PropaInsight: Toward Deeper Understanding of Propaganda in Terms of Techniques, Appeals, and Intent

Jiateng Liu[1, *], Lin Ai[2, *], Zizhou Liu[2], Payam Karisani[1], Zheng Hui[2], May Fung[1], Preslav Nakov[3], Julia Hirschberg[2], Heng Ji[1]
[1]University of Illinois Urbana-Champaign [2]Columbia University
[3]Mohamed bin Zayed University of Artificial Intelligence
{jiateng5, hengji}@illinois.edu, {lin.ai, julia}@cs.columbia.edu

Abstract

Propaganda plays a critical role in shaping public opinion and fueling disinformation. While existing research primarily focuses on identifying propaganda techniques, it lacks the ability to capture the broader motives and the impacts of such content. To address these challenges, we introduce ***PropaInsight***, a conceptual framework grounded in foundational social science research, which systematically dissects propaganda into techniques, arousal appeals, and underlying intent. *PropaInsight* offers a more granular understanding of how propaganda operates across different contexts. Additionally, we present ***PropaGaze***, a novel dataset that combines human-annotated data with high-quality synthetic data generated through a meticulously designed pipeline.

Our experiments show that off-the-shelf LLMs struggle with propaganda analysis, but training with *PropaGaze* significantly improves performance. Fine-tuned Llama-7B-Chat achieves 203.4% higher text span IoU in technique identification and 66.2% higher BertScore in appeal analysis compared to 1-shot GPT-4-Turbo. Moreover, *PropaGaze* complements limited human-annotated data in data-sparse and cross-domain scenarios, showing its potential for comprehensive and generalizable propaganda analysis.[1]

1 Introduction

In an era of unbounded digital information, the deliberate dissemination of propaganda has proliferated, shaping public opinion and influencing political events (Stanley, 2015). Propaganda is also a key component of disinformation, where false information is intentionally crafted and distributed to deceive or mislead (Da San Martino et al., 2020). Detecting and analyzing propaganda is essential to maintain the integrity of public discourse and to ensure that individuals make informed, unbiased decisions (Da San Martino et al., 2020).

Most current research on propaganda detection focuses on identifying and categorizing the specific techniques used to persuade the audience (Da San Martino et al., 2019; Martino et al., 2020a). However, simply recognizing these techniques does not fully capture the motives behind the propaganda or its broader impact. As noted by Hobbs (2020); Lord and Vogt (2021), understanding the sources and the intentions behind the information is critical to effectively combat disinformation. Therefore, there is a growing need for moving beyond simply recognizing techniques to deepen understanding of propaganda.

Moreover, existing studies often rely on expert annotations because non-expert annotators struggle with separating personal biases from their assessments of propaganda (Da San Martino et al., 2019). This dependence on expert annotations leads to relatively small datasets, which may be insufficient for training large, generalizable models, and limits their applicability in cross-domain contexts where propaganda usage varies. For instance, strategies in military content can differ greatly from such in political content, highlighting the need for broader, more diverse datasets.

To address these challenges, we build on foundational social science research on propaganda (Nelson, 1997; Jowett and O'donnell, 2018; Ellul, 2021) and identify three key elements behind propaganda attempts: *propaganda techniques*, *arousal appeals*, and *underlying intent*. Consequently, we introduce a new conceptual framework, *PropaInsight*, that systematically analyzes these elements. Additionally, we leverage the strong context understanding ability of large language models (LLMs) to generate synthetic data, resulting in *PropaGaze*, a novel dataset for propaganda analysis. Our motivations are twofold: **(1)** to develop a comprehensive frame-

---

[*] These authors contribute to this work equally.

[1] *PropaGaze* and code are available at this GitHub repository: https://github.com/Lumos-Jiateng/PropaInsight.

