LLM-Driven Implicit Target Augmentation and Fine-Grained Contextual Modeling for Zero-Shot and Few-Shot Stance Detection

Yanxu Ji[1], **Jinzhong Ning**[1]*, **Yijia Zhang**[1]*,
Zhi Liu[1], **Hongfei Lin**[2]
[1] Dalian Maritime University, [2] Dalian University of Technology

Abstract

Stance detection aims to identify the attitude expressed in text towards a specific target. Recent studies on zero-shot and few-shot stance detection focus primarily on learning generalized representations from explicit targets. However, these methods often neglect implicit yet semantically important targets and fail to adaptively adjust the relative contributions of text and target in light of contextual dependencies. To overcome these limitations, we propose a novel two-stage framework: First, a data augmentation framework named Hierarchical Collaborative Target Augmentation (HCTA) employs Large Language Models (LLMs) to identify and annotate implicit targets via Chain-of-Thought (CoT) prompting and multi-LLM voting, significantly enriching training data with latent semantic relations. Second, we introduce DyMCA, a Dynamic Multi-level Context-aware Attention Network, integrating a joint text-target encoding and a content-aware mechanism to dynamically adjust text-target contributions based on context. Experiments on the benchmark dataset demonstrate that our approach achieves state-of-the-art results, confirming the effectiveness of implicit target augmentation and fine-grained contextual modeling. Our code is publicly available at https://github.com/EliaukoaYoW/DyMCA.

1 Introduction

Stance detection aims to identify the stance (e.g., *Favor*, *Against*, or *Neutral*) that a speaker holds toward a specific target (e.g., a person, policy, or concept) within a text (Mohammad et al., 2016; Küçük and Can, 2020; ALDayel and Magdy, 2021). By uncovering latent attitudinal tendencies in the text, stance detection facilitates a deeper understanding of textual information and is particularly crucial in analyzing significant events, such as public policy debates or presidential elections. Early research

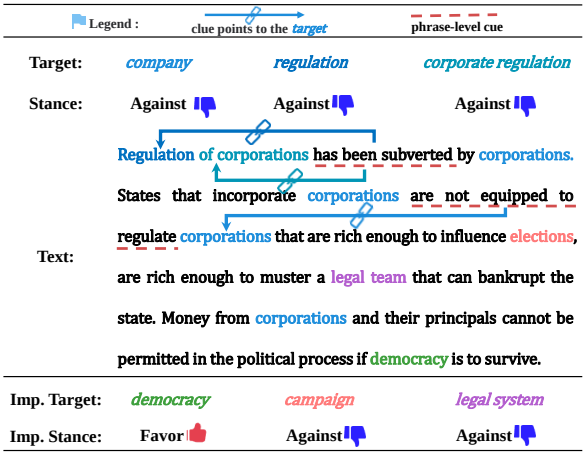

Figure 1: Example from VAST illustrating the omission of implicit targets. Explicitly discussed targets with clear textual cues are annotated, whereas implicit but semantically important targets are overlooked.

in stance detection can primarily be divided into two categories: in-target stance detection and cross-target stance detection. **In-target stance detection** focuses on in-depth analysis of a specific target (Hasan and Ng, 2014; Mohammad et al., 2016; Graells-Garrido et al., 2020), where the data for model training and testing both share identical target (e.g., "*Donald Trump*"). **Cross-target stance detection** (Augenstein et al., 2016; Xu et al., 2018; Wei and Mao, 2019), which emphasizes transfer learning between semantically related but distinct targets (e.g., training on "*Donald Trump*" and testing on "*Hillary Clinton*"). However, it is difficult to acquire all possible (in-target) or related (cross-target) targets in real scenarios. Therefore, Allaway and McKeown (2020) introduces zero-shot stance detection (ZSSD) and few-shot stance detection (FSSD), which aim to predict stances of unseen targets with little or no annotated data, promoting the field toward more practical direction.

Current research in ZSSD and FSSD primarily focus on learning stance representations of known

*corresponding author.

Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, pages 5884–5896
November 4-9, 2025

targets through various techniques, including attention mechanisms (Allaway and McKeown, 2020), knowledge graphs (Liu et al., 2021), and contrastive learning (Liang et al., 2022a), and then generalize the learned features to unseen targets. While these approaches have achieved notable success, two fundamental limitations remain. **Issue1: Under-utilization of Implicit Targets.** Existing dataset annotation guidelines prioritize explicitly-discussed targets with clear textual cues, systematically omitting implicitly-inferred yet semantically crucial targets (e.g., the target "*democracy*" in Figure 1). This omission may hamper the model's ability to learn rich semantic representations, weakening its capacity for semantic inference, especially in handling indirect scenarios like metaphors or irony. **Issue2: Insufficient Contextual Adaptation Capability of Models.** As stance prediction inherently relies on both the text and its associated target, it is imperative to dynamically calibrate their respective roles according to contextual dependencies. For instance, when a text explicitly expresses its stance, the target information may serves only as an auxiliary cue, while in implicit stance cases, the target's semantic guidance becomes crucial. However, existing state-of-the-art methods (Li et al., 2023) typically employ fixed-weight interactions, which fail to capture such contextual variation. This limitation reduces the adaptability of the models and hinders their ability to generalize to unseen targets.

To address the aforementioned issues, we propose solutions along two complementary axes: target augmentation and model design.

For issue1, we propose an LLM-based **HCTA** (**H**ierarchical **C**ollaborative **T**arget **A**ugmentation) framework that leverages the semantic reasoning capabilities of Large Language Models (**LLMs**) to automatically discover and utilize implicit target information in existing dataset. The framework operates through two collaborative stages: **First**, we employ Chain-of-Thought (**CoT**) prompting to guide an LLM in identifying implicit targets from the given text that were missed during annotation; **Then** multiple LLMs work in parallel to predict stances for these newly identified targets, with their predictions subsequently aggregated via voting to mitigate individual LLM bias. This strategy facilitates to make previously overlooked information accessible for model training, while enhancing the model's semantic reasoning capabilities.

