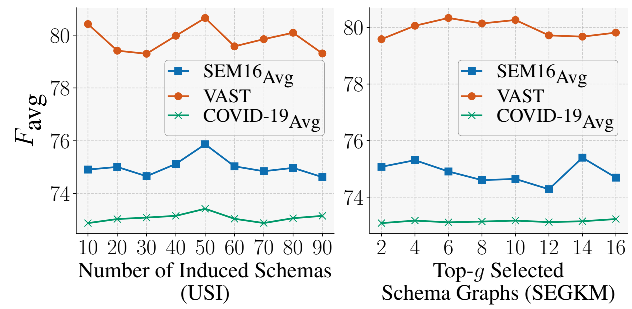

Figure 5: Impact of schema number (left) and selected schema graph number (right).

Effect of Top-g Schema Graph Selection. We analyze how varying the number of schema graphs dynamically selected during inference affects CIRF's zero-shot stance detection performance. As shown in Figure 5, performance across all datasets remains highly stable as the number of selected schema graphs increases from 2 to 16, with only minor fluctuations observed for each dataset. The absence of any clear trend or performance drop demonstrates that CIRF is robust to this hyperparameter, and that a modest number of selected schema graphs is sufficient to capture the essential reasoning patterns for each instance. Increasing the number beyond this range offers little additional benefit. These findings support our design hypothesis that stance reasoning can be abstracted by a compact set of logic-based schema graphs, and further simplify model tuning and deployment, confirming that CIRF is practical and easy to adapt to diverse domains and resource settings.

Conclusion

In this work, we introduce the CIRF, a schema-driven approach for zero-shot stance detection that bridges linguistic input and abstract reasoning through automatic schema induction and schema-guided inference. By leveraging unsupervised abstraction of first-order logic patterns and schema-enhanced graph kernel alignment, CIRF enables robust generalization and interpretability across diverse and previously unseen targets. Extensive experiments on multiple benchmarks demonstrate that CIRF not only achieves new state-of-the-art performance, but also maintains competitive results with minimal labeled data, highlighting its practical value in low-resource settings. Despite these advances, our framework still faces challenges, such as scaling schema induction for extremely large or noisy corpora, and integrating richer world knowledge or multimodal signals. Future work includes exploring more efficient schema induction algorithms, adaptive schema selection strategies, and extending CIRF to broader reasoning tasks in natural language understanding. Looking ahead, we will refine cognitive schema generation and extend CIRF to broader linguistic and cultural contexts to further improve its generalizability and real-world effectiveness.

Acknowledgements

This research is supported by the National Natural Science Foundation of China No. 62306184, Natural Science Foundation of Top Talent of SZTU (grant no. GDRC202320), and Shenzhen Science and Technology Program (No. JCYJ20240813113218025).

References

Allaway, E.; and Mckeown, K. 2020. 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)*, 8913–8931.

Barwise, J. 1977. An introduction to first-order logic. In *Studies in Logic and the Foundations of Mathematics*, volume 90, 5–46. Elsevier.

Binz, M.; and Schulz, E. 2023. Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences*, 120(6): e2218523120.

Cambria, E.; Zhang, X.; Mao, R.; Chen, M.; and Kwok, K. 2024. SenticNet 8: Fusing emotion AI and commonsense AI for interpretable, trustworthy, and explainable affective computing. In *International Conference on Human-Computer Interaction*, 197–216. Springer.

Cheng, Z.-Q.; Dong, Y.; Shi, A.; Liu, W.; Hu, Y.; O'Connor, J.; Hauptmann, A. G.; and Whitefoot, K. 2024. SHIELD: LLM-Driven Schema Induction for Predictive Analytics in EV Battery Supply Chain Disruptions. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*, 303–333.

Cocchi, L.; Zalesky, A.; Fornito, A.; and Mattingley, J. B. 2013. Dynamic cooperation and competition between brain systems during cognitive control. *Trends in cognitive sciences*, 17(10): 493–501.

Dai, G.; Liao, J.; Zhao, S.; Fu, X.; Peng, X.; Huang, H.; and Zhang, B. 2025. Large Language Model Enhanced Logic Tensor Network for Stance Detection. *Neural Networks*, 183: 106956.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186.