Proceedings of the 31st International Conference on Computational Linguistics, pages 5607–5628
January 19–24, 2025. ©2025 Association for Computational Linguistics

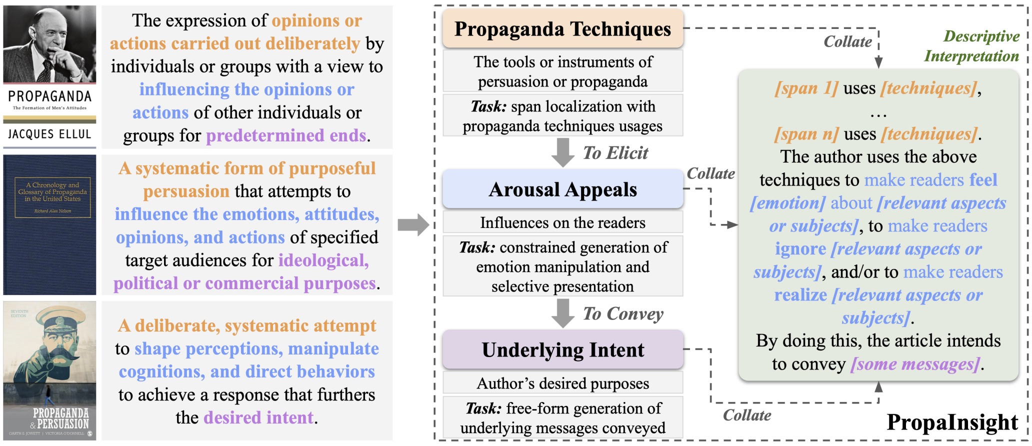

PROPAGANDA
The Formation of Men's Attitudes
JACQUES ELLUL
The expression of opinions or actions carried out deliberately by individuals or groups with a view to influencing the opinions or actions of other individuals or groups for predetermined ends.
A Chronology and Glossary of Propaganda in the United States
A systematic form of purposeful persuasion that attempts to influence the emotions, attitudes, opinions, and actions of specified target audiences for ideological, political or commercial purposes.
PROPAGANDA & PERSUASION
A deliberate, systematic attempt to shape perceptions, manipulate cognitions, and direct behaviors to achieve a response that furthers the desired intent.
Propaganda Techniques
The tools or instruments of persuasion or propaganda
Task: span localization with propaganda techniques usages
To Elicit
Arousal Appeals
Influences on the readers
Task: constrained generation of emotion manipulation and selective presentation
To Convey
Underlying Intent
Author's desired purposes
Task: free-form generation of underlying messages conveyed
Collate
Collate
Collate
Descriptive Interpretation
[span 1] uses [techniques],
…
[span n] uses [techniques].
The author uses the above techniques to make readers feel [emotion] about [relevant aspects or subjects], to make readers ignore [relevant aspects or subjects], and/or to make readers realize [relevant aspects or subjects].
By doing this, the article intends to convey [some messages].
PropaInsight

Figure 1: We abstract key elements of propaganda from social science literature. A comprehensive propaganda frame includes the techniques employed, the appeals evoked in readers, and the author's underlying intent.

work that goes beyond identifying techniques, and **(2)** to explore the use of synthetic data to supplement limited human-annotated data. Our contributions are as follows:

1. We propose *PropaInsight*, a conceptual framework for granular and comprehensive propaganda analysis that identifies propaganda techniques, arousal appeals, and underlying intent in news articles.
2. We introduce *PropaGaze*, a novel dataset for fine-grained propaganda analysis, consisting of a human-annotated news sub-dataset and two high-quality synthetic sub-datasets: one focused on the Russia-Ukraine conflict and one on the political domain.
3. We demonstrate that *PropaGaze* enhances LLMs' ability to analyze propaganda within the *PropaInsight* framework, paving the way for more nuanced and generalizable propaganda analysis methods.

2 *PropaInsight*: A Propaganda Analysis Framework

We introduce *PropaInsight*, a new conceptual framework for comprehensive propaganda analysis. In contrast to previous methods which ignore the underlying purposes and only focus on techniques, *PropaInsight* delves into the more subtle and hidden elements of propaganda. Drawing from foundational social science research on propaganda (Nelson, 1997; Jowett and O'donnell, 2018; Ellul, 2021), we identify three key elements of each propaganda attempt: *propaganda techniques*, *arousal appeals*, and *underlying intent*. As shown in Figure 1, for a given article, we first identify and classify the techniques used. We then infer the arousal appeals these techniques evoke, and we further deduce the underlying intent of the article. To ensure interpretability and consistency, we consolidate these elements into a clear, structured natural language paragraph using a descriptive template, as shown in Figure 1. Below, we provide a detailed explanation of each element of our proposed framework.