For issue2, we propose **DyMCA** (**Dy**namic **M**ulti-level **C**ontext-aware **A**ttention Network) to enhance the contextual adaptability of text-target interactions. DyMCA adopts a progressive strategy that transitions from global comprehension to localized refinement: (1) **The Global module** constructs a joint semantic representation to captures holistic text-target relations; (2) **The Local module** introduces a Content-aware Mechanism (**CM**) to dynamically adjust the relative influence of text and target according to contextual demands. This hierarchical architecture not only enhances the model's sensitivity to contextual variation but also improves its generalization to unseen targets while maintaining robust performance under resource-constrained conditions.

Our contributions are summarized as follows:

- We propose a novel LLM-driven framework (HCTA) that automatically identifies and leverages implicit target information through CoT prompting and multi-LLM voting [1].
- We propose the Dynamic Multi-level Context-aware Attention Network (DyMCA) that dynamically adjusts text-target interactions weights through fine-grained contextual modeling, enabling adaptive allocation of text and target influence based on context, improving model performance.
- Extensive experiments on benchmark datasets show that our proposed methods achieves state-of-the-art performance on both ZSSD and FSSD tasks. Furthermore, we extend our methods to related challenges to validate their versatility and effectiveness.

2 Related Work

Zero-Shot and Few-Shot Stance Detection focuses on identifying stances toward unseen or rarely seen targets. Allaway and McKeown (2020) constructed the VAried Stance Topics (VAST) dataset for ZSSD, which contains thousands of targets spanning diverse domains including politics and education. Building upon VAST, Allaway and McKeown (2020) employed an unsupervised clustering method to learn generalizable target representations. Liu et al. (2021) incorporated a commonsense knowledge graph and utilized Graph Convolution Networks (GCN) to model implicit connec-

[1]To the best of our knowledge, this represents the first work in stance detection that explicitly addresses and utilizes implicit target information requiring deep semantic reasoning.

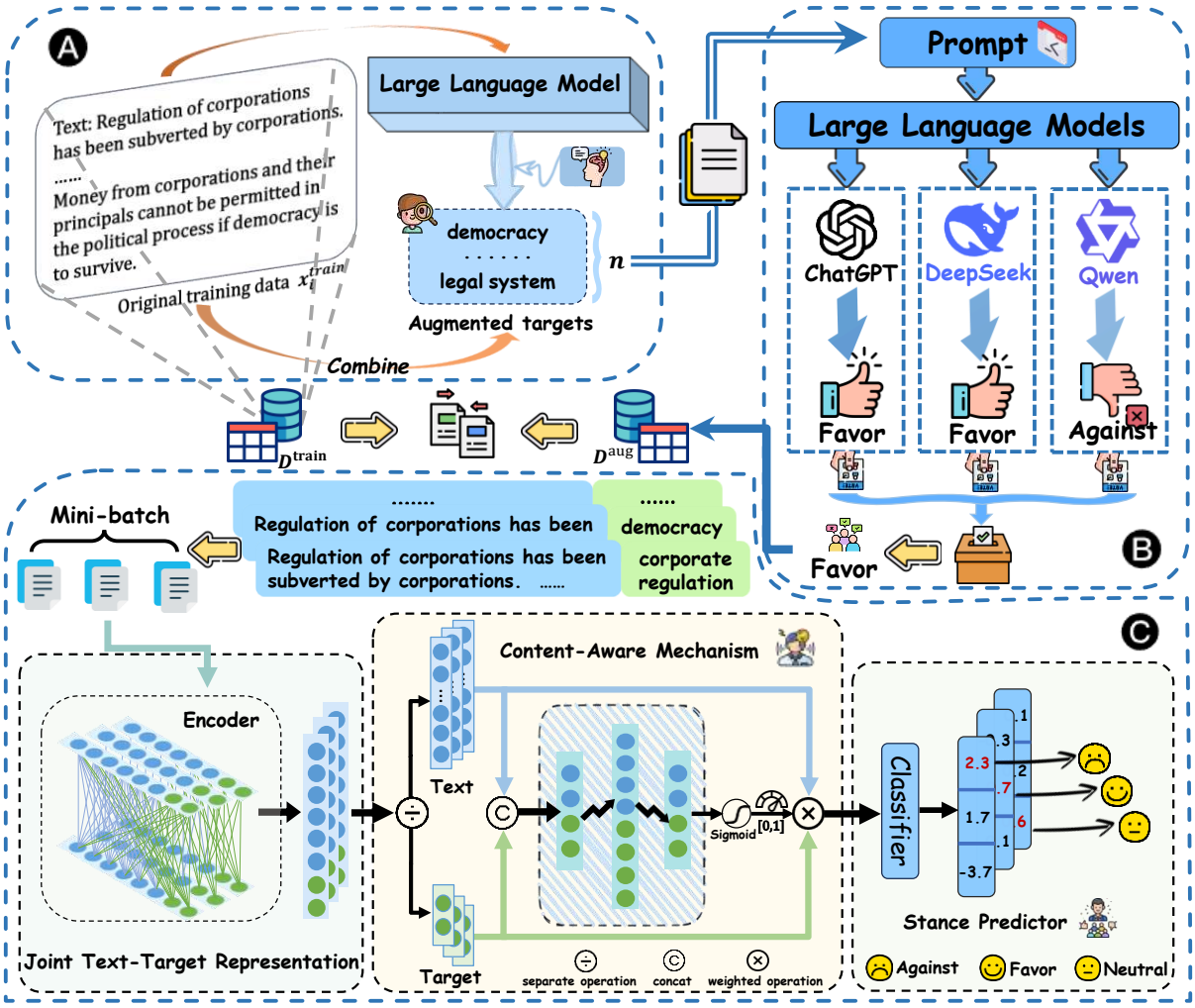

Figure 2: Overview of the proposed framework and model. (a) and (b) illustrate the HCTA process, where LLMs are guided to identify implicit targets and collaboratively assign pseudo-labels via prompt-based reasoning. (c) shows the DyMCA architecture, which employs a global awareness to localized focus for context-aware stance prediction.

tions between texts and targets. Liang et al. (2022a) leveraged contrastive learning to model both target-invariant and target-specific stance features. Despite progress in modeling unseen targets, a key limitation in ZSSD remains the limited availability of annotated data. To alleviate this, recent work has turned to data augmentation as a promising direction. Zhang et al. (2023) constructed augmented instances by extracting target-relevant text segments. Li et al. (2023) proposed extracting additional targets via small pre-trained model and assigns pseudo labels through a self-training. Most recently, Zhang et al. (2024) utilized LLMs to extract target-text relational knowledge for context enrichment. Notably, our work makes the first systematic attempt to leverage LLMs for discovering implicit targets and corresponding underlying stance information for enhanced model training. This fundamentally differs from prior approaches by: (1) leveraging CoT reasoning to identify annotation-omitted yet semantically crucial implicit targets, and (2) establishing a multi-LLMs collaborative framework to ensure prediction reliability, ultimately advancing both performance and generalization.