Dror, R.; Wang, H.; and Roth, D. 2023. Zero-Shot On-the-Fly Event Schema Induction. In *Findings of the Association for Computational Linguistics: EACL 2023*, 693–713.

Du, J.; Xu, R.; He, Y.; and Gui, L. 2017. Stance Classification with Target-specific Neural Attention. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, 3988–3994.

Edwards, C.; and Ji, H. 2023. Semi-supervised New Event Type Induction and Description via Contrastive Loss-Enforced Batch Attention. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, 3805–3827.

Feng, A.; You, C.; Wang, S.; and Tassiulas, L. 2022. Kergnns: Interpretable graph neural networks with graph kernels. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, 6614–6622.

Finch, J. D.; Josyula, Y.; and Choi, J. D. 2025. Generative Induction of Dialogue Task Schemas with Streaming Refinement and Simulated Interactions. *arXiv preprint arXiv:2504.18474*.

Glandt, K.; Khanal, S.; Li, Y.; Caragea, D.; and Caragea, C. 2021. 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 (Volume 1: Long Papers)*, 1596–1611.

Halpern, J. Y. 1990. An analysis of first-order logics of probability. *Artificial intelligence*, 46(3): 311–350.

Hu, Z.; Ma, X.; Liu, Z.; Hovy, E.; and Xing, E. 2016a. Harnessing Deep Neural Networks with Logic Rules. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2410–2420.

Hu, Z.; Yang, Z.; Salakhutdinov, R.; and Xing, E. 2016b. Deep neural networks with massive learned knowledge. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 1670–1679.

Huang, H.; Zhang, B.; Jing, L.; Fu, X.; Chen, X.; and Shi, J. 2022. Logic tensor network with massive learned knowledge for aspect-based sentiment analysis. *Knowledge-Based Systems*, 257: 109943.

Huang, L.; Cassidy, T.; Feng, X.; Ji, H.; Voss, C.; Han, J.; and Sil, A. 2016. Liberal event extraction and event schema induction. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 258–268.

Lan, X.; Gao, C.; Jin, D.; and Li, Y. 2024. Stance detection with collaborative role-infused llm-based agents. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, 891–903.

Lee, A. J.; Webb, T.; Bihl, T.; Holyoak, K.; and Lu, H. 2025. Few-Shot Learning of Visual Compositional Concepts through Probabilistic Schema Induction. *arXiv preprint arXiv:2505.09859*.

Li, A.; Liang, B.; Zhao, J.; Zhang, B.; Yang, M.; and Xu, R. 2023a. Stance Detection on Social Media with Background Knowledge. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 15703–15717.

Li, S.; Reddy, R. G.; Nguyen, K.; Wang, Q.; Fung, Y.; Han, C.; Han, J.; Natarajan, K.; Voss, C.; and Ji, H. 2024. Schema-Guided Culture-Aware Complex Event Simulation with Multi-Agent Role-Play. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 372–381.

Li, S.; Zhao, R.; Li, M.; Ji, H.; Callison-Burch, C.; and Han, J. 2023b. Open-Domain Hierarchical Event Schema Induction by Incremental Prompting and Verification. In *Proceed-*

ings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 5677–5697.

Li, Y.; Zhao, C.; and Caragea, C. 2023. Tts: A target-based teacher-student framework for zero-shot stance detection. In *Proceedings of the ACM Web Conference 2023*, 1500–1509.

Liang, B.; Chen, Z.; Gui, L.; He, Y.; Yang, M.; and Xu, R. 2022a. Zero-Shot Stance Detection via Contrastive Learning. In *Proceedings of the ACM Web Conference 2022*, 2738–2747.

Liang, B.; Zhu, Q.; Li, X.; Yang, M.; Gui, L.; He, Y.; and Xu, R. 2022b. 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)*, volume 1, 81–91. Association for Computational Linguistics.

Mohammad, S.; Kiritchenko, S.; Sobhani, P.; Zhu, X.; and Cherry, C. 2016. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, 31–41.

Peng, S.; Hu, X.; Yi, Q.; Zhang, R.; Guo, J.; Huang, D.; Tian, Z.; Chen, R.; Du, Z.; Guo, Q.; et al. 2024. Hypothesis, verification, and induction: grounding large language models with self-driven skill learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 14599–14607.

Ragni, M.; and Knauff, M. 2013. A theory and a computational model of spatial reasoning with preferred mental models. *Psychological review*, 120(3): 561.

Rong, H.; Chen, Z.; Lu, Z.; Xu, X.-k.; Huang, K.; and Sheng, V. S. 2025. Pred-ID: Future event prediction based on event type schema mining by graph induction and deduction. *Information Fusion*, 117: 102819.

Shen, J.; Zhang, Y.; Ji, H.; and Han, J. 2021. Corpus-based Open-Domain Event Type Induction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 5427–5440.

Shi, X.; Xue, S.; Wang, K.; Zhou, F.; Zhang, J.; Zhou, J.; Tan, C.; and Mei, H. 2024. Language models can improve event prediction by few-shot abductive reasoning. *Advances in Neural Information Processing Systems*, 36.

Su, Y.; Zhang, H.; Zhang, G.; Wang, Y.; Fan, Y.; Li, R.; and Wang, Y. 2025. Enhancing Event Causality Identification with LLM Knowledge and Concept-Level Event Relations. In *Proceedings of the 31st International Conference on Computational Linguistics*, 7403–7414.

Tang, J.; Lin, H.; Li, Z.; Lu, Y.; Han, X.; and Sun, L. 2023. Harvesting Event Schemas from Large Language Models. *arXiv preprint arXiv:2305.07280*.

Tao, Z.; Jin, Z.; Zhang, Y.; Chen, X.; Zhao, H.; Li, J.; Liang, B.; Tao, C.; Liu, Q.; and Wong, K.-F. 2025. A comprehensive evaluation on event reasoning of large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 25273–25281.

Tenenbaum, J. B.; Kemp, C.; Griffiths, T. L.; and Goodman, N. D. 2011. How to grow a mind: Statistics, structure, and abstraction. *science*, 331(6022): 1279–1285.

Wang, Z.; Sun, Q.; Li, S.; Zhu, Q.; and Zhou, G. 2020. Neural stance detection with hierarchical linguistic representations. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28: 635–645.

Weinzierl, M.; and Harabagiu, S. 2024. Tree-of-Counterfactual Prompting for Zero-Shot Stance Detection. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 861–880.

Yin, S.; and Zhong, G. 2024. Textgt: A double-view graph transformer on text for aspect-based sentiment analysis. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 19404–19412.

Zhang, B.; Ding, D.; Huang, Z.; Li, A.; Li, Y.; Zhang, B.; and Huang, H. 2024. Knowledge-Augmented Interpretable Network for Zero-Shot Stance Detection on Social Media. *IEEE Transactions on Computational Social Systems*.

Zhang, B.; Fu, X.; Ding, D.; Huang, H.; Li, Y.; and Jing, L. 2023. Investigating Chain-of-thought with ChatGPT for Stance Detection on Social Media. *arXiv preprint arXiv:2304.03087*.

Zhang, B.; Huang, X.; Huang, Z.; Huang, H.; Zhang, B.; Fu, X.; and Jing, L. 2022. Sentiment interpretable logic tensor network for aspect-term sentiment analysis. In *Proceedings of the 29th International Conference on Computational Linguistics*, 6705–6714.

Zhang, B.; Li, X.; Ma, J.; Zhang, X.; Dai, G.; and Ye, J. 2025a. Zero-shot Stance Detection with Logically Consistent Data Augmentation. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

Zhang, B.; Ma, J.; Fu, X.; and Dai, G. 2025b. Logic Augmented Multi-Decision Fusion framework for stance detection on social media. *Information Fusion*, 103214.

Zhao, C.; Li, Y.; Caragea, C.; and Zhang, Y. 2024. ZeroStance: Leveraging ChatGPT for Open-Domain Stance Detection via Dataset Generation. In *Findings of the Association for Computational Linguistics ACL 2024*, 13390–13405.

Zhu, Q.; Liang, B.; Sun, J.; Du, J.; Zhou, L.; and Xu, R. 2022. 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*, 2070–2075.