GRASP-ChoQ: Knowledge Graph-Based Retrieval Augmentation for Stance Detection in Political Texts with Chain-of-Questions Reasoning

Md Rasel Mahmud
USTC, China
raselmahmud@mail.ustc.edu.cn

Md. Abdur Rakib Mollah
Independent Researcher
rakib1703115@gmail.com

Aninda Kumar Sharma
University of Adelaide
aiubanik18@gmail.com

Omar Faruq Osama
SUNY Binghamton
oosama@binghamton.edu

Abstract

Political stance detection in understudied socio-political contexts presents a persistent challenge for language models. It is because dynamic contexts and indirect relationships between political entities complicate the accurate alignment of opinions. To address this, we introduce **GRASP-ChoQ**, an approach that combines structured knowledge graphs with a chain-of-questions reasoning to break down interactions in political texts. We support this with **BPDisC**, a novel dataset of politically charged tweets from Bangladesh during and after the July 2024 protests. Instead of making direct predictions, our approach relies on intermediate reasoning steps facilitated by contextually retrieved subgraphs from the knowledge graph. We evaluated our method on BPDisC and six additional benchmark datasets. Across all datasets, it consistently outperformed baseline LLMs. Notably, DeepSeek R1 combined with GRASP-ChoQ achieved the highest performance on the BPDisC dataset, with a **40%** improvement in F1 score over the zero-shot baseline. These results highlight our method's ability to integrate context and effectively handle complex, low-resource, and evolving political scenarios.[1].

1 Introduction

Since large language models (LLMs) are usually trained on static datasets containing predetermined temporal cutoffs (Liu et al., 2024), they are not aware of events or novel ideas that emerge following the training period. As an example, a pre-trained model might have difficulty identifying a change in political ideology (Liu et al., 2022). Retraining these models by incorporating new data is computationally challenging and often results in updated models losing their previous knowledge. Since LLMs cannot naturally adapt to situations that change quickly, it is hard to figure out political positions in environments that are constantly changing politically and socially (Rozado, 2024).

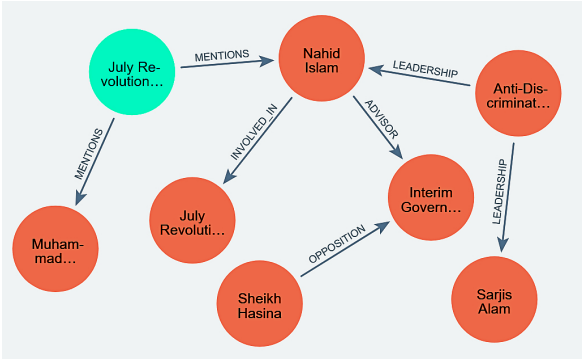

Figure 1: A knowledge graph containing political entities as nodes and their relationships as edges

Identifying political stances often requires understanding indirect relationships—such as informal alliances or ideological disagreements—that are implied rather than explicitly stated in texts (Leifeld and Brandenberger, 2019; Oswald et al., 2021). LLMs struggle to detect these subtle cues, especially in tweets, due to limited context and a focus on surface patterns (Cheng et al., 2024). While Retrieval-Augmented Generation (RAG) aids in providing contexts, it does not always effectively uncover these nuanced political connections.

This study presents **GRASP-ChoQ**, a hybrid approach which improves LLM-based stance detection by combining knowledge graphs (KGs) and multistep reasoning. This approach addresses the dual challenges of adapting to dynamic political contexts and inferring indirect relational

[1] https://github.com/Programming-Dude/GRASP-ChoQ

Proceedings of the Second Workshop on Bangla Language Processing (BLP-2025), pages 8–30

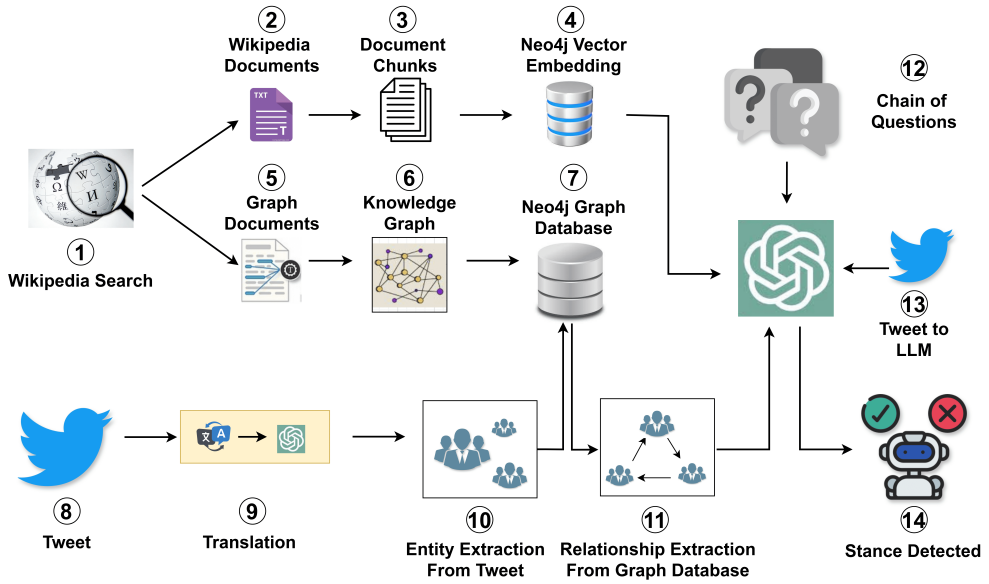

Figure 2: Our approach, GRASP-ChoQ, detects stance using two main components: Knowledge Graph and Chain-of-Questions Reasoning. The pipeline integrates Wikipedia search, a Neo4j graph database, and language models. Wikipedia documents are processed into chunks and stored in the knowledge graph. The graph database extracts entities and relationships, and a language model generates queries to determine the tweet's stance.

knowledge.

A significant aspect of the structure it contains is the incorporation of relational principles compared to **Social Balance Theory (SBT)** (Heider, 2013), which are frequently utilized to analyze harmony or stress in undirected triadic interactions (e.g., "a friend of my friend is my friend" or "a foe of my foe is my friend").

In order to describe asymmetric interactions using signed, directional edges, recent extensions of SBT to directed graphs have been devised (Aref et al., 2020). According to the principles of SBT, knowledge graph integration may be very helpful in detecting latent linkages by organizing the data in a manner that makes it possible to uncover hidden connections based on interactions that already exist (Chepurova et al., 2023). That is why our approach integrates this.

Approaching stance detection as a one-step decision limits the ability to reason through complex, indirect relationships. We address this by providing a set of questions (Zhu et al., 2023) that divides the process into smaller, more controllable phases, therefore facilitating the correct decision.

While prior works have separately explored knowledge-graphs, retrieval augmentation (e.g., RAG with static subgraph retrieval) and Chain-of-Thought (CoT) or Chain-of-Questions prompting for enhanced reasoning, no existing approach integrates multi-hop KG traversal with question-guided reasoning for political stance detection.

GRASP-ChoQ is particularly effective for complex political environments such as Bangladesh (Chowdury, 2024), a geopolitically significant region in recent times that remains underexplored in the context of target-based political stance detection (Bestvater and Monroe, 2023).

The main contributions of our research paper are as follows:

- We introduce Knowledge **G**raph-Based **R**etrieval **A**ugmentation for **S**tance Detection in **P**olitical Texts with **Ch**ain-**o**f-**Q**uestions Reasoning, shortly called **GRASP-ChoQ**. This approach combines two essential components: Knowledge Graph (KG) and Chain-of-Questions (ChoQ) to detect political stances effectively.
- We present **BPDisC** (**B**angladesh **P**olitical **Dis**course **C**orpus)-a novel dataset comprising tweets that explicitly express political stances related to the political climate in Bangladesh during and following the July 2024 protests.
- Our proposed method substantially improves the **reasoning capabilities** of baseline LLMs

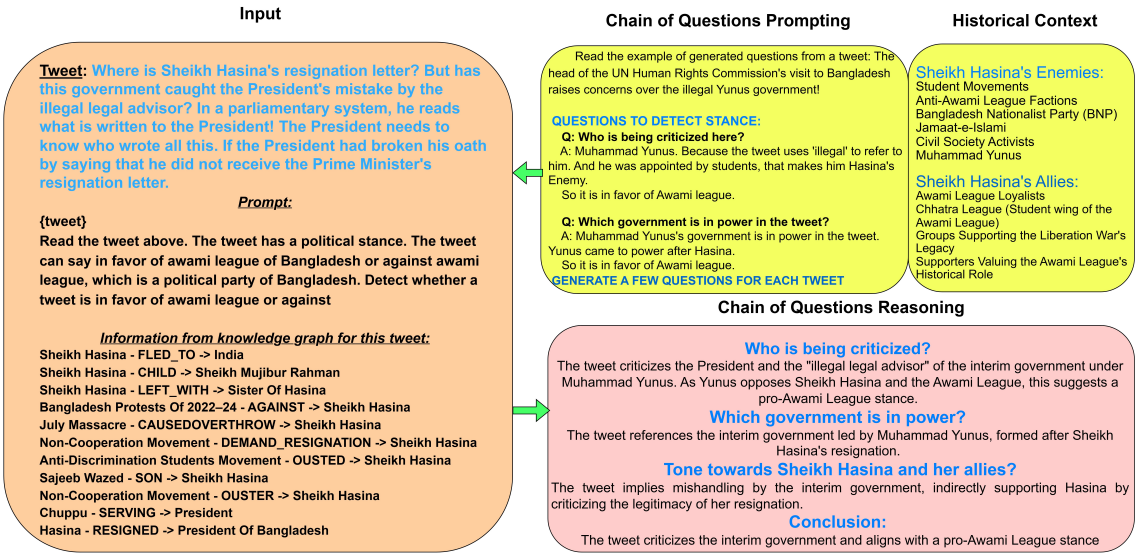

Figure 3: This figure illustrates a structured approach which integrates information from knowledge graph, chain-of-questions prompting, historical context, and reasoning to classify tweets

in detecting political stances by addressing the dual challenges of adapting to dynamic political contexts and inferring indirect relational knowledge, as confirmed through an ablation study.