3 Methods

3.1 Task Description

Formally, let $D = \{(x_i, t_i, y_i)_{i=1}^{N}\}$ denotes the training dataset which contains N examples, where x_i is a text, t_i is the corresponding target, and y_i is the associated stance label. The goal of the zero-shot and few-shot stance detection is to infer stance label y_i given (x_i, t_i) pairs, where zero-shot assumes that test examples are entirely unseen during training, whereas few-shot allows limited.

Text	Target
Regulation of corporations has been subverted by corporations. States that incorporate corporations are not equipped to regulate corporations that are rich enough to influence elections, are rich enough to muster a legal team that can bankrupt the state. Money from corporations and their principals cannot be permitted in the political process if democracy is to survive.	regulation political process democracy legal system

Table 1: Sample Predictions using LLM for Implicit Target Extraction. This table compares the performance of the model in extracting implicit targets with and without the use of CoT. The target is highlighted as: Without CoT, With CoT. Best viewed in color.

3.2 LLM-driven Hierarchical Collaborative Target Augmentation

To mitigate the under-utilization of implicit target information in the dataset, we propose a Hierarchical Collaborative Target Augmentation framework, which leverages LLMs' superior semantic reasoning capabilities to identify and exploit implicit target information. The framework comprises two main stages: Implicit Target Extraction and Stance Prediction via LLM Collaboration.

Implicit Target Extraction (ITE): We employ CoT prompting (Wei et al., 2022) to guide an LLM in discovering semantically but annotation-omitted targets. As demonstrated in Table 1, in a text discussing "corporations", the LLM equipped with CoT reasoning significantly outperforms its non-CoT counterpart, successfully identifying implicit targets like "democracy" that lack explicit textual cues but maintain crucial semantic relevance. This process generates a augmented dataset $D_{\text{ITE}} = \{(x_i, t_j)\}$, where each original text x_i is paired with one or more implicit targets $\{t_j\}$. While these text-target pairs lack initial stance labels, they capture valuable latent semantic relationships that prove essential for following module.

Stance Prediction via LLM Collaboration: For each text-target pair $(x_i, t_j) \in D_{\text{ITE}}$, we employ multiple LLMs to predict its stances. However, prior studies indicate that individual LLMs may introduce the bias (Gonçalves and Strubell, 2023). To mitigate the such biases, we adopt a multi-LLM voting strategy: L distinct LLMs independently predict stances through customized prompts:

$$s_{i,j}^k = \text{LLM}^k(x_i, t_j, \text{prompt}) \tag{1}$$

where $s_{i,j}^k$ is the stance predicted by the k-th LLM and $k \in \{1, 2, \cdots, L\}$. The resulting predictions are aggregated using a majority voting to produce pseudo-labels $\tilde{s}_{i,j}$:

$$\tilde{s}_{i,j} = \text{argmax}_{s \in \mathcal{Y}} \sum_{k=1}^{L} \mathbb{I}(s_{i,j}^k = s) \tag{2}$$

where $\mathcal{Y} \in \{\text{Favor}, \text{Against}, \text{Neutral}\}$ denotes the set of predefined stance labels and $\mathbb{I}(\cdot)$ is the indicator function. To ensure label quality, text-target pairs without clear consensus (fewer than $L/2$ votes) on the predicted label are discarded. The remaining instances are used to construct the final augmented dataset $D_{\text{aug}}^{final} = \{(x_i, t_j, \tilde{s_{i,j}})\}$, which incorporates previously underutilized implicit targets while maintaining annotation reliability. This enriched augmented dataset significantly improves the model's semantic inference, thereby improving the model's generalization to complex semantic scenarios.

3.3 Dynamic Multi-level Context-aware Attention Network (DyMCA)

Existing approaches for ZSSD and FSSD predominantly adopt fixed-weight text-target interaction processes, which fail to adapt to contextual dynamics (Zhang et al., 2023; Li et al., 2023). To address this limitation, We propose the DyMCA that enhances contextual adaptability by progressively transitioning from global understanding to local adjustment. As illustrated in Figure 2, DyMCA's architecture comprises three key modules: Joint Text-Target Representation, Content-aware Mechanism and Stance Predictor.

Joint Text-Target Representation (JTTR): To establish a holistic semantic basis, we first concatenate the text sequence X and target sequence T along the token dimension, and then passed through the BART (Lewis et al., 2019) encoder to obtain a joint contextual token representations:

$$[\text{h}_{\text{text}}; \text{h}_{\text{target}}] = \text{Encoder}([X; T]) \tag{3}$$

where $[\cdot;\cdot]$ denotes concatenation along the token sequence dimension. This yields text and target representations $\mathrm{h}_{\text{text}} \in \mathbb{R}^{L_{\text{text}} \times d}$ and $\mathrm{h}_{\text{target}} \in \mathbb{R}^{L_{\text{target}} \times d}$ both enriched with mutual contextual priors respectively. L_{text} and L_{target} being their respective sequence lengths, and d represents the hidden dimension size. For implementation, sequences are padded to a maximum length L_{max} for efficient computation.

Content-aware Mechanism (CM): Stance prediction requires not only an comprehensive understanding but also the capacity to regulate how the text-target respectively contribute to the stance decision. To address this need, we propose **CM**, which dynamically adjust the contributions of the text and target according to their contextual relevance.

Specifically, we first concatenate the text and target representations to compute a relevance score that reflects indicates the extent to which the stance can be directly inferred from the text:

$$\mathrm{h}_{\text{concat}} = \mathrm{h}_{\text{text}} \oplus \mathrm{h}_{\text{target}} \tag{4}$$

where $h_{\text{concat}} \in \mathbb{R}^{2d}$. This joint representation is then fed into a single-layer perceptron with a sigmoid activation:

$$w = \text{sigmoid}\left(W\mathrm{h}_{\text{concat}} + b\right) \tag{5}$$

where $W \in \mathbb{R}^{1\times 2d}$, $b \in \mathbb{R}$. The resulting scalar $w \in [0, 1]$ represents the relative contribution of the text. The final fused representation is then obtained by adaptively weighting text and target:

$$\mathrm{F} = w \cdot \mathrm{h}_{\text{text}} + (1 - w) \cdot \mathrm{h}_{\text{target}} + \mathrm{h}_{\text{target}} \tag{6}$$

This adaptive fusion enables the model to contextually prioritize either the text or target. Moreover, the residual target term serves as a stable semantic anchor that supports reliable inference even when stance expressions are ambiguous.

Stance Predictor: A classifier transforms the fused representation F into stance probability scores:

$$p = \text{Classifier}(\mathrm{F}) \tag{7}$$

Where $p \in \mathbb{R}^3$ represents the predicted probabilities over the three stance classes (Against, Favor, Neutral).

Training Phase: The mode optimization employs multi-class cross-entropy:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N} y^{(i)} \log(p^{(i)}) \tag{8}$$

	Train	Dev	Test
# Examples	13,477	2,062	3,006
# Unique Comments	1,845	682	786
# Zero-shot Topics	4,003	383	600
# Few-shot Topics	638	114	159

Table 2: Statistics of VAST Dataset.

where $p^{(i)}$ represents the predicted probability for the i-th stance category, $y^{(i)}$ indicates the corresponding ground-truth label, and N specifies the predefined number of stance categories. the model is trained on a mixture of original annotations D_{train} and augmented instances D_{aug}^{final}.

Inference phase: The augmented data D_{aug}^{final} is only used during training, while model evaluation is conducted solely on the original annotated datasets D_{dev} and D_{test} to assess stance prediction performance.

4 Experiments

4.1 Dataset and Evaluation Metrics

We evaluate our model on the VAried Stance Topics (**VAST**) dataset, a large-scale benchmark with diverse targets for ZSSD and FSSD (Allaway and McKeown, 2020). The statistics of VAST are demonstrated in Table 2. Following the previous work (Allaway and McKeown, 2020), the macro average of F1-score is adopted as evaluation metric.

4.2 Experimental Settings

We conduct all experiments on a single NVIDIA RTX A6000 GPU. For implicit target extraction, we employ GPT-4o[2]. For stance prediction, we leverage multiple LLMs, including GPT-4o, DeepSeek-V3[3] and Qwen-Max[4] to independently infer stance labels. We adopt the encoder of the BART[5] model pretrained on the MNLI dataset (Williams et al., 2018) as the Encoder. The learning rates of the Encoder and fully-connected layers are set to 2e-5 and 1e-3. AdamW (Loshchilov and Hutter, 2017) is utilized as the optimizer with a weight decay of 1e-4. The batch-size is 64, the maximum sequence length is 200 and hidden dimension of 1024.

We compare our model with several state-of-the-art baselines into two groups:

[2]https://chatgpt.com/
[3]https://platform.deepseek.com/
[4]https://bailian.console.aliyun.com/
[5]https://huggingface.co/facebook/bart-large-mnli

Model	Zero-Shot				Few-Shot				ALL			
	Con	Pro	Neu	ALL	Con	Pro	Neu	ALL	Con	Pro	Neu	ALL
BiCond ♮	0.475	0.459	0.349	0.427	0.463	0.454	0.259	0.392	0.468	0.457	0.306	0.410
Cross-Net ♮	0.434	0.462	0.404	0.434	0.505	0.508	0.410	0.474	0.471	0.486	0.408	0.455
BERT-joint ♮	0.584	0.546	0.853	0.660	0.597	0.543	0.796	0.646	0.591	0.545	0.823	0.653
TGA-Net ♮	0.585	0.568	0.858	0.666	0.595	0.589	0.805	0.663	0.590	0.573	0.831	0.665
PT-HCL ♭	0.617	0.63.5	0.89.6	0.716	0.623	0.670	0.843	0.712	–	–	–	–
JointCL ♭	0.649	0.632	0.889	0.723	0.632	0.667	0.846	0.715	–	–	–	–
DyMCA w/o HCTA	**0.735**	**0.726**	**0.916**	**0.792**	**0.708**	**0.676**	**0.871**	**0.751**	**0.715**	**0.666**	**0.893**	**0.758**
SEKT ♯	0.442	0.504	0.308	0.418	0.479	0.510	0.215	0.474	0.462	0.507	0.263	0.411
CKE-Net ♯	0.612	0.612	0.880	0.702	0.622	0.644	0.835	0.701	0.617	0.629	0.857	0.701
BS-RGCN ♯	0.674	0.608	0.895	0.726	0.665	0.600	0.839	0.702	0.669	0.604	0.866	0.713
CR-MTR ♭	0.652	0.601	0.886	0.713	0.638	0.636	0.860	0.712	0.645	0.620	0.873	0.713
TTS ♭	0.751	0.725	**0.925**	0.801	0.684†	0.669†	0.878†	0.744†	0.696†	0.684†	0.893†	0.758†
LKI-BART ♭	0.751	0.729	0.907	0.796	–	–	–	–	–	–	–	–
TTS with HCTA	0.755	0.734	0.921	0.803	0.692	0.691	0.883	0.755	0.712	0.694	0.893	0.766
DyMCA with HCTA	**0.758**	**0.746** *	**0.925**	**0.810** *	**0.702** *	**0.719** *	**0.891**	**0.771** *	**0.723** *	**0.705** *	**0.899**	**0.776** *

Table 3: Performance comparison of models on stance detection across zero-shot, few-shot and all settings. ♮ denotes results are taken from (Liu et al., 2021), ♯ are taken from (Zhang et al., 2023), † indicates results from our reproduction, ♭ are taken from the original papers and * improves the best baseline at $p < 0.05$ with paired t-test. Results are highlighted with blue (↑) for improvements, purple (→) for stable and red (↓) for degradations relative to corresponding baselines. Best viewed in color.

Model	Zero-Shot 10% Train			
	Con	Pro	Neu	ALL
BiCond ♮	0.401	0.298	0.346	0.348
Cross-Net ♮	0.329	0.373	0.385	0.362
TGA-Net ♮	0.582	0.476	0.864	0.641
BERT-joint ♮	0.552	0.496	0.888	0.645
TTS ♮	0.721	0.719	**0.913**	0.784
DyMCA	**0.741**	**0.732**	0.908	**0.794**
TTS with HCTA	0.735	0.727	0.916	0.793
DyMCA with HCTA	**0.747**	**0.738**	**0.919**	**0.801**

Table 4: Zero-Shot 10% Train. ♮ denotes numbers are taken from (Li et al., 2023).

(1) Models without External Information: **BiCond** (Augenstein et al., 2016) and **CrossNet** (Xu et al., 2018) predict the class label based on the conditional encoding of the BiLSTM model. **BERT-joint** (Allaway and McKeown, 2020)and **TGA-Net** (Allaway and McKeown, 2020) encode the texts and targets using the BERT model, followed by classification with two fully-connected layers. **PT-HCL** (Liang et al., 2022a) and **JointCL** (Liang et al., 2022b) adopt a contrastive learning strategy to mine the relationships and differences within stance features.

(2) Models with External Information: **SEKT** (Zhang et al., 2020), **CKE-Net** (Liu et al., 2021) and **BS-RGCN** (Luo et al., 2022) apply the GCN to incorporate external knowledge into the stance detection process. **CR-MTR** (Zhang et al., 2023) and **TTS** (Li et al., 2023) both devise custom-made augmentation strategies from diverse dimensions. **LKI-BART** (Zhang et al., 2024) integrates contextual knowledge generated by LLM.

4.3 Results

4.3.1 Results of Different Scenarios

We conduct comprehensive evaluations of our proposed DyMCA across three distinct scenarios: Zero-Shot, Few-Shot and All. The overall results of our model and main baselines are summarized in Table 3, our model achieves state-of-the-art performance with macro F1-scores of **0.81**, **0.771** and **0.776** respectively, representing improvements of **1.12%**, **3.63%** and **2.37%** over the strongest competitor (TTS). Our analysis reveals three key findings:

First, compared to models without leveraging external knowledge or augmented strategy (shown in the first block of Table 3) DyMCA without HCTA strategy achieves a remarkable **9.5%** improvement in zero-shot performance over the best competitor in this category. This highlights the critical role of dynamic contribution regulation and demonstrates DyMCA's ability to adaptively adjust its focus under different contexts, thereby enhancing overall performance.

Second, when incorporating HCTA strategy (re-

Model	Zero-Shot	Few-Shot	ALL
CR-MTR	0.713	0.712	0.713
TTS	0.801	0.744	0.758
LKI-BART	0.796	–	–
DyMCA with HCTA	**0.810**	**0.771**	**0.776**
HCTA			
w/o All	0.792	0.751	0.758
w/o CoT	0.796	0.753	0.762
w/o Vote	0.793	0.746	0.754
Model Architecture			
w/o JTTR	0.791	0.753	0.766
w/o CM	0.787	0.756	0.751
- residual target	0.790	0.757	0.757

Table 5: Ablation study of HCTA framework and DyMCA architecture. 'w/o All' removes the HCTA entirely; 'w/o CoT' excludes Chain-of-Thought in implicit target extraction; 'w/o Vote' uses a single model for stance prediction; 'w/o JTTR' removes Joint Text-Target Representation, where text and target are encoded independently; 'w/o CM' omits content-aware mechanism including residual target.

Model	Imp	mlT	mlS	Qte	Sarc
BERT-joint	57.1	59.0	52.4	63.4	60.1
TGA-Net	59.4	60.5	53.2	66.1	63.7
CKE-Net	62.5	63.4	55.3	69.5	68.2
BS-RGCN	62.1	64.7	55.6	70.1	71.7
CR-MTR	61.0	65.9	55.8	67.1	70.8
TTS †	63.1	66.5	57.4	71.3	71.9
DyMCA	**65.3**	**67.9**	**59.8**	**72.5**	**73.1**
TTS with HCTA	66.4	68.5	60.4	73.3	73.2
DyMCA with HCTA	**67.6**	**69.8**	**62.5**	**75.1**	**75.4**

Table 6: Accuracy (%) on five phenomena in the VAST.

sults in the second block of Table 3), DyMCA additional performance gains, achieves new state-of-the-art results across all scenarios. In particular, it obtains an overall macro-score of **0.81** in the Zero-Shot setting, outperforming all prior augmented baselines with improvements of **1.12%**, demonstrating the complementary benefits of our model design and augmentation strategy.

Third, further experiments shows that applying HCTA to the TTS baseline yields improvements of **1.48%** (Few-Shot) and **1.06%** (ALL), confirming the general effectiveness of our augmentation method. However, the performance gap between enhanced TTS and DyMCA highlights the fundamental advantages of our model's architectural innovations in stance detection tasks.

4.3.2 Low-resource scenario Evaluation

We conduct experiments under low-resource conditions using only 10% of the training data (Table 4) to evaluate the effectiveness of the proposed DyMCA framework and HCTA strategy. Results demonstrate that DyMCA (without HCTA) achieves significant performance improvements over the state-of-the-art TTS baseline, particularly for Con and Pro labels with gains of **2.77%** and **1.81%** respectively. A slight decrease on Neu label might arise because neutral expressions often lack explicit stance signals, this adaptive weighting may introduce noise that interferes with the accurate recognition of neutrality. Furthermore, the HCTA strategy proves particularly beneficial in this resource-constrained setting: while TTS with HCTA shows a **1.15%** improvement, DyMCA with HCTA achieves a new state-of-the-art macro-F1 score of **0.801**. This results confirm that the implicit target information captured by HCTA provides valuable supplementary signals that effectively enhance model generalization, complementing DyMCA's architectural strengths. Overall, our framework demonstrates strong robustness and effectiveness even with severely limited training data.

4.3.3 Ablation Study

We conduct ablation studies to assess the impact of key components in DyMCA and HCTA, as shown in Table 5.

Within the HCTA pipeline, we complete removal of the augmentation strategy (w/o All) leads to significant performance degradation, demonstrating the importance of LLM-driven implicit target information mining. While eliminating the Chain-of-Thought prompting (w/o CoT) causes modest performance drops, confirming its value in eliciting comprehensive reasoning from LLMs, and removing the multi-LLM voting mechanism (w/o Vote) also reduces performance, indicating that single LLM predictions may introduce stance bias and result in reduced predictive reliability.

On the architectural side, DyMCA exhibits performance declines when critical components are ablated. Removing the joint text-target representation (w/o JTTR) degrades performance, validating the benefit of unified global context modeling. Additionally, eliminating the Content-aware Mechanism (w/o CM) or its residual target design causes performance drops, supporting our hypothesis about the relative contributions of text and target must vary with context. The residual target addi-

Text	Target	TTS	DyMCA	DyMCA with HCTA
If we keep using fossil fuels like we're doing now, future generations are definitely going to enjoy our smog-filled skies. What a legacy to leave behind!	renewable energy	AGAINST	FAVOR	FAVOR
Online learning has changed the way students access education, especially during the past few years.	online education	FAVOR	FAVOR	NEUTRAL

Table 7: Case study comparing stance prediction results across different models. We evaluate the strongest competitor TTS, our DyMCA model, and DyMCA enhanced with the HCTA strategy. Predictions are highlighted as CORRECT and INCORRECT.

tionally proves indispensable by providing a stable semantic reference across diverse scenarios. Overall, these findings confirm the complementary roles of global and local context modeling in DyMCA' architecture.

4.3.4 Results of Different Phenomena

We conduct a experiment evaluation of DyMCA and HCTA across five challenging phenomena: (1) ***Imp***, the target is not contained in the text and the label is non-neutral, (2) ***mlT***, multiple samples share the same text but differ in target, (3) ***mlS***, the same text appears varying non-neutral stance labels, (4) ***Qte***, the text includes quoted content, and (5) ***Sarc***, the text contains sarcasm. As shown in Table 6, DyMCA both with and without HCTA augmentation, achieves the best performance across all challenging phenomena. Notably, when equipped with HCTA, the baseline TTS model also shows significant performance improvements in all complex scenarios. These results validate the robustness of both DyMCA's adaptive architecture and HCTA's augmentation strategy in handling diverse and challenging target-stance phenomena.

4.4 Case Study

We present comparative case analyzing our model with TTS trained under the zero-shot setting (as shown in Table 7) to highlight the effectiveness of DyMCA and HCTA. The first example reflects implicit sentiment toward the target "renewable energy", expressed through sarcasm about fossil fuels. While TTS incorrectly predicts an Against stance, DyMCA correctly identifies a Favor stance, showing that its adaptive mechanism can recalibrate text–target contributions based on context and capture implicit stance signals more effectively. The improvement is further solidified by the HCTA, showing the benefit of augmented contextual grounding. In the second case, the target "online education" presented through factual statements without explicit stance indicators. While both TTS and DyMCA erroneously predict Favor, the DyMCA with HCTA correctly outputs Neutral. These cases collectively demonstrate DyMCA's advancements in processing both explicit and implicit stance indicators through its integrated architecture and augmentation strategy.

5 Conclusion

In this work, we address two key challenges in ZSSD and FSSD: under-utilization of implicit targets and insufficient contextual adaptation capability. To this end, we propose **HCTA**, a hierarchical framework that leverages LLMs to extract and incorporate implicit target information, and **DyMCA**, a dynamic multi-level context-aware network that adaptively balances text–target contributions under different contexts. Extensive experiments on benchmark dataset demonstrate that our approach achieves state-of-the-art performance on both ZSSD and FSSD tasks, underscoring the effectiveness of integrating LLM-driven target augmentation and context-sensitive interaction modeling to enhance stance detection beyond explicit target supervision and rigid interaction schemes.

Limitations

Despite the demonstrated effectiveness, our approach has several limitations. First, the reliability of extracted implicit targets and corresponding pseudo-labels is inherently tied to the reasoning quality and consistency of LLMs, which may vary across model families and domain shifts. Second, the computational cost of multiple LLMs collaboration could pose challenges for large-scale applications. Future work may explore more efficient yet reliable target augmentation strategies, as well as principled methods for uncertainty-aware filtering

of noisy LLM outputs to ensure robust augmentation strategy.

References

Abeer ALDayel and Walid Magdy. 2021. Stance detection on social media: State of the art and trends. *Information Processing & Management*, 58(4):102597.

Emily Allaway and Kathleen McKeown. 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)*, pages 8913–8931, Online. Association for Computational Linguistics.

Isabelle Augenstein, Tim Rocktäschel, Andreas Vlachos, and Kalina Bontcheva. 2016. Stance detection with bidirectional conditional encoding. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 876–885, Austin, Texas. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Gustavo Gonçalves and Emma Strubell. 2023. Understanding the effect of model compression on social bias in large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 2663–2675, Singapore. Association for Computational Linguistics.

Eduardo Graells-Garrido, Ricardo Baeza-Yates, and Mounia Lalmas. 2020. Representativeness of abortion legislation debate on twitter: a case study in argentina and chile. In *Companion Proceedings of the Web Conference 2020*, pages 765–774.

Kazi Saidul Hasan and Vincent Ng. 2014. Why are you taking this stance? identifying and classifying reasons in ideological debates. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 751–762, Doha, Qatar. Association for Computational Linguistics.

Dilek Küçük and Fazli Can. 2020. Stance detection: A survey. *ACM Computing Surveys (CSUR)*, 53(1):1–37.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.

Yingjie Li, Chenye Zhao, and Cornelia Caragea. 2023. Tts: A target-based teacher-student framework for zero-shot stance detection. In *Proceedings of the ACM Web Conference 2023*, WWW '23, page 1500–1509, New York, NY, USA. Association for Computing Machinery.

Bin Liang, Zixiao Chen, Lin Gui, Yulan He, Min Yang, and Ruifeng Xu. 2022a. Zero-shot stance detection via contrastive learning. In *Proceedings of the ACM Web Conference 2022*, WWW '22, page 2738–2747, New York, NY, USA. Association for Computing Machinery.

Bin Liang, Qinlin Zhu, Xiang Li, Min Yang, Lin Gui, Yulan He, and Ruifeng Xu. 2022b. Jointcl : a joint contrastive learning framework for zero-shot stance detection. In *The 60th Annual Meeting of the Association for Computational Linguistics (ACL)*, volume 1, pages 81–91. Association for Computational Linguistics.

Rui Liu, Zheng Lin, Yutong Tan, and Weiping Wang. 2021. Enhancing zero-shot and few-shot stance detection with commonsense knowledge graph. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3152–3157, Online. Association for Computational Linguistics.

Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Yun Luo, Zihan Liu, Yuefeng Shi, Stan Z. Li, and Yue Zhang. 2022. Exploiting sentiment and common sense for zero-shot stance detection. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 7112–7123, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. SemEval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41, San Diego, California. Association for Computational Linguistics.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Penghui Wei and Wenji Mao. 2019. Modeling transferable topics for cross-target stance detection. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR'19, page 1173–1176, New York, NY, USA. Association for Computing Machinery.

Adina Williams, Nikita Nangia, and Samuel R. Bowman. 2018. A broad-coverage challenge corpus for sentence understanding through inference. *Preprint*, arXiv:1704.05426.

Chang Xu, Cécile Paris, Surya Nepal, and Ross Sparks. 2018. Cross-target stance classification with self-attention networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 778–783, Melbourne, Australia. Association for Computational Linguistics.

Bowen Zhang, Min Yang, Xutao Li, Yunming Ye, Xiaofei Xu, and Kuai Dai. 2020. Enhancing cross-target stance detection with transferable semantic-emotion knowledge. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3188–3197, Online. Association for Computational Linguistics.

Jiarui Zhang, Shaojuan Wu, Xiaowang Zhang, and Zhiyong Feng. 2023. Task-specific data augmentation for zero-shot and few-shot stance detection. In *Companion Proceedings of the ACM Web Conference 2023*, WWW '23 Companion, page 160–163, New York, NY, USA. Association for Computing Machinery.

Zhao Zhang, Yiming Li, Jin Zhang, and Hui Xu. 2024. LLM-driven knowledge injection advances zero-shot and cross-target stance detection. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*, pages 371–378, Mexico City, Mexico. Association for Computational Linguistics.

Appendix

A CoT Prompt Details of Implicit Target Extraction

Task Description: Your task is to read a given segment of text and identify the target entity or entities that are implied but not explicitly mentioned. To do this effectively, please think through the problem step by step, using the following process steps as a guide.
Steps:

1. Understand Context: Briefly grasp the text's main topic and argument.
2. Identify Clues: Look for keywords or phrases suggesting related but unmentioned entities.
3. Infer Potential Targets: Deduce possible implicit targets based on these clues and the overall context.
4. Verify Rationality: Briefly justify why the inferred target is relevant to the text's meaning.
5. Formulate Target: Clearly state the identified implicit target(s).

Text: < The given text >
Annotation Targets: $[T_1, T_2, ..., T_n]$
Output Format: [A comma-separated list of the inferred implicit targets]

B LLM-Prediction Prompt Details

Task: As an expert in sentiment analysis within Natural Language Processing, your task is to understand and analyze the stance that the given text expresses towards the corresponding target.
Text: < The given text >
Target: < The specified target >
Output Format: [FAVOR | AGAINST | NEUTRAL]

C Impact of Implicit Target Augmented Data Quantity

First, we present comprehensive statistics of the implicit target augmented data. The augmented dataset contains 6,988 instances with a well-balanced distribution across three stance categories: FAVOR (29%), AGAINST (38%), and NEUTRAL (33%).

We further investigate how varying quantities of implicit target augmented data affect model performance in zero-shot stance detection. The line chart in Figure 3 reveals a clear positive correlation between the amount of augmented data used (from 0% to 100% of the augmented set) and the model's macro average F1 score when combined with the original training data. This consistent performance improvement demonstrates that our implicit target augmentation strategy effectively uncovers latent beneficial patterns in the dataset, leading to enhanced model performance and generalization capability. The results suggest that incorporating more augmented data helps the model better capture the nuanced relationships between implicit targets and their corresponding stances.

D Comparative Analysis of Language Model Scales for Target Augmentation and Stance Prediction

Prior work (Li et al., 2023) has predominantly employed small pre-trained language models for both

Target-Extract	Stance-Predict	Zero-Shot	Few-Shot	ALL
Small-LM	Small-LM	0.773	0.749	0.755
Small-LM	Large-LM	0.786	0.758	0.758
Large-LM	Small-LM	0.785	0.747	0.753
Large-LM	Large-LM	0.810	0.771	0.776

Table 8: Performance comparison of different combinations of target extraction and stance prediction modules using Small Language Models and Large Language Models under zero-shot, few-shot, and combined (ALL) settings.

Method Combination		Annotation Targets	Implicit Targets	Stance Prediction
SLM (Targets)	SLM (Stance)	company	corporations	AGAINST
		corporate regulation	regulation	AGAINST
		regulation corporation	democracy	AGAINST
SLM (Targets)	LLMs (Stance)	company	corporations	AGAINST
		corporate regulation	regulation	AGAINST
		regulation corporation	democracy	FAVOR
LLM (Targets)	SLM (Stance)	company	democracy	AGAINST
		corporate regulation	campaign	AGAINST
		regulation corporation	legal system	AGAINST
LLM (Targets)	LLMs (Stance)	company	democracy	FAVOR
		corporate regulation	campaign	AGAINST
		regulation corporation	legal system	AGAINST

Table 9: Exploration of Implicit Target Identification and Stance Prediction Using Different Model Combinations. SLM denotes the Small Language Model, and LLM(s) denotes Large Language Models. This analysis evaluates the effectiveness of combining these methods to uncover implicit targets and predict stances, with a focus on leveraging LLMs for both tasks due to their advanced contextual understanding and predictive capabilities. This text corresponds to the text of first example in Table 11.

target mining and stance prediction to achieve data augmentation. However, we argue that compared to Large Language Models (LLMs), these smaller models exhibit weaker semantic reasoning capabilities, making them inadequate for comprehensively capturing complex and implicit target patterns. To validate this claim, we present the following experimental analysis and case study.

Based on the results shown in Table 8, we conducted an analysis to assess the impact of different model combinations on implicit target identification and stance prediction. The results clearly demonstrate that using Large Language Models (LLMs) for either target extraction or stance prediction leads to noticeable improvements. Furthermore, combining LLMs for both tasks results in significant performance gains. This highlights the complementary strengths of LLMs in capturing implicit semantics and performing stance inference.

We provide a specific example for each model combination to further illustrate the conclusions mentioned above, as shown in Table 9. When using small pre-trained language model for both tasks (**SLM+SLM**), the model generates limited and partially redundant targets, resulting in consistent but incorrect stance predictions. Substituting the stance module with an LLM (**SLM+LLMs**) yields more accurate stance labeling for implicit target like "democracy", despite target limitations. In contrast, replacing the target identification module with an LLM (**LLM+SLM**) produces more diverse and semantically rich targets, though stance predictions remain constrained by the small language model's limited inference capacity. The full LLM combination (**LLM+LLMs**) achieves the most coherent target-statement alignment and accurate stance recognition, highlighting the advantage of leveraging LLMs for both contextual comprehension and predictive reasoning in implicit stance scenarios.

Text	Implicit Target	ChatGPT	DeepSeek	Qwen
Regulation of corporations has been subverted by corporations. States that incorporate corporations are not equipped to regulate corporations that are rich enough to influence elections, are rich enough to muster a legal team that can bankrupt the state. Money from corporations and their principals cannot be permitted in the political process if democracy is to survive.	democracy	FAVOR	AGAINST	FAVOR
"...one must ask how much money they must make to demonstrate that they are among the best managed companies on the planet." They must make enough money to insure that they can never fail and threaten the stability of the worlds economy again. That much money.	management	AGAINST	AGAINST	FAVOR

Table 10: Illustrative stance predictions for implicit targets from various Large Language Models. Correct predictions are highlighted: CORRECT , INCORRECT . Optimal viewing in color.

E Motivation for Multi-LLM Voting in Stance Prediction

We showcase in Table 10 the stance predictions produced by multiple LLMs (ChatGPT, DeepSeek, and Qwen) for texts toward implicit targets. The results from different LLMs are not always consistent, with some models yielding incorrect stance predictions due to varied interpretations or reasoning paths. This variation highlights a core challenge in leveraging LLMs for complex semantic tasks: despite their strong individual capabilities, no single model can guarantee reliable performance across all instances. To address this, we adopt a simple yet effective voting strategy within our HCTA framework to aggregate predictions from multiple LLMs. By relying on collective agreement that can mitigate individual model biases or occasional errors, yielding more robust and accurate final predictions. This ensemble approach leverages the complementary strengths of different models and introduces a form of redundancy that is particularly beneficial when dealing with ambiguous or under-annotated data. The examples in the table clearly demonstrate cases where voting corrects individual misjudgments, further justifying the necessity of collaborative inference in our framework.

F Case Study: Small-LM vs. Large-LM Collaboration in Stance Prediction

We presents representative examples (Table 11) to illustrate the effectiveness of our LLMs-collaboration strategy in predicting stances toward implicit targets. Specifically, we compare the predictions made by a standard small pre-trained language model (BART-large-mnli) and our collaborative LLM-based approach. The implicit targets shown in the examples are not explicitly mentioned in the input texts but are instead inferred through deeper semantic understanding. As demonstrated, the pre-trained model tends to misclassify the stance due to its limited capacity for indirect reasoning. In contrast, the LLMs-collaboration approach correctly identifies the intended stance by leveraging richer contextual inference. This underscores the advantage of our HCTA framework, which equips models with more fine-grained comprehension, particularly in cases involving abstract, metaphorical, or ideologically charged content. These examples further validate the necessity of incorporating implicit supervision signals and highlight the limitations of relying solely on off-the-shelf pre-trained encoders for challenging stance detection scenarios.

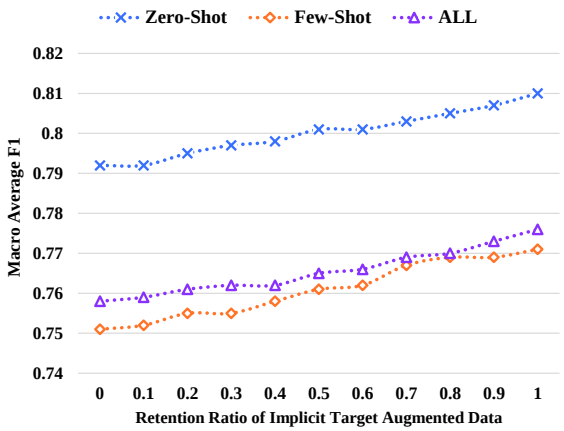

Figure 3: Variation in performance when combing different proportions of augmented data with original training dataset.

Text	Implicit Target	Small-LM	LLM-Collaboration
Regulation of corporations has been subverted by corporations. States that incorporate corporations are not equipped to regulate corporations that are rich enough to influence elections, are rich enough to muster a legal team that can bankrupt the state. Money from corporations and their principals cannot be permitted in the political process if democracy is to survive.	democracy	AGAINST	FAVOR
"...one must ask how much money they must make to demonstrate that they are among the best managed companies on the planet." They must make enough money to insure that they can never fail and threaten the stability of the worlds economy again. That much money.	management	FAVOR	AGAINST

Table 11: Comparative examples of stance prediction, contrasting the output of a Small Language Model (BART-large-mnli) with the LLM-collaboration approach. Predictions are color-coded: CORRECT , INCORRECT . Best viewed in color.