Propaganda Techniques Propaganda techniques are systematic, deliberate strategies used to craft persuasive content (Jowett and O'Donnell, 2012). Domain experts typically define these techniques as pre-defined labels like 'loaded language'. While the specific techniques may vary across different shared tasks (Torok, 2015), we follow the set of propaganda techniques defined in (Da San Martino et al., 2019), where each technique can be evaluated intrinsically. The full list of the 16 propaganda techniques we use is provided in Appendix C.

Arousal Appeals Appeals directly influence a reader's emotions, opinions, and actions after consuming propagandistic content (Nelson, 1997; Jowett and O'Donnell, 2012). A common propaganda device is to evoke strong emotions, such as hate or fear, in readers (Miller, 1937). Another approach involves selectively presenting evidence and facts to shape the audience's perception (Walton, 1997; O'Shaughnessy, 2004). To capture these effects, we design three templates (detailed in Appendix C) that identify the emotions evoked and the aspects readers are guided toward or distracted from while reading an article.

Underlying Intent Intent represents the ideological, political, or other underlying goal the author seeks to convey or achieve. To handle diverse real-world scenarios, we frame intent prediction as a free-text generation task, similar to approaches used for open intent generation in dialogue systems (Şimşek and Fensel, 2018; Wagner, 2022). The advantage of this novel formulation in propaganda intent analysis is its flexibility in capturing complex, nuanced intent that predefined labels cannot easily categorize, allowing greater freedom to generate more detailed and context-specific interpretations of intent.

Propaganda Analysis Task The design of *PropaInsight* introduces a new propaganda analysis task: generating a descriptive natural language paragraph explaining the techniques used, the appeals aroused, and the underlying intent. To avoid overlooking individual elements and to simplify evaluation, we divide the task into three sub-tasks:

1. ***Propaganda Technique Identification:*** Detect the spans where propaganda techniques are applied and which specific technique(s) correspond to each span, following prior task settings (Martino et al., 2020a,b).
2. ***Appeal Analysis:*** Generate the descriptions of emotions and feelings evoked using a template-based approach (see Appendix C).
3. ***Intent Analysis:*** Generate a free-form explanation of the article's underlying intent.

3 *PropaGaze*: A Dataset for Systematically Analyzing Propaganda

Existing propaganda datasets (Martino et al., 2020a; Heppell et al., 2023) primarily focus on identifying propaganda techniques and their associated text spans, but lack insights into appeal and intent. We introduce *PropaGaze*, a new dataset specifically designed for comprehensive propaganda analysis, consisting of three sub-datasets: *PTC-Gaze*, *RUWA-Gaze*, and *Politifact-Gaze*.

3.1 *PTC-Gaze*: Human-Annotated Dataset

PTC-Gaze builds on the existing PTC dataset (Martino et al., 2020a), which includes human-written news articles annotated for propaganda techniques and spans. We reannotate this dataset by hiring human annotators to label appeals and intent independently. For appeals, annotators review propaganda-containing sentences along with their context and describe the feelings evoked. To reduce cognitive load, we provide GPT-4 generated candidate annotations for assistance. Annotators then evaluate whether the generated candidates accurately reflect their interpretations and reactions, and if not, they rewrite the descriptions based on the template in § 2 and Appendix C. For intent, annotators read the full article and infer its underlying intent in a single free-form sentence, and we leave the multi-intent scenarios for future work.We used Label Studio to design the annotation interface, which is later shown in Appendix D. Two professional annotators from Kitware Inc. are in charge of the annotation task. We only have one annotator for each annotation task so no agreement rate is computed. As shown inTable 1, this annotated sub-dataset contains 79 articles, with an average of 12.77 propaganda techniques per article. Additional information, data examples, and analysis of the annotation quality are given in Appendix D.