2 Related Works

Kopacheva et al. (2023) explored the structural roles of social networks in predicting the participation of protests in the context of ecological protests in Russia. The research emphasized how personal social networks predict political participation by highlighting network density and size as major attributes affecting the likelihood of participation in a protest action.

Panda et al. (2024) proposed HOLMES, a distilled knowledge graph approach for multi-hop QA, reducing token use and improving Exact Match on HotpotQA and MuSiQue, though challenged by graph incompleteness. Chen et al. (2024b) introduced a three-stage LLM-based pipeline for KGR without fine-tuning, enhancing KG enrichment and entity reranking, but limited by cost and scalability. Edge et al. (2024) presented Graph RAG for query-focused summarization via graph-based retrieval and indexing, requiring broader evaluation. Wu et al. (2024) introduced MedGraphRAG, a medical RAG model using triple-linked graphs and U-Retrievals, needing clinical validation and real-time updates. Garg and Caragea (2024) addressed target prioritization failures in stance detection with Stanceformer, a target-aware transformer that boosts self-attention scores, though constrained by LLM size and resources. Weinzierl and Harabagiu (2024) proposed "Tree-of-Counterfactual Prompting" (TR-ZSSD), a zero-shot stance detection method reliant on culturally informed models, limiting application in diverse settings. Chen et al. (2024a) enhanced FSA on complex financial texts using a four-step Chain-of-Thought LLM pipeline, showing top performance. Mei et al. (2024) proposed a contrastive learning approach for hateful meme detection, with an updatable embedding space that avoids retraining.

Shui et al. (2024) proposed a Llama3-8b-based emotion text classification model enhanced with LoRA and FlashAttention. Focused on stance detection in political tweets, it leverages knowledge graph principles for improved classification of complex opinions. However, dataset limitations restricted sentiment intensity regression to just two datasets, and hardware constraints prevented fine-tuning on larger models like Llama 3-70B, limiting results to 8B and 7B variants. Sreekala et al. (2024) introduced a news classification scheme using hierarchical clustering with ensemble methods. Documents were clustered with various aggregative techniques and classified using Gradient Boosting, Bagging, and Random Forests. A key limitation is its reliance on the BBC News dataset, which

Datasets	Target(s)	Source	Size	Time	Language	Annotation Classes
SemEval-2016 Task 6	Atheism, Climate Change is Concern, Feminist Movement, Hillary Clinton, Legalization of Abortion, Donald Trump	Twitter	4,870 tweets	2016	English	Favor, Against, Neither
P-stance	Donald Trump, Joe Biden, Bernie Sanders	Twitter	21,574 tweets	2020	English	Favor, Against, Neither
Mawqif	COVID-19 vaccine, digital transformation, women empowerment	Twitter	4,121 tweets	2022	Arabic	Favor, Against, None
Swami et al. (2018)	Demonetisation in India in 2016	Twitter	3,545 tweets	2018	English-Hindi	Favor, Against, None
Lai et al. (2018).	2016 referendum on reform of the Italian Constitution	Twitter	993 triplets (2,889 tweets)	2018	Italian	Favor, Against, None
Lüüsi et al. (2024)	Immigration-related sentences from Estonian news articles from Ekspress Grupp and Uued Uudised	News media	3,261 sentences	2015–2022	Estonian	Supportive, Neutral, Against
BPDisC (Ours)	**Awami League (Bangladeshi Political Party)**	**Twitter**	**2,847 tweets**	**2024**	**Bangla-English**	**Favor, Against**

Table 1: Stance detection datasets for the experiments. The datasets cover political topics across different languages, including English (Mohammad et al., 2016a; Li et al., 2021), Arabic (Alturayeif et al., 2022), Hindi, Italian, Estonian. For our experiments we only used FAVOR and AGAINST classes.

may not fully represent diverse news sources, affecting classification performance. Shimin (2024) explored book content classification using LLaMA's self-attention and word embedding mechanisms, demonstrating improved efficiency and accuracy in book classification.

3 Methodology

3.1 Retrieval Augmented Generation

RAG retrieves documents by measuring semantic similarity—typically via vector embeddings and metrics like cosine similarity—treating each document as an isolated point in embedding space (Barnett et al., 2024). This, however, ignores relational context such as citations or knowledge graph links.

For instance, in a citation network (Clough et al., 2015), papers P_1, P_2, P_3 may be connected via citation edges $P_1 \rightarrow P_2$ and $P_2 \rightarrow P_3$. If the logical citation path ($Q \rightarrow P1 \rightarrow P2 \rightarrow P3$) is more informative, a retrieval system depending only on embeddings would prioritize retrieving *P3* if it has a high embedding similarity to a given query *Q*.

3.2 Knowledge Graph

Graph-based structures link pieces of information into nodes and edges, making it easy to organize and find what is needed (Xie et al., 2024). Just the relevant subgraph is then pulled from the knowledge graph to show how entities connect (Li et al., 2024). This turns long passages into a small, clear graph (Dong et al., 2024).

From Table 2 and Table 3, we can see how a long text is summarized and the key relationships are extracted. Next, we performed subgraph retrieval from the larger knowledge graph by identifying key entities within the tweet with LLM. This process involved extracting relevant entities and formulating a graph query to retrieve the most contextually appropriate subgraph.

Let:

- q be the input tweet.
- $E(q)$ represent the set of entities extracted from q.
- $g(e)$ represent the function that generates a query from entity e.
- $G(q)$ represent the result of querying the graph with query q.

The function can be expressed as:

$$\text{retriever}(q) = \bigcup_{e_i \in E(q)} G(g(e_i))$$

3.3 Chain of Questions

The primary motivation for adopting a Chain-of-Questions approach is to enable systematic decomposition into sub-questions. (Dua et al., 2022). By decomposing complex questions into simpler sub-questions, the model can focus on answering each sub-component accurately, thereby reducing the accumulation of errors that often occur in single-steps. As shown in Figure 3, the Chain-of-Questions prompting section features two open-ended questions (Ling et al., 2023) along with their corresponding answers, while another question is posed without an answer. In this scenario, the reasoning follows a clear and logical path.

3.4 Proposed Method

For each tweet, we extracted a relevant subgraph of information and generated a chain of questions through initial prompting designed to infer the stance of the tweet based on its contextual associations.

Figure 2 describes the overall workflow and association between modules and Figure 3 how the context is provided to LLM.

Let $\mathbf{D} = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N\}$ be the dataset, where $\mathbf{r}_i$ represents the i-th row. The iteration over the dataset is represented as:

$$\forall i \in \{1, 2, \ldots, N\}, \quad y_i = f(t_i, r, g(S_i), U_i, M)$$

Where y_i represents the stance for the i-th row. t_i is the translated tweet for the i-th row. r means text with historical relationships which is constant across all rows. S_i denotes structured data retrieved from knowledge graph for the i-th row. U_i is unstructured chunks of texts for the i-th row. M represents the baseline LLM model used by f for prediction. $g(S_i)$ function converts S_i (structured data) into a string representation.

After the stance function is iterated over all rows:

$$Y = \{y_i \mid i \in \{1, 2, \ldots, N\}\}$$

Here, Y is the final collection of political stances. y_i is predicted stance for the i-th row. $i \in \{1, 2, \ldots, N\}$ iterates over all N rows in the dataset.

4 Experimental Results

This section presents the experimental outcomes evaluating the proposed methods through quantitative metrics and qualitative analysis, emphasizing their strengths and limitations.

4.1 Experimental Settings

The experiment setup describes the data used, how it was prepared, and the steps involved. A brief overview is given here, with more details in Section A of the Appendix.

Initial Data Processing We collected tweets and extracted key entities such as persons, locations, and organizations.

Build & Store Knowledge Graph Using these entities, we pulled information from Wikipedia, structured it into a Knowledge Graph, and stored it in Neo4j.

Baseline Performance We tested baseline language models on the dataset to obtain initial results.

Integrating Our Method For each tweet, we retrieved relevant subgraphs and combined them with chain-of-questions in the model for enhanced stance detection.

Other Datasets We repeated the process on BPDisC and six benchmark datasets, consistently outperforming baselines.

4.1.1 BPDisC Dataset

Data Collection The dataset covers large-scale political discourse on Twitter during the 2024 Bangladesh Protests, collected between July 1 and December 31, 2024. We gathered around 10,000 publicly available tweets by employing keyword searches, hashtags, regional settings, and specific timestamps. Each tweet in the dataset is accompanied by metadata, including tweet id, handle name, timestamps, replies, retweets, likes. Tweets were obtained through the X (Twitter) API, this methodology aligns with the approaches utilized in earlier research studies (Mohammad et al., 2016b) and (Hossny and Mitchell, 2018).

Pre-processing After collecting tweets, we applied preprocessing by excluding image-based posts, users without clear political affiliation, and tweets under five words. Following these exclusions, we refined the dataset to 2,847 tweets. The dataset contained tweets in both Bangla and English. For non-English text within this final dataset, we utilized machine translation via GPT-4o to ensure data clarity and uniform analysis.

Node ID	Type	Context
Ahidul Islam	Person	In 2021, seven children of freedom fighters, including Ahidul Islam, filed a writ petition challenging the quota system decision.
High Court	Organization	On 5 June 2024, High Court declared the decision to scrap the quota system invalid.
Students	Group	Key protest action staged by students during the intensified July demonstrations.
Police	Organization	Police used excessive force during protests, leading to clashes while attempting to quell student agitations.
Bangla Blockade	Event	Immediately after the verdict announcement, students protested in various universities. The movement intensified in July with blockades like "Bangla Blockade".

Table 2: Structured Extraction of Entities and Relationships from Wikipedia using LLM—The table represents extracted nodes (entities) and their contextual relationships from Wikipedia text.

Source Node	Relation	Target Node
Ahidul Islam	FILED	High Court
High Court	DECLARED	Quota System
Students	STAGED	Bangla Blockade
Police	USED_FORCE	Students

Table 3: Extracted Relationships using LLM Demonstrating Connections Between Key Entities.

Annotation Strategy Following prior strategies on inferring political affiliation from online profiles (Baran et al., 2022; Huszár et al., 2022; Johnson and Goldwasser, 2016), we manually annotated a subset of Twitter user accounts based on political cues found in their bios. Specifically, we identified users' stances in favor or against the Awami League, the most frequently referenced political party in our dataset. 32 Twitter accounts were labeled as in favor of the Awami League, while 17 accounts were labeled as against. These annotations were based on explicit expressions in user bios, such as slogans, hashtags, or mentions of party leaders. Accounts without such information were excluded, along with celebrity and news media outlets, as they generally maintain a neutral stance. Hence, we had to drop a massive portion of originally collected ~10k tweets.

Once annotated, the political stance label of each account was propagated to all tweets posted by that user in the dataset. This approach allowed us to construct a larger dataset of labeled tweets based on account-level annotations. In total, this yielded 1,512 tweets labeled as class 1 (in favor) and 1,335 tweets as class 0 (against), forming a final dataset suitable for supervised learning and analysis.

4.1.2 Inter-Annotator Agreement

We conducted a two-stage inter-annotator agreement process to ensure the reliability of the BPDisC dataset. This validation was performed first at the account level to confirm the user-level annotations and subsequently at the tweet level to validate the propagation of those labels. We used Cohen's Kappa(Cohen, 1960) to measure the consistency at both stages.

Account Level For the user-level annotations, two coders independently assessed 49 Twitter accounts (32 labeled as 'in favor' and 17 as 'against'), based on the explicit political cues in their bios. It was also confirmed that none of the 49 accounts were linked to recognized media outlets in Bangladesh which are generally regarded as politically neutral. The two coders reached an agreement on 47 out of the 49 accounts. To measure the consistency of this process, we calculated the inter-annotator agreement, achieving a **Cohen's Kappa of 0.91**, which indicates strong agreement and confirms the reliability of our account-level annotation.

Tweet Level To further validate the dataset's annotation, we also assessed the reliability of propagating account-level labels to individual tweets. A random subset of 300 tweets was independently annotated by two coders to determine if the tweet's content matched the account's propagated stance. The annotators agreed on 281 of the 300 tweets. This resulted in a **Cohen's Kappa of 0.87**. This strong level of agreement confirms that the stances expressed

Methods	LLMs	Accuracy	Precision	Recall	F1 Score
Zero-shot	GPT-4	58.00	80.04	27.85	41.30
	GPT-4o	59.61	84.81	29.17	43.30
	Gemini 1.5 Pro	59.64	75.17	35.85	48.54
	Mistral Large Latest	53.39	70.79	20.83	32.19
	DeepSeek R1	61.40	75.40	40.54	52.73
Few-shot	GPT-4	51.07	63.53	18.33	28.39
	GPT-4o	53.07	63.66	27.12	38.03
	Gemini 1.5 Pro	51.46	64.77	18.85	29.24
	Mistral Large Latest	50.09	60.76	17.00	26.56
	DeepSeek R1	62.59	81.26	38.43	52.18
Few Shot with Naive-RAG	GPT-4	56.41	78.60	24.54	37.50
	GPT-4o	63.96	83.11	40.34	54.32
	Gemini 1.5 Pro	57.50	81.29	25.86	39.24
	Mistral Large Latest	54.97	74.47	23.15	35.32
	DeepSeek R1	64.56	75.90	48.74	59.36
GRASP-ChoQ (Ours)	GPT-4	88.55	88.61	90.01	89.30
	GPT-4o	85.04	87.81	83.40	85.85
	Gemini 1.5 Pro	75.10	87.49	61.97	72.55
	Mistral Large Latest	72.60	83.83	59.99	69.93
	DeepSeek R1	93.12	96.14	90.67	**93.33**

Table 4: Comparison of different methods (Zero-shot, Few-shot, Naive-RAG, and GRASP-ChoQ) across models (Mistral, GPT-4, GPT-4o, Gemini 1.5 Pro, and DeepSeek R1) in terms of Accuracy, Precision, Recall, and F1 Score on **BPDisC** Dataset.

in individual tweets are highly consistent with the account-level labels, validating the reliability of the BPDisC dataset.

4.1.3 Other Datasets

In addition to our **BPDisC** dataset, we evaluated our approach on six other datasets.

For English-language, we utilized SemEval-2016 Task-6 and P-stance dataset (Mohammad et al., 2016a; Li et al., 2021). The dataset from (Swami et al., 2018) includes English–Hindi code-mixed tweets related to India's 2016 demonetization.

Beyond English, we employed non-English datasets: *Mawqif*, an Arabic-dialect Twitter corpus (Alturayeif et al., 2022); the dataset from (Lai et al., 2018), comprising Italian tweets related to the 2016 referendum on constitutional reform; and (Lüüsi et al., 2024), which features Estonian news media sentences discussing immigration. The datasets are detailed in Table 1.

We first removed stances labeled as *None*, *Neither*, or *Neutral* from the datasets. Next, we performed translation, entity extraction, and knowledge graph construction. Finally, we tailored prompts to each dataset and applied both zero-shot and GRASP-ChoQ methods.

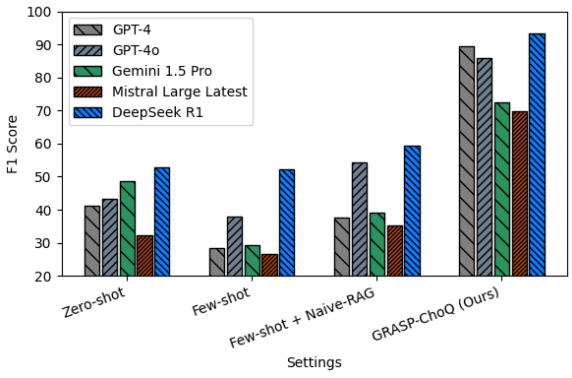

Figure 4: DeepSeek R1 Demonstrates Superior Performance Across All Input Settings.

4.1.4 Neo4j Configuration

Texts sourced from Wikipedia (Yu et al., 2021) were transformed into a graph database using Neo4j (Miller, 2013), a platform tailored for managing highly interconnected knowledge graphs. The database was represented as a graph comprising 883 nodes and 2,281 relationships for the BPDisC dataset.

In this graph, nodes represented key entities such

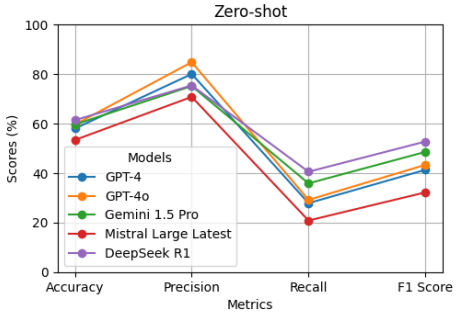

Figure 5: The performance of five AI models (GPT-4, GPT-4o, Gemini 1.5 Pro, Mistral Large Latest, and DeepSeek R1) across four evaluation metrics in a zero-shot setting.

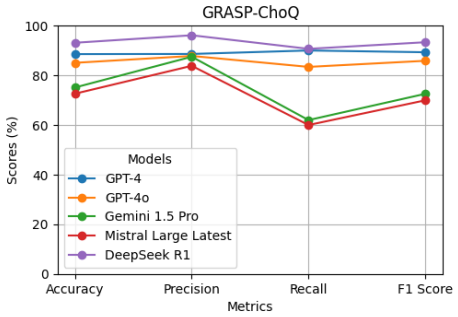

Figure 6: The performance of five AI models across four evaluation metrics for GRASP-ChoQ.

as users, topics, or hashtags, while relationships captured interactions.

Neo4j's Cypher query language enabled extraction of related subgraphs (Anuyah et al., 2024).

4.1.5 Baseline LLMs

Mistral Large Latest Developed by French startup Mistral AI (Jiang et al., 2023), this top-tier model builds on earlier versions with billions of parameters.

OpenAI GPT-4 The latest in the GPT series, GPT-4 surpasses GPT-3.5 in context handling, reasoning, and comprehension (OpenAI, 2023).

OpenAI GPT-4o An optimized GPT-4 variant offering faster, more accurate, and versatile performance (OpenAI, 2024).

Gemini 1.5 Pro DeepMind's multimodal model improves on Gemini 1.0 with longer context, faster processing, and enhanced performance (DeepMind, 2023).

DeepSeek R1 Launched in 2023 by DeepSeek AI, this open-source model is adopted by major firms like Microsoft and Amazon for its cost-effectiveness and accessibility(DeepSeekAI, 2025).

4.2 Ablation Study

We performed an ablation study by excluding two fundamental elements of GRASP-ChoQ:

- Retrieval and knowledge graph data integration into the reasoning process.
- ChoQ framework for breaking down intricate searches into sequential, linked inquiries.

Our analysis reveals that simple RAG is insufficient for complex political texts. While the DeepSeek-R1 baseline achieved a 52.73 F1 score, our full GRASP-ChoQ framework reached 93.33 F1. This massive +40.6 point improvement confirms the critical synergy between our components. The high performance is not attributable to just one part, but to the powerful combination of graph-based retrieval and structured reasoning.

Our ablation study confirms the components are complementary. From Table 5, we can clearly see the improvements of baseline models with both components.

Knowledge Graph provides superior, structured information between entities. Removing it untethered the model from factual evidence, causing it to rely on simple keywords rather than contextual relationships. The essential reasoning structure is supplied by Chain-of-Questions. When removed, the model lost its ability to perform multi-step logic, defaulting to single-turn predictions that failed to interpret subtle or indirect viewpoints.

These results prove that the model's success stems directly from the synergy between its parts: the Knowledge Graph provides the external understanding, while ChoQ enforces the disciplined reasoning required to use it. Both are essential for maintaining accuracy and interpretability.

4.3 Performance and Results

Initial evaluations on the BPDisC dataset (Table 4) showed that the DeepSeek-R1 model, enhanced by our proposed GRASP-ChoQ approach, achieved superior performance across all metrics compared to other models. Figure 4 and Figure 5 illustrate DeepSeek R1's strong baseline

Dataset	Model	Zero-shot	GRASP-ChoQ
BPDisC	GPT-4	41.30	89.30 *(+48.0)*
	GPT-4o	43.30	85.85 *(+42.5)*
	Gemini 1.5 Pro	48.54	72.55 *(+24.0)*
	Mistral Large Latest	32.19	69.93 *(+37.7)*
	DeepSeek R1	**52.73**	**93.33** *(+40.6)*
SemEval 2016	DeepSeek R1	95.07	96.81 *(+1.7)*
P-Stance	DeepSeek R1	97.61	98.33 *(+0.7)*
Mawqif	DeepSeek R1	84.37	95.63 *(+11.3)*
Swami et al.	DeepSeek R1	87.31	92.56 ***(+5.3)***
Lai et al.	DeepSeek R1	74.21	90.83 *(+16.6)*
Lüüsi et al.	DeepSeek R1	79.35	92.94 *(+13.6)*

Table 5: F1 Score Comparison of Models on Various Datasets in Zero-shot and GRASP-ChoQ methods

performance across various settings. Based on this finding, we proceeded to evaluate only the DeepSeek-R1 model on the remaining six datasets. GRASP-ChoQ significantly enhanced DeepSeek R1's performance on these datasets as well. Table 5 highlights this improvement.

Understanding Performance Gains: From Table 5, we can see that English language datasets have strong baseline performance, our approach increases F1 scores slightly. This happens because the target entities of these datasets are well-known and LLMs have their information in the pre-training data already. But for other language datasets, where target entities are not so well known, our approach improves DeepSeek-R1's baseline performance.

So, from the experiments and results, it can be observed that, our approach is beneficial for low-resource political texts whose target entities are not well-studied.

5 Conclusion

Large Language Models (LLMs) often demonstrate suboptimal performance when tasked with processing low-resource political information. To date, no prior research has attempted to classify political stances in the context of Bangladesh, presenting a unique opportunity for exploration. Our proposed approach GRASP-ChoQ, offers a novel strategy for addressing this gap. This approach was applied to the novel BPDisC and six other datasets. By fostering critical thinking, this method has the potential to be adapted to other political contexts. This can be achieved by constructing a tailored knowledge base and designing a carefully curated chain of questions specific to each contextual framework.

Limitations

The GRASP-ChoQ method has several limitations. It relies solely on textual data, excluding multimodal inputs like images or videos, which are increasingly important in political communication these days. This restricts the method's ability to capture visual contexts. Moreover, the exclusion of image data means the final dataset is smaller than it could have been. Additionally, our main focus was to reduce the cost of fine-tuning large models, so no additional training was performed, which may limit adaptability. Furthermore, the limited knowledge base may restrict the approach's ability to generalize across diverse political contexts. Accurately identifying political sarcasm—a subtle yet pervasive feature of political discourse—also remains a significant challenge. While the inter-annotator agreement process demonstrated strong reliability at both the account and tweet levels, limitations exist in terms of the small sample sizes. The focus on accounts without recognized media links may not fully capture the political diversity present in the broader Twitter ecosystem.

Ethics and Bias Mitigation

This paper guarantees openness, fairness, and privacy protection by following strict ethical criteria in stance identification for political debate. All the information used in this study came from publicly accessible sources, including X (Twitter). The final dataset was relatively balanced. This distribution reduces the risk of systematic bias in classification outcomes based on label imbalance. However, large language models (LLMs) used in this study may still reflect biases inherited during pretraining on broad internet data.

We ensured that both Pro-Awami League and Anti-Awami League classifications were treated equitably. We have meticulously documented all

data preprocessing steps and KG made to facilitate independent verification. We conducted this research solely for academic and research purposes, with a commitment to upholding the ethical standards established for public data analysis. No aspect of this research is intended to be used for political profiling, influence campaigns, or any activities that may lead to the manipulation or misrepresentation of political discourse. We acknowledge the ethical implications of automated stance detection in sensitive political contexts and emphasize the responsible use of such technologies.

References

Nora Saleh Alturayeif, Hamzah Abdullah Luqman, and Moataz Aly Kamaleldin Ahmed. 2022. Mawqif: A multi-label arabic dataset for target-specific stance detection. In *Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 174–184.

Sydney Anuyah, Victor Bolade, and Oluwatosin Agbaakin. 2024. Understanding graph databases: a comprehensive tutorial and survey. *arXiv preprint arXiv:2411.09999*.

Samin Aref, Ly Dinh, Rezvaneh Rezapour, and Jana Diesner. 2020. Multilevel structural evaluation of signed directed social networks based on balance theory. *Scientific reports*, 10(1):15228.

Joanna Baran, Michał Kajstura, Maciej Ziółkowski, and Krzysztof Rajda. 2022. Does twitter know your political views? politweets dataset and semi-automatic method for political leaning discovery. *arXiv preprint arXiv:2207.07586*.

Scott Barnett, Stefanus Kurniawan, Srikanth Thudumu, Zach Brannelly, and Mohamed Abdelrazek. 2024. Seven failure points when engineering a retrieval augmented generation system. In *Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering-Software Engineering for AI*, pages 194–199.

Samuel E Bestvater and Burt L Monroe. 2023. Sentiment is not stance: Target-aware opinion classification for political text analysis. *Political Analysis*, 31(2):235–256.

Tianyu Chen, Yiming Zhang, Guoxin Yu, Dapeng Zhang, Li Zeng, Qing He, and Xiang Ao. 2024a. EFSA: Towards event-level financial sentiment analysis. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7455–7467. Association for Computational Linguistics.

Zhongwu Chen, Long Bai, Zixuan Li, Zhen Huang, Xiaolong Jin, and Yong Dou. 2024b. A new pipeline for knowledge graph reasoning enhanced by large language models without fine-tuning. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 1366–1381, Miami, Florida, USA. Association for Computational Linguistics.

Ning Cheng, Zhaohui Yan, Ziming Wang, Zhijie Li, Jiaming Yu, Zilong Zheng, Kewei Tu, Jinan Xu, and Wenjuan Han. 2024. Potential and limitations of llms in capturing structured semantics: A case study on srl. In *International Conference on Intelligent Computing*, pages 50–61. Springer.

Alla Chepurova, Aydar Bulatov, Yuri Kuratov, and Mikhail Burtsev. 2023. Better together: Enhancing generative knowledge graph completion with language models and neighborhood information. *arXiv preprint arXiv:2311.01326*.

Saeyd Rashed Hasan Chowdury. 2024. The role of political parties in bangladesh's july revolution of 2024: Insights from sufi perspectives. *International Journal of Research and Innovation in Social Science*, 8(11):2077–2093.

James R Clough, Jamie Gollings, Tamar V Loach, and Tim S Evans. 2015. Transitive reduction of citation networks. *Journal of Complex Networks*, 3(2):189–203.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.

Google DeepMind. 2023. Gemini 1.5: Unlocking multimodal understanding across millions of tokens.

DeepSeekAI. 2025. Deepseek-r1: Incentivizing reasoning capability in large language models via reinforcement learning.

Jialin Dong, Bahare Fatemi, Bryan Perozzi, Lin F Yang, and Anton Tsitsulin. 2024. Don't forget to connect! improving rag with graph-based reranking. *arXiv preprint arXiv:2405.18414*.

Dheeru Dua, Shivanshu Gupta, Sameer Singh, and Matt Gardner. 2022. Successive prompting for decomposing complex questions. *arXiv preprint arXiv:2212.04092*.

Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, and Jonathan Larson. 2024. From local to global: A graph rag approach to query-focused summarization. *arXiv preprint arXiv:2404.16130*.

Karan Garg and Cornelia Caragea. 2024. Stanceformer: Target-aware transformer for stance detection. *arXiv preprint*, arXiv:2410.07083.

Fritz Heider. 2013. *The psychology of interpersonal relations*. Psychology Press.

Ahmad Hany Hossny and Lewis Mitchell. 2018. Event detection in twitter: A keyword volume approach. In *2018 IEEE international conference on data mining workshops (ICDMW)*, pages 1200–1208. IEEE.

Ferenc Huszár, Sofia Ira Ktena, Conor O'Brien, Luca Belli, Andrew Schlaikjer, and Moritz Hardt. 2022. Algorithmic amplification of politics on twitter. *Proceedings of the national academy of sciences*, 119(1):e2025334119.

Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. 2023. Mistral 7b. *arXiv preprint arXiv:2310.06825*.

Kristen Johnson and Dan Goldwasser. 2016. Identifying stance by analyzing political discourse on twitter. In *Proceedings of the First Workshop on NLP and Computational Social Science*, pages 66–75.

Elizaveta Kopacheva, Masoud Fatemi, and Kostiantyn Kucher. 2023. Using social-media-network ties for predicting intended protest participation in russia. *OsnEM*, n/a:n/a.

Mirko Lai, Viviana Patti, Giancarlo Ruffo, and Paolo Rosso. 2018. Stance evolution and twitter interactions in an italian political debate. In *Natural Language Processing and Information Systems: 23rd International Conference on Applications of Natural Language to Information Systems, NLDB 2018, Paris, France, June 13-15, 2018, Proceedings 23*, pages 15–27. Springer.

Philip Leifeld and Laurence Brandenberger. 2019. Endogenous coalition formation in policy debates. *arXiv preprint arXiv:1904.05327*.

Y. Li, R. Zhang, and J. Liu. 2024. An enhanced prompt-based llm reasoning scheme via knowledge graph-integrated collaboration. In *International Conference on Artificial Neural Networks*, pages 251–265, Cham. Springer Nature Switzerland.

Yingjie Li, Tiberiu Sosea, Aditya Sawant, Ajith Jayaraman Nair, Diana Inkpen, and Cornelia Caragea. 2021. P-stance: A large dataset for stance detection in political domain. In *Findings of the association for computational linguistics: ACL-IJCNLP 2021*, pages 2355–2365.

Chen Ling, Xuchao Zhang, Xujiang Zhao, Yanchi Liu, Wei Cheng, Mika Oishi, Takao Osaki, Katsushi Matsuda, Haifeng Chen, and Liang Zhao. 2023. Open-ended commonsense reasoning with unrestricted answer candidates. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 8035–8047.

Yang Liu, Jiahuan Cao, Chongyu Liu, Kai Ding, and Lianwen Jin. 2024. Datasets for large language models: A comprehensive survey. *arXiv preprint arXiv:2402.18041*.

Yujian Liu, Xinliang Frederick Zhang, David Wegsman, Nick Beauchamp, and Lu Wang. 2022. Politics: Pretraining with same-story article comparison for ideology prediction and stance detection. *arXiv preprint arXiv:2205.00619*.

Lauri Lüüsi, Uku Kangur, Roshni Chakraborty, and Rajesh Sharma. 2024. Political stance detection in estonian news media. In *Proceedings of the 9th International Workshop on Computational Linguistics for Uralic Languages*, pages 12–28.

Jianzong Mei, Jing Chen, Wei Lin, Bill Byrne, and Marcus Tomalin. 2024. Improving hateful meme detection through retrieval-guided contrastive learning. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5333–5347. Association for Computational Linguistics.

Justin J Miller. 2013. Graph database applications and concepts with neo4j. In *Proceedings of the southern association for information systems conference, Atlanta, GA, USA*, volume 2324, pages 141–147.

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016a. 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.

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016b. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pages 31–41.

OpenAI. 2023. Gpt-4 technical report.

OpenAI. 2024. Gpt-4o system card.

Michael T Oswald, Meike Fromm, and Elena Broda. 2021. Strategic clustering in right-wing-populism¿green policies' in germany and france. *Zeitschrift für Vergleichende Politikwissenschaft*, 15(2):185–205.

Pranoy Panda, Ankush Agarwal, Chaitanya Devaguptapu, Manohar Kaul, and Prathosh A P. 2024. Holmes: Hyper-relational knowledge graphs for multi-hop question answering using llms. In *Proceedings of the 2024 Annual Meeting of the Association for Computational Linguistics (ACL 2024)*.

David Rozado. 2024. The political preferences of llms. *arXiv preprint arXiv:2402.01789*.

F. Shimin. 2024. Application research on large language model attention mechanism in automatic classification of book content. In *2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA)*, pages 343–350, Shenyang, China. IEEE.

H. Shui, Y. Zhu, F. Zhuo, Y. Sun, and D. Li. 2024. An emotion text classification model based on llama3-8b using lora technique. In *2024 7th International Conference on Computer Information Science and Application Technology (CISAT)*, pages 380–383, Hangzhou, China. IEEE.

K. Sreekala, M. Sivajyothi, D. Chiranjevi, G. Sunitha, and M. Swetha. 2024. A novel integration of hierarchical clustering and ensemble classification algorithms for news classification. In *2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC - ROBINS)*, pages 524–528, Coimbatore, India. IEEE.

Sahil Swami, Ankush Khandelwal, Vinay Singh, Syed Sarfaraz Akhtar, and Manish Shrivastava. 2018. An english-hindi code-mixed corpus: Stance annotation and baseline system. *arXiv preprint arXiv:1805.11868*.

The Economist. 2024. The economist's country of the year for 2024. *The Economist*.

Michael Weinzierl and Sanda Harabagiu. 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)*, pages 861–880. Association for Computational Linguistics.

J. Wu, J. Zhu, Y. Qi, J. Chen, M. Xu, F. Menolascina, and V. Grau. 2024. Medical graph rag: Towards safe medical large language model via graph retrieval-augmented generation. *arXiv preprint arXiv:2408.04187*.

Weijian Xie, Xuefeng Liang, Yuhui Liu, Kaihua Ni, Hong Cheng, and Zetian Hu. 2024. Weknow-rag: An adaptive approach for retrieval-augmented generation integrating web search and knowledge graphs. In *Proceedings of the KDD Cup CRAG Workshop 2024*, Barcelona, Spain. ACM.

Haoze Yu, Haisheng Li, Dianhui Mao, and Qiang Cai. 2021. A domain knowledge graph construction method based on wikipedia. *Journal of Information Science*, 47(6):783–793.

Wang Zhu, Jesse Thomason, and Robin Jia. 2023. Chain-of-questions training with latent answers for robust multistep question answering. *arXiv preprint arXiv:2305.14901*.

Appendix

A Experimental Setup

All prompting experiments were conducted using GPT-4, GPT-4o, Gemini 1.5 Pro, DeepSeek R1 and Mistral Large (latest), with a temperature setting of 0 and all other parameters maintained at their default values to ensure consistency and reproducibility. We used the OpenAI SDK for interfacing with all models, thereby streamlining the experimental process and ensuring a uniform execution environment. Consistent hyperparameters were applied across the board, ensuring that any performance differences could be confidently attributed to the intrinsic capabilities of each model rather than to variations in the configuration or interfacing method.

A.1 Dataset Details

Group	Distinct Handles
Against	@ZulkarnainSaer, @MushfiqulFazal, @mrforayeji, @redwanxyz, @theBDarmy, @tasneem, @UNinIndia, @support_yunus, @shafiqalam2024, @EUinBangladesh, @BNPBdMediaCell, @volker_turk, @muktadirnewage, @Oliver_Tomarket, @JonFDanilowicz, @ChiefAdviserGoB, @dhruvrahtee
Favor	@alliance29464, @sajeebwazed, @Chellaney, @albd1971, @bdperspectives, @ATeam_1971, @bdwatch2024, @k_shayera, @VNouka, @NeenaRai, @Asifurrahman71, @pressxpresspx, @sumon_tarek, @MAarafat71, @BD_DiGEST, @AsadZam89687147, @RUSI_org, @CJBdingo25, @istiak_ahmmad, @Udashi_Pothik, @MaryMillben, @OnceAgainHasina, @SushantaDGupta, @INSIGHTUK2, @amnestysasia, @IndiaToday, @iindrojit, @Bangladesh_Fact, @TimesAlgebraIND, @CChoddogram, @FreedomRightsRL, @sagor250

Table 6: Distinct Twitter handles grouped by stance

A.2 Tweet Translation Prompt

A.2.1 For BPDisC Dataset

For translating tweets from Bangla to English, preserving proper nouns (like names of people and organizations), and leaving English tweets unchanged, we use the following prompt structure. Few-shot examples are included to guide the Large Language Model (LLM).

LOC	Count	PER	Count	ORG	Count
bangladesh	2661	yunus	991	government	921
country	484	hasina	717	league	873
dhaka	276	sheikh	608	awami	731
india	242	rahman	135	police	441
bangladeshi	183	bangabandhu	128	bnp	366
state	145	muhammad	122	army	309
bangladeshcrisis	133	asif	108	jamaat	197
uk	116	nazrul	101	chhatra	148

Table 7: Top named entities by type with mention frequency

Task: Translate the following tweet from Bangla to English. If the tweet is already in English, output the original tweet. Do NOT translate proper nouns (e.g., names of people, organizations, specific places).

Examples: Use the following examples to understand the task better.

Tweet (Bangla): প্রধানমন্ত্রী শেখ হাসিনা আজ একটি নতুন প্রকল্পের উদ্বোধন করেছেন।
Translated Tweet (English): Prime Minister Sheikh Hasina inaugurated a new project today.
Tweet (English): Just attended the Google I/O conference.
Translated Tweet (English): Just attended the Google I/O conference.

Tweet (Bangla): BRAC বিশ্বের অন্যতম বৃহত্তম এনজিও।
Translated Tweet (English): BRAC is one of the world's largest NGOs.

A.2.2 For Other Datasets

To support multilingual datasets, we use the following prompt structure. The Language Model (LLM) is instructed to translate non-English tweets into English, while preserving proper nouns (e.g., names of people, organizations, or specific places), and to leave tweets already in English unchanged. Few-shot examples are included to guide the model's behavior.

Figure 7: Most frequent entities from the collected tweets

Task: Translate the following tweet into English. If the tweet is already in English, output the original tweet. Do **NOT** translate proper nouns (e.g., names of people, organizations, specific places).

Examples: Use the following examples to understand the task better.

Tweet (Non-English): [Example tweet in a non-English language with proper nouns]
Translated Tweet (English): [Correct English translation, preserving proper nouns]

Tweet (English): [Example English tweet]
Translated Tweet (English): [Same English tweet]

Tweet (Non-English): [Another example tweet in a different non-English language]
Translated Tweet (English): [Correct English translation, preserving proper nouns]

A.3 Graph Construction from Supporting Documents

To build a knowledge graph from unstructured text, we use Wikipedia as a source of supporting documents. We retrieve relevant content based on a user query, split the documents into smaller chunks, convert them into graph-structured representations using a language model, and finally populate the graph. This structured format helps capture both entities and relationships for downstream

reasoning.

Step 1: Load Wikipedia content using a query.

```
from langchain.document_loaders
import WikipediaLoader

raw_documents = WikipediaLoader(
query="bangladesh student uprising
of july 2024"
).load()
```

Step 2: Split documents into token-sized chunks.

```
from langchain.text_splitter
import TokenTextSplitter

text_splitter = TokenTextSplitter(
chunk_size=256, chunk_overlap=50)
documents = text_splitter.
split_documents(
raw_documents)
```

Step 3: Convert document chunks into graph-structured format.

```
from langchain_experimental.
graph_transformers
import LLMGraphTransformer
llm_transformer =
LLMGraphTransformer(llm=llm)

graph_documents = llm_transformer.
convert_to_graph_documents(
[raw_documents])
```

Step 4: Add graph documents to the knowledge graph.

```
graph.add_graph_documents(
    graph_documents,
    baseEntityLabel=True,
    include_source=True
)
```

A.4 Hybrid Vector Indexing and Full-Text Search

To support both dense and keyword-based retrieval, we create a hybrid vector index using sentence embeddings from a pre-trained model. These embeddings are stored in a Neo4j-backed vector store. Additionally, a full-text index is created for entity-level lookup to complement semantic search with symbolic filtering.

Step 1: Define the embedding model.

```
from langchain.embeddings import
HuggingFaceEmbeddings
embeddings = HuggingFaceEmbeddings(
model_name="sentence-transformers/
all-MiniLM-L6-v2"
)
```

Step 2: Create a hybrid vector index from the graph.

```
from langchain.vectorstores import
Neo4jVector
vector_index                    =
Neo4jVector.from_existing_graph(
embeddings,
search_type="hybrid",
node_label="Document",
text_node_properties=["text"],
embedding_node_property="embedding"
)
```

Step 3: Create a full-text index on entity nodes.

```
graph.query("CREATE FULLTEXT INDEX
entity IF NOT EXISTS
FOR (e:__Entity__) ON EACH [e.id]")
```

A.5 Entity Extraction and Full-Text Query Generation

To extract entities (person, organization, business) from the input text and generate full-text queries for a Neo4j-based search, we use the following approach.

Step 1: Define the entity extraction model.

```
from    langchain_core.pydantic_v1
import BaseModel, Field
from typing import List
class Entities(BaseModel):
# Identifying information about
entities.
names:    List[str]    =    Field(...,
description="All    the    person,
organization,  location  entities
that appear in the text")
```

Step 2: Set up the prompt and model chain for extraction.

```
from langchain_core.prompts import ChatPromptTemplate
prompt = ChatPromptTemplate.from_messages([
("system", "You are extracting location, organization and person entities from the text."),
("human", "Use the given format to extract information from the following input: question")
])
entity_chain = prompt | llm.with_structured_output(Entities)
```

Step 3: Define the full-text query generation function.

```
from langchain_community.vectorstores.neo4j_vector import remove_lucene_chars
def generate_full_text_query(input: str) -> str:
full_text_query = ""
words = [el for el in remove_lucene_chars(input).split() if el]
for word in words[:-1]:
full_text_query += f" word 2 AND"
full_text_query += f" words[-1] 2"
return full_text_query.strip()
```

Step 4: Implement structured retrieval using entity names.

```
def structured_retriever(question: str) -> str:
time.sleep(0.5)
result = ""
entities = entity_chain.invoke("question": question)
for entity in entities.names:
response = graph.query(
"CALL db.index.fulltext.queryNodes(
'entity', $query, limit:2)"
YIELD node,score
CALL {
WITH node
MATCH (node)-[r:!MENTIONS]->(neighbor)
RETURN node.id + ' - ' + type(r) +
' -> ' + neighbor.id AS output
UNION ALL
WITH node
MATCH (node)<-[r:!MENTIONS]-(neighbor)
RETURN neighbor.id + ' - ' + type(r)
+ ' -> ' + node.id AS output

RETURN output LIMIT 50
)
result += "
n".join([el['output'] for el in response])
return result
```

B Baseline & GRASP-ChoQ Prompts

B.1 Prompt used in zero-shot

Target Entity: Awami League
Task: Read the tweet and determine whether it expresses a stance **in FAVOR of** or **AGAINST** the specified target entity.

Tweet: [tweet]

B.2 Prompt used in few-shot

Target Entity: Awami League
Task: Analyze the following tweets and determine if the author's stance is **in FAVOR of** or **AGAINST** the specified target entity.

Tweet: "The country is moving forward under the leadership of Sheikh Hasina. #AwamiLeague"
Stance: Favor

Tweet: "Corruption is rampant, and the government is not listening to the people. #Bangladesh"
Stance: Against

Tweet: "{New tweet text goes here}"
Stance:

B.3 Prompt used in few-shot with Naive-RAG

Target Entity: Awami League
Context: [retrieved context]
Task: Analyze the following tweets and determine if the author's stance

is **in FAVOR of** or **AGAINST** the specified target entity, using the tweet and context provided.

Tweet: "The country is moving forward under the leadership of Sheikh Hasina. #AwamiLeague"
Context: "Sheikh Hasina is the leader of the Awami League and has been praised for infrastructure development."
Stance: Favor

Tweet: "Corruption is rampant, and the government is not listening to the people. #Bangladesh"
Context: "The Awami League has been criticized in the media for alleged corruption and authoritarian practices."
Stance: Against

Tweet: "{New tweet text goes here}"
Context: "{Retrieved context goes here}"
Stance:

B.4 Prompts Used in GRASP-ChoQ For BPDisC Dataset

We design prompts to evaluate whether a given tweet exhibits a political stance **in favor of** or **against** the **Awami League**, a major political party in Bangladesh. The prompt integrates contextual reasoning and guided questioning to emulate how annotators or language models might use external political knowledge to infer implicit stance.

Tweet: The head of the UN Human Rights Commission's visit to Bangladesh raises concerns over the illegal Yunus government!

Read the tweet above. The tweet has a political stance. It may express a view either **in favor of the Awami League of Bangladesh** or **against it**. Detect the stance of the tweet with respect to the Awami League. Use reasoning based on political references or implied affiliations.

ASK QUESTIONS TO DETECT STANCE:

Q: Who is being criticized here? A: Muhammad Yunus. Because the tweet uses "illegal" to describe him. Since Yunus is opposed to Sheikh Hasina (leader of Awami League), this implies support for Awami League.
Q: Which government is depicted in power in the tweet? A: Muhammad Yunus's government. As he is seen to follow Hasina, and is portrayed negatively, the stance favors the Awami League.

To aid your decision, general background knowledge about political figures and affiliations is provided.

INFO_FROM_KNOWLEDGE_GRAPH:
Sheikh_Hasina - LEADER -> Awami_League
Sheikh_Hasina - RESIGNATION_DUE_TO -> Bangladesh_Protests_Of_2022–24
Sheikh_Hasina - PREDECESSOR -> Muhammad_Yunus
Student–People'S_Uprising - OVERTHROWN -> Sheikh_Hasina
Non-Cooperation_Movement - RESULTED_IN_RESIGNATION -> Sheikh_Hasina
Muhammad_Yunus - LEADER -> Interim_Government_Of_Bangladesh
Muhammad_Yunus - INTERIM_LEADER -> Anti-Discrimination_Students_Movement
Anti-Discrimination_Students_Movement - PROPOSED -> Muhammad_Yunus
Mohammed_Shahabuddin - ADMINISTERED_THE_OATH -> Muhammad_Yunus
Anti-Discrimination_Students_Movement - CALLEDTOLEAD -> Muhammad_Yunus
Hasnat_Abdullah - REQUESTED_TO_HEAD -> Muhammad_Yunus

HISTORICAL_CONTEXT:
Sheikh Hasina's Enemies:
Student Movements
Anti-Awami League Factions
Bangladesh Nationalist Party (BNP)
Jamaat-e-Islami
Civil Society Activists
Muhammad Yunus

Sheikh Hasina's Allies:
Awami League Loyalists
Chhatra League (Student wing of the Awami League)
Groups Supporting the Liberation War's Legacy
Supporters Valuing the Awami League's Historical Role

Tweet: Cancel your Netflix subscription and donate to @realDonaldTrump's campaign for entertainment #Trump #SemST

Original Stance: Against Donald Trump

Response After Zero Shot

The tweet encourages canceling a Netflix subscription to donate to Donald Trump's campaign, using supportive hashtags. This directly promotes financial support for Trump, indicating a stance **FAVOR** of him.

Response after GRASP-ChoQ

The tweet is **AGAINST Donald Trump**.

Key Analysis:
- **Mismatched Tone & Context:** The call to "donate to @realDonaldTrump's campaign for entertainment" uses sarcasm. Donating to a political campaign is not typically framed as "entertainment," implying that supporting Trump is absurd or laughable.
- **Trivialization:** Equating political support with "entertainment" undermines the seriousness of Trump's campaign, suggesting it is a spectacle rather than a legitimate effort.
- **Hashtag Ambiguity:** While #Trump is neutral, the sarcastic framing inverts the literal message.

The tweet mocks Trump by treating his campaign as a joke, signaling opposition.

Figure 8: Comparison of stance classification outputs for a sarcastic tweet originally labeled as "Against Donald Trump." The Zero-Shot response misinterprets the literal content, incorrectly inferring a stance in favor of Trump. In contrast, the GRASP-ChoQ model accurately identifies sarcasm and contextual cues, correctly classifying the stance as against Trump. Key interpretive factors include mismatched tone, trivialization, and hashtag ambiguity.

Tweet: #JoeBiden #Biden #WhistleblowerComplaint Why a surprise, nothing new with moving transcript to top secret server.

Original Stance: Against Joe Biden

Response After Zero Shot

The tweet is in FAVOR of Joe Biden.
They argue that moving the transcript to a top-secret server—a decision now under scrutiny—is a normal way to protect sensitive information, not a sign of wrongdoing. This framing downplays the controversy, defends Biden's transparency, and casts critics' claims as exaggerated or politically driven.

Response after GRASP-ChoQ

The tweet is AGAINST Joe Biden. The sarcastic tone ("Why a surprise, nothing new") and the reference to a "whistleblower complaint" involving moving a transcript to a "top secret server" imply criticism of Biden's actions, suggesting unethical or secretive behavior. This aligns with common conservative critiques of Biden's transparency.

Figure 9: Comparison of stance classification outputs for a sarcastic tweet originally labeled as "Against Joe Biden". The tweet is from SemEval 2016 dataset. The Zero-Shot response fails to detect sarcasm, incorrectly interpreting the tweet as supportive of Biden. In contrast, the GRASP-ChoQ model correctly infers opposition by identifying the sarcastic tone and contextual cues such as references to secrecy and a whistleblower complaint—framing consistent with common criticisms of Biden's transparency.

C Examples of LLM Outputs

These examples highlight the significant challenges Large Language Models (LLMs) face in zero-shot stance classification, particularly when dealing with nuanced or non-literal language such as sarcasm or hyperbole. As illustrated in Figure 8 and Figure 9, the zero-shot LLM frequently misinterprets sarcastic tweets, inferring a stance directly opposite to the author's true intent by focusing on the literal meaning while failing to detect the ironic tone and contextual cues. Similarly, Figure 10 demonstrates how a zero-shot approach can be misled by repeated negative phrasing, overlooking the broader positive sentiment and contextual nuances that indicate a favorable stance despite momentary frustration. In stark contrast, our proposed GRASP-ChoQ method consistently demonstrates a more robust understanding of

Tweet: More than 3 months ago, I withdrew 100 riyals and said to keep it in the car in case I needed cash at any time.. Until now, I have not even touched my ATM cards and all my ID cards since then.. The electronic transfer has become terrible, terrible. You do not need to take cards or a wallet.. except your phone and your key only.

Original Stance: Favor of Digital Transformation

Response After Zero Shot

The tweet is AGAINST Digital Transformation. The user expresses frustration with electronic transfers, calling them "terrible," which indicates dissatisfaction despite the convenience of not needing physical cards. The negative experience outweighs the benefits, leading to a stance against digital transformation.

Response after GRASP-ChoQ

The tweet expresses a stance in **FAVOR of Digital Transformation**.

Key reasoning:

- The user highlights the convenience of relying solely on their phone (digital tools) for transactions, eliminating the need for physical cards or wallets.

- While they initially call electronic transfers "terrible, terrible," this appears to be hyperbolic or ironic frustration, as the overall narrative emphasizes seamless adoption of digital solutions (e.g., not touching physical cards for months).

- The conclusion reinforces reliance on digital infrastructure ("phone and your key only"), framing it as a net positive shift.

Figure 10: Comparison of stance classification for a tweet originally labeled as "In Favor of Digital Transformation". The tweet is from P-stance dataset. The Zero-Shot model misclassifies the stance as negative, focusing on the repeated phrase “terrible, terrible” and overlooking the tweet’s broader context. In contrast, GRASP-ChoQ correctly interprets the stance as favorable by recognizing the user’s emphasis on the convenience of digital transactions and framing the frustration as possibly hyperbolic or ironic. This nuanced understanding reflects support for digital infrastructure despite momentary dissatisfaction.

these complexities. By leveraging its analytical framework, GRASP-ChoQ accurately identifies sarcasm, interprets contextual references, and discerns underlying tones, leading to correct stance classifications even in challenging cases where the zero-shot LLM fails. Further examples in Figure 11, Figure 12, and Figure 13 show how GRASP-ChoQ successfully revises the zero-shot interpretation, aligning with the true stance by capturing subtleties missed by the baseline model. These comparisons underscore the limitations of zero-shot LLMs for nuanced stance detection and emphasize the superior performance achieved by incorporating the structured analysis provided by our GRASP-ChoQ approach.

D Libraries Used

In this research, we employed the `TokenTextSplitter` from the LangChain library to preprocess and split large raw text documents into smaller, manageable chunks for efficient processing by language models. We configured the `TokenTextSplitter` with a `chunk_size` of 256 tokens and a `chunk_overlap` of 50 tokens. By overlaying 50 tokens, our arrangement guaranteed contextual continuity between subsequent chunks and let us produce small enough chunks for effective processing. The raw documents were then subjected to the `split_documents` technique to produce a list of tokenized sections. The result was examined to confirm the proper amount of pieces and their semantic coherence, therefore verifying that the approach worked for downstream language model tasks.

In our research paper, we utilized the **LangChain** library, specifically the experimental module for graph transformations, to streamline the extraction and structuring of entity relationships from unstructured textual data. By employing the `LLMGraphTransformer` class, we efficiently converted documents into graph-based representations, enabling a more detailed analysis of interconnected entities. Additionally, we leveraged the **LangChain OpenAI integration** to utilize advanced language models like GPT-4, ensuring the accurate identification of nodes and relationships. This approach facilitated the

Tweet: @ArvindKejriwal @JantaKaReporter You are doing such crimes as if you are a note-taking nation..shame on you.
Original Stance: FAVOR of Demonetization

Response After Zero Shot

The tweet is expressing disapproval towards Kejriwal or the media outlet, possibly for their stance or actions related to demonetisation. Since the tweet is critical, it's likely against the demonetisation policy. People who opposed demonetisation often criticized the government's handling of it, the sudden implementation, and the resulting chaos.

The tweet is AGAINST the 2016 Indian banknote demonetisation. The language used is critical and disapproving, suggesting frustration with the policy's implementation and its effects.

Response after GRASP-ChoQ

The tweet is **FAVOR** of the 2016 demonetization.

Key Reasons:

- The tweet attacks Arvind Kejriwal, a prominent critic of demonetization, implying alignment with Modi's pro-demonetization stance.

- The sarcastic phrase "note-taking nation" likely mocks Kejriwal's opposition to demonetization (a policy centered on invalidating banknotes).

- The critical tone toward Kejriwal and the political context (Modi vs. Kejriwal) suggest indirect support for Modi's policies, including demonetization.

While the tweet does not explicitly mention demonetization, the adversarial targeting of Kejriwal—a key opponent of the policy—signals a pro-demonetization stance.

Figure 11: Comparison of tweet interpretations concerning the 2016 Indian demonetization from (Swami et al., 2018). The top section presents the original tweet and its initial stance. The middle section shows an interpretation ("Response After Zero Shot") indicating opposition to demonetization. The bottom section ("Response After GRASP-ChoQ") reveals a revised interpretation favoring demonetization after applying the GRASP-ChoQ analytical framework.

generation of knowledge graphs that supported our research objectives by providing a structured and interpretable format for complex datasets.

This study produced dense vector representations of text using the pre-trained model `sentence-transformers/all-MiniLM-L6-v2` by way of the `Hugging FaceEmbeddings` class from LangChain. These embeddings were coupled with a Neo4j graph database using the `Neo4jVector` class to enable hybrid search capabilities, integrating semantic correspondence alongside graph traversal. We extracted embeddings from network nodes labeled as "Document" and employed precomputed embeddings and text content to allow efficient and context-aware retrieval of information. This method leverages the structured relationships of graph databases to harness the power of semantic embeddings.

This approach generates Lucene-compatible full-text search queries for Neo4j by utilizing the `remove_lucene_chars` method from the `langchain_community.vectorstores.neo4j_vector` library to enable approximate matching with fuzzy search. The initial step sanitizes the input text by removing special characters that could interfere with Lucene syntax. Next, the sanitized text is split into individual words, and each word is appended with the fuzzy search operator ($\sim$2) to allow a Levenshtein distance of up to 2, accommodating minor spelling errors or variations. These words are then combined using the `AND` operator to ensure all terms appear in the search results with fuzzy matching. This robust query-generation method is particularly valuable for Neo4j graph databases with Lucene-based indexing in applications requiring error-tolerant and flexible search capabilities.

E Neo4j Workflow

We created a function `structured_retriever` to process an input tweet and return a string containing structured data extracted from a graph

Tweet: Those who #iovotosi and #bastaunSi of the #referendumcostituzionale #4dicembre, know why migrants arrive in Italy? https://t.co/JLVLJG6eXg

Original Stance: Against the 2016 referendum on the reform of the Italian Constitution

Response After Zero Shot

The tweet is in FAVOR of the Referendum held in Italy on December 4, 2016, about a reform of the Italian Constitution.

Step-by-Step Explanation:

1. **Hashtags Analysis:**
   - The tweet uses #iovotosi ("I vote yes") and #bastaunSi ("enough with yes"), which are pro-referendum hashtags.
   - These hashtags indicate support for the "yes" campaign in the constitutional referendum.

2. **Content Context:**
   - The tweet questions whether supporters know why migrants arrive in Italy, possibly highlighting an issue but does not express opposition to the referendum itself.

3. **Overall Stance:**
   - The use of supportive hashtags suggests the tweet aims to encourage a "yes" vote, aligning with the pro-referendum stance.

Conclusion:
The tweet supports the referendum, encouraging a "yes" vote despite raising a question about migration issues.

Response after GRASP-ChoQ

The tweet is **AGAINST** the 2016 referendum on the reform of the Italian Constitution.

Key indicators:

- **Hashtags**: While the user includes pro-reform hashtags like *#iovotosi* ("I vote yes") and *#bastaunSi* ("enough with a yes"), they are used sarcastically or critically to address supporters of the "Yes" camp, not to express genuine support.

- **Rhetorical framing**: The tweet poses a confrontational question to "Yes" voters about migration, implying they are ignorant of the root causes of migration (a hot-button issue in Italy at the time). This links opposition to the referendum to broader anti-establishment or populist narratives (e.g., distrust of Renzi's government and its policies).

- **Tone**: Dismissive and provocative, aiming to undermine the "Yes" camp by associating the reform with unresolved societal issues (migration), a common tactic among "No" campaigners.

- **Implicit ideological conflict**: Suggests the referendum is part of a disconnect between the pro-reform elite (Renzi) and public concerns, a core "No" campaign theme.

The tweet does not explicitly endorse the reform and instead critiques its supporters, aligning it with the **AGAINST** stance.

Figure 12: Comparison of tweet interpretations regarding the 2016 Italian constitutional referendum from (Lai et al., 2018). The top section displays the original tweet and its initial stance. The middle section presents an interpretation ("Response After Zero Shot") indicating support for the referendum. The bottom section ("Response After GRASP-ChoQ") reveals a revised interpretation opposing the referendum after applying the GRASP-ChoQ analytical framework.

database. Next, the function invokes an entity extraction process via `entity_chain.invoke` to identify key entities within the question. For each extracted entity, it generates a full-text search query using `generate_full_text_query` and executes a corresponding Cypher query on the graph database. This query retrieves nodes that match the entity from a full-text index and then traverses the graph to find related nodes via relationships, deliberately excluding those of type `MENTIONS`. The results from the query—formatted to include node identifiers and relationship types—are concatenated together into a single string, which is returned as the final output.

F Supplementary Datasets History

Extensive data gathered during the student-led protest movement in Bangladesh, spanning **July 15 to August 5, 2024**, marks a turning point in the political history of the country. This

Tweet: Refugees are one thing, but does this also extend to economic migrants? Currently, the latter are not even a topic of discussion.

Original Stance: Against Immigration

Response After Zero Shot

The tweet is in FAVOR of Immigration.

The author highlights the exclusion of economic migrants from discussions, suggesting their inclusion is necessary, indicating support for broader immigration policies.

Response after GRASP-ChoQ

The tweet is **AGAINST** Immigration.

Key Analysis:

- **Distinction Between Groups:** The speaker separates refugees (implied as more acceptable) from economic migrants, suggesting the latter are a distinct concern.

- **Implied Criticism:** By noting economic migrants are "not even a topic of discussion," there's an implicit call to scrutinize or restrict this category, framing them as a potential issue needing attention.

- **Anti-Stance Indicators:** While not overtly hostile, the contrast ("but") and focus on excluding economic migrants from current debates align with anti-immigration narratives that differentiate "deserving" refugees from "undeserving" economic migrants.

Answer: AGAINST Immigration.

Figure 13: Comparison of tweet interpretations regarding immigration from (Lüüsi et al., 2024). The top section presents the original tweet and its initial stance. The middle section shows an interpretation ("Response After Zero Shot") indicating support for immigration. The bottom section ("Response After GRASP-ChoQ") reveals a revised interpretation opposing immigration after applying the GRASP-ChoQ analytical framework.

research effort extracts the data. Inspired by university students opposing the reintroduction of a 30% government job quota for the children of independence war veterans a measure revoked by the Supreme Court in late June 2024 the movement rapidly acquired momentum. Originally subdued, the demonstrations—mostly involving students of Dhaka University—became more intense when members of the ruling Awami League's student branch, the Bangladesh Chatra League (BCL), assaulted protestors. With Abu Sayed, a student shot by police, rising as a symbol of opposition, police actions involving tear gas and live bullets led to over 200 deaths and thousands of injuries. From particular quota changes to more general demands for government responsibility and human rights, the movement's emphasis changed and resulted in the Student-People's Uprising, which inspired popular support across social lines. A large-scale protest in Dhaka on August 5, 2024, drove Prime Minister Sheikh Hasina to resign under government crackdowns and worldwide criticism of violations of human rights. Honoring the resiliency and tenacity of its youth-led movement, Bangladesh was named The **Economist's "Country of the Year** for 2024, acknowledged for its great influence on democracy (The Economist, 2024).

.

Tweet: '"Very stimulating meeting with Hon'ble Adviser in charge of ICT, Post and Telecom, and Information & Broadcasting this afternoon #Connectivity #InternetForAll #StudentsPower #RiskPrevention #FreedomOfExpression #MediaFreedom"

Original Stance: Against Awami League

Response After Zero Shot

'Analyzing the tweet:\n\n### Tweet:\n"Very stimulating meeting with Hon\'ble Adviser in charge of ICT, Post and Telecom, and Information & Broadcasting this afternoon #Connectivity #InternetForAll #StudentsPower #RiskPrevention #FreedomOfExpression #MediaFreedom"\n\n---\n\n### Analysis:\n\n- The tweet is neutral in tone and focuses on a meeting with an adviser responsible for ICT and media-related topics.\n- The hashtags emphasize positive and progressive themes like "Connectivity," "FreedomOfExpression," and "MediaFreedom," which can align with initiatives typically promoted by governments to showcase achievements or policies.\n- There is no direct mention of Sheikh Hasina, Muhammad Yunus, or the Awami League by name.\n- The tone is optimistic, not critical, which may align with a pro-government stance but does not overtly support or oppose the Awami League or their policies.\n\n### Stance:\nThe tweet does not explicitly take a stance in favor of or against the Awami League. If additional context suggests the adviser mentioned is aligned with Awami League policies, it could lean in favor of Awami League.'

Response After Few Shot

'Let\'s analyze the tweet:\n\n**Tweet:** Very stimulating meeting with Hon\'ble Adviser in charge of ICT, Post and Telecom, and Information & Broadcasting this afternoon #Connectivity #InternetForAll #StudentsPower #RiskPrevention #FreedomOfExpression #MediaFreedom\n\n**Questions:**\n\n1. **Who is being praised or highlighted in the tweet?** The tweet mentions a "Hon\'ble Adviser" in charge of key ministries. This suggests a positive portrayal of this individual and their role.\n\n2. **What are the hashtags associated with the tweet?** The hashtags focus on positive themes like connectivity, internet access, student empowerment, risk prevention, freedom of expression, and media freedom. These are generally aligned with progressive and development-oriented agendas.\n\n3. **Considering the general political landscape of Bangladesh, who might benefit from these positive associations?** The Awami League government, under Sheikh Hasina, has often emphasized digital development, internet access, and freedom of expression as key priorities. \n\n**Answering the Questions:**\n\n* The tweet praises an individual in a position of power within the government.\n* The hashtags promote positive themes often associated with the Awami League\'s policies.\n* The positive portrayal aligns with the Awami League\'s focus on development and progressive values.\n\n\n**Stance:**\n\nBased on the analysis, the tweet appears to be **in favor of the Awami League**. \n'

Response after GRASP-ChoQ

Key considerations:
Who is being discussed?
The tweet refers to an "Hon'ble Adviser in charge of ICT, Post and Telecom, and Information & Broadcasting." Based on the provided information, this appears to align with a key member of the interim Yunus government.
Tone and context:
The tweet uses positive language ("stimulating meeting," "Hon'ble Adviser"), suggesting a favorable tone towards the individual or administration.
Hasina's allies vs. enemies:If the tweet praised Awami League or Hasina, it would likely highlight their policies, achievements, or victimization.
The mention of "StudentsPower" and "FreedomOfExpression" aligns with movements and values often associated with the student-led opposition or interim government under Yunus.
Key identifiers from general info:
The current government in the tweet is linked to Muhammad Yunus and his allies. By praising an adviser from this government, the tweet indirectly supports Yunus and opposes the Awami League.
Conclusion:

This tweet is **against Awami League** as it praises a key figure in the Yunus government, which is a known adversary of Sheikh Hasina and the Awami League.

Figure 14: Using zero-shot, few-shot, and GRASP-ChoQ methods, this chart displays stance categorization findings. Zero-shot marks the tweet as neutral, while few-shot indicates a pro-Awami League stance. GRASP-ChoQ coincides with the original label, marking it as opposed to the Awami League.

Tweet: 'Where is Sheikh Hasina's resignation letter? But has this government caught the President's mistake by the illegal legal advisor? In a parliamentary system, he reads what is written to the President! The President needs to know who wrote all this. If the President had broken his oath by saying that he did not receive the Prime Minister's resignation letter'
Original Stance: Favour of Awami League

Response After Zero Shot

"The tweet provided is **critical of Sheikh Hasina and the Awami League government**. It implies that the government has engaged in **manipulative or illegal actions** concerning the President's conduct. The focus on **resignation letters and accusations of oath-breaking** indicates **distrust and criticism of Sheikh Hasina's government**. There is no praise or victim narrative for Sheikh Hasina or the Awami League in this message.\n\nHence, the **political stance** of the tweet is:\n\n**Stance: Against Awami League**"

Response After Few Shot

Analyzing the given tweet and the guidelines:

The tweet asks about Sheikh Hasina's resignation letter, implying skepticism or criticism about her resignation process. It suggests the government is using an "illegal legal advisor" to manipulate decisions, which is accusatory. The criticism of the parliamentary system and the President's actions is indirectly questioning the integrity of the government associated with Sheikh Hasina. The tone and content criticize the Awami League's governance and Sheikh Hasina's role.

Stance: Against Awami League
The tweet aligns with an anti-Awami stance, as it criticizes Sheikh Hasina and her government's actions, painting them as manipulative or unlawful.

Response after GRASP-ChoQ

Questions to detect stance:

1. **Who is being criticized here?**
   - The tweet criticizes the President and the 'illegal legal advisor,' who are part of the interim government led by Muhammad Yunus. Since Muhammad Yunus is an enemy of Sheikh Hasina and the Awami League, criticism of his government suggests a pro-Awami League stance.
2. **Which government is in power in the tweet?**
   - The tweet refers to the interim government led by Muhammad Yunus, which came to power after Sheikh Hasina's resignation. This indicates that the tweet is discussing the political dynamics following Hasina's departure.
3. **What is the tone towards Sheikh Hasina and her allies?**
   - The tweet does not directly praise Sheikh Hasina but implies that the current government is mishandling the situation and is illegitimate. This indirectly supports the idea that Hasina's resignation was not handled properly, which could be seen as sympathetic to her situation.

Conclusion:
The tweet is critical of the interim government led by Muhammad Yunus and the President, which aligns with a pro-Awami League stance. Therefore, the tweet is **in favour of the Awami League**."

Figure 15: An instance of a misclassified stance in BPDisC dataset. Stance categorization results are compared here utilizing zero-shot, few-shot, and GRASP-ChoQ approaches. Although zero-shot and few-shot label the tweet as against the Awami League, GRASP-ChoQ rightly notes that it favors the Awami League, so matching with the initial label.