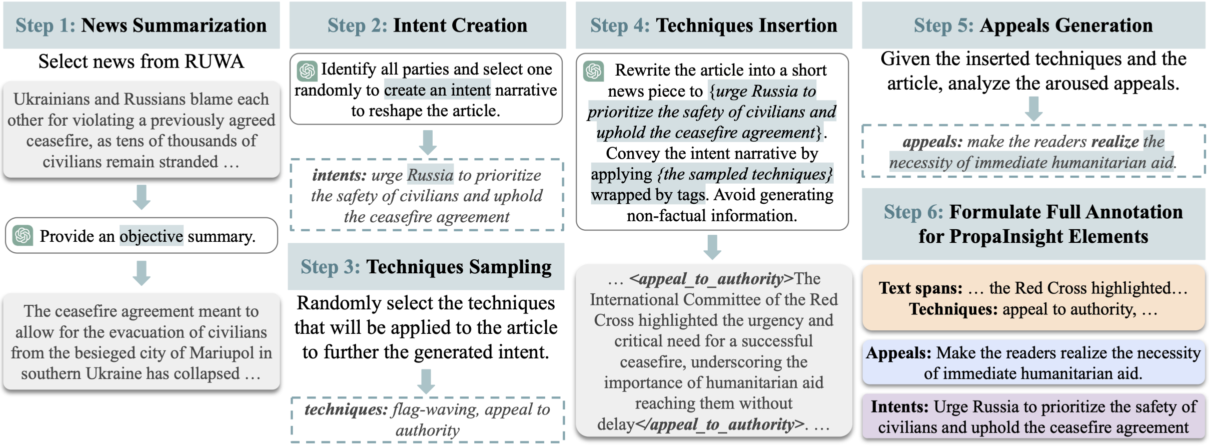
Sub-Dataset ↓	**#Articles**	**Avg. Article Length (words)**	**Avg. Propa Usage / Article**
PTC-Gaze	79	885.16	12.77
RUWA-Gaze	497	266.04	3.46
Politifact-Gaze	593	339.05	3.47

Table 1: Statistics about the *PropaGaze* dataset.

3.2 *RUWA-Gaze* and *Politifact-Gaze*: Synthetic Datasets

One limitation of the fully human-annotated dataset is that its usualy expensive, due to the challenging nature of the annotation tasks. This makes it insufficient for training large, generalizable models and limits its cross-domain applicability. Sparse data is a common issue in propaganda analysis research. To address this, we leverage LLMs such as LLaMA (Touvron et al., 2023) and GPT (Ye et al., 2023; OpenAI, 2023) to synthesize data, using their strong prior knowledge and context understanding. These models have shown effectiveness in data augmentation for tasks like propaganda techniques identification, such as fallacy recognition (Alhindi et al., 2024). These synthetic datasets are created mainly for training and can also serve as silver-standard benchmarks for propaganda analysis.

We construct *RUWA-Gaze* and *Politifact-Gaze* using a partially controlled data generation pipeline, as illustrated in Figure 2. Specifically, *RUWA-Gaze* is built upon a dataset of real-world news articles focused on the Russia-Ukraine War (Khairova et al., 2023), while *Politifact-Gaze* is constructed using the PolitiFact partition of the FakeNewsNet dataset (Shu et al., 2020).



Step 1: News Summarization

Select news from RUWA

Ukrainians and Russians blame each other for violating a previously agreed ceasefire, as tens of thousands of civilians remain stranded …

Provide an objective summary.

The ceasefire agreement meant to allow for the evacuation of civilians from the besieged city of Mariupol in southern Ukraine has collapsed …

Step 2: Intent Creation

Identify all parties and select one randomly to create an intent narrative to reshape the article.

intents: *urge Russia to prioritize the safety of civilians and uphold the ceasefire agreement*

Step 3: Techniques Sampling

Randomly select the techniques that will be applied to the article to further the generated intent.

techniques: *flag-waving, appeal to authority*

Step 4: Techniques Insertion

Rewrite the article into a short news piece to *{urge Russia to prioritize the safety of civilians and uphold the ceasefire agreement}*. Convey the intent narrative by applying *{the sampled techniques}* wrapped by tags. Avoid generating non-factual information.

… ***<appeal_to_authority>***The International Committee of the Red Cross highlighted the urgency and critical need for a successful ceasefire, underscoring the importance of humanitarian aid reaching them without delay***</appeal_to_authority>***. …

Step 5: Appeals Generation

Given the inserted techniques and the article, analyze the aroused appeals.

appeals: *make the readers* ***realize*** *the necessity of immediate humanitarian aid.*

Step 6: Formulate Full Annotation for PropaInsight Elements

Text spans: … the Red Cross highlighted…
Techniques: appeal to authority, …

Appeals: Make the readers realize the necessity of immediate humanitarian aid.

Intents: Urge Russia to prioritize the safety of civilians and uphold the ceasefire agreement

Figure 2: Partially controlled data generation pipeline: We first collect real-world news articles and derive an objective summary to extract events. Then we generate event-based intent, and randomly sample specific propaganda techniques to insert into the event descriptions. Lastly, we generate appeals from a reader's perspective, aiming at making the appeals grounded to the text.

Data Generation Pipeline Figure 2 shows the data creation pipeline. Initially, we use GPT-3.5 to summarize human-written, published news articles and to identify key events and objective facts. These summaries are intended to be objective, as the original articles may reflect various biases that could influence the creation of new propaganda pieces. Following this, we use GPT-3.5 to extract all focal entities involved in the events. We then randomly select one entity's perspective and set an intent to guide the revision of the article. We also randomly choose a set of propaganda techniques to be inserted into the article, reshaping its narrative.

Subsequently, we use GPT-4 as an intermediary author to craft intentional propaganda articles based on real-world events by injecting sampled propaganda techniques into an objective summary. We also ask the model to self-analyze the appeals the rewritten article might evoke to ensure alignment with the established intent. Human readers then verify the data quality for any obvious errors. The prompts for each step are provided in Appendix B.

As illustrated in Table 1, *RUWA-Gaze* consists of 497 articles, and *Politifact-Gaze* consists of 593 articles. While we generated moderate data due to the computational cost. We believe the data generation pipeline is generalizable. The language models can be replaced with cheaper or open-source LLMs to reduce costs and, in turn, generate larger-scale datasets. In addition, we identify that these two subsets come from different domains (Military & War and Politics), and they differ significantly in both content and the use of propaganda techniques.

4 Experiments

LLMs have strong prior knowledge and advanced context understanding, which makes them ideal for synthesizing propaganda-rich datasets and potentially effective for analyzing propaganda. In this section, we explore three research questions: **(1)** how off-the-shelf LLMs perform on propaganda analysis, **(2)** how much the *PropaGaze* dataset improves performance when used for training or fine-tuning, and **(3)** whether propaganda analysis is transferable across domains.

4.1 Experimental Setup

Sub-Tasks and Metrics As outlined in § 2, *PropaInsight* makes it possible to break the propaganda analysis task into three sub-tasks to ensure detailed evaluation and capture key elements:

1. ***Propaganda Techniques Identification:*** We use Intersection over Union (**IoU**) to measure the overlap between the identified text spans and the actual propaganda spans, and **F1** scores to evaluate propaganda technique classification, following prior task settings (Martino et al., 2020a,b).
2. ***Appeals Analysis:*** We evaluate the quality of the generated responses using **BertScore** (Zhang et al., 2019) to measure semantic similarity.
3. ***Intent Analysis:*** Similarly, we use **BertScore** for this sub-task.

Models We experiment with the following:

1. **GPT-4-Turbo**: One of the top-performing OpenAI models for various tasks. We use it in both zero-shot and few-shot prompting settings across all sub-tasks. The specific prompts used for each

sub-task are given in Appendix C.

2. **Llama-7B-Chat**: A popular open-source LLM. Due to its smaller size and relatively worse performance compared to GPT-4-Turbo, we fine-tune it for our sub-tasks. Specifically, we instruction-tune it to predict whether each sentence contains propaganda, and if so, identify the techniques and the appeals used, and predict the article's intent. See Appendix C for the fine-tuning prompts.
3. **Multi-Granularity Neural Networks (MGNN model)** (Da San Martino et al., 2019): A benchmark method for the propaganda techniques identification sub-task. We train MGNN from scratch for this specific task, as it is not designed for text generation and cannot be applied to the other two sub-tasks.

Data-Rich and Data-Sparse Training Settings In real-world scenarios, obtaining a large volume of well-annotated data for analyzing propaganda is challenging, as discussed in § 3. 6tt

For all *PropaGaze* sub-datasets, we split the articles into training and testing sets using a 70:30 ratio. *PTC-Gaze*, with only 79 articles, represents a data-sparse condition. In contrast, the synthetic sub-datasets, *RUWA-Gaze* and *Politifact-Gaze*, contain a total of over 1,000 articles. To simulate data-sparse scenarios with these two sub-datasets, we sample subsets matching the size of the full *PTC-Gaze* training set. For data-rich conditions, we use the full training sets of *RUWA-Gaze* and *Politifact-Gaze*, reserving one-seventh as the validation set.

4.2 How Do Off-the-Shelf LLMs Perform on Propaganda Analysis Tasks?

As shown in Tables 2 and 3, **zero-shot LLMs struggle with propaganda analysis**. For example, in identifying propaganda techniques, zero-shot GPT-4-Turbo underperforms compared to the trained MGNN, even in data-sparse conditions, despite MGNN being much smaller in size. Zero-shot LLMs often struggle to pinpoint sentences containing propaganda. Similarly, in appeal analysis, zero-shot GPT-4-Turbo achieves relatively low BertScores. However, these models perform better at inferring intent, as shown by their stronger performance in the intent analysis sub-task (Table 3).

Few-shot prompting improves LLM performance in analyzing propaganda elements. Specifically, in identifying propaganda techniques, one-shot GPT-4-Turbo shows an 80.8% improvement in average IoU on *RUWA-Gaze*, a 20.4% increase on *Politifact-Gaze*, and a 33.1% higher IoU on *PTC-Gaze* compared to a zero-shot setting. Similarly, in appeal analysis, one-shot GPT-4-Turbo achieves 14.9% higher BertScore on *RUWA-Gaze*, 15.8% higher on *Politifact-Gaze*, and 45.2% higher on *PTC-Gaze*. In intent analysis, zero-shot GPT-4-Turbo already performs well. The improvements compared to one-shot prompting are minor, with the highest increase being 3.5% on *RUWA-Gaze*.

4.3 How Much Does *PropaGaze* Enhance Model Performance?

***PropaGaze* substantially improves the overall propaganda analysis performance**, especially in identifying propaganda techniques, under both data-sparse and data-rich training conditions. In the data-sparse setting, fine-tuned LLaMA-7B-Chat outperforms one-shot GPT-4-Turbo, achieving an average of 65.8% higher text span IoU and 33.7% higher technique identification F1 score, as shown in Table 2. In the data-rich setting, the performance increases even further, with LLaMA-7B-Chat showing 90.9% higher text span IoU and 125.1% higher F1 score compared to the data-sparse results.Table 3 shows similar improvements in appeals and intent analysis. For the appeals sub-task, data-rich fine-tuning leads to an average 70.1% increase in BertScore, while for intent analysis there is a smaller 8.5% gain compared to data-sparse training. This is likely due to the already high baseline performance. These results demonstrate that the synthetic sub-datasets effectively complement the limited human-annotated data, significantly improving the model's performance in analyzing propaganda elements.

We also compare the performance of LLaMA-7B-Chat with the baseline benchmark MGNN on propaganda identification. In the data-sparse setting, fine-tuned LLaMA-7B-Chat substantially outperforms trained MGNN, achieving 158.43% higher IoU on *RUWA-Gaze* and 58.1% higher IoU on PolitiFact-Gaze. However, in data-rich scenarios, MGNN, benefiting from the larger amount of training data, surpassing LLaMA-7B-Chat. This may be due to the fact that smaller models, such as MGNN, can overfit when exposed to excessive training data, while larger LLMs, such as LLaMA-7B-Chat, generalize better in data-sparse conditions. These findings suggest that **LLMs are more suited for the task with limited training data, while smaller, dedicated models like MGNN could benefit more from the synthetic**

Data Setting ↓	Dataset → Model ↓	*RUWA-Gaze* Span Avg. IoU	*RUWA-Gaze* Techniques Macro F1	*Politifact-Gaze* Span Avg. IoU	*Politifact-Gaze* Techniques Macro F1	*PTC-Gaze* Span Avg. IoU	*PTC-Gaze* Techniques Macro F1
No Training Data	GPT-4-Turbo$_{0s}$	0.073	0.097	0.152	0.226	0.124	0.068
	GPT-4-Turbo$_{1s}$	0.132	0.145	0.183	0.269	0.165	0.171
Data-Sparse Training	MGNN	0.089	0.139	0.160	0.159	0.140	**0.206**
	Llama-7B-Chat$_{ft}$	0.230	0.210	0.253	0.281	**0.179**	0.191
Data-Rich Training	MGNN	**0.545**	0.591	**0.449**	**0.461**	-	-
	Llama-7B-Chat$_{ft}$	0.506	**0.607**	0.409	0.453	-	-

Table 2: Model performance on the propaganda technique identification sub-task under different training data settings. We report the performance of trained MGNN model and both k-shot (ks) and fine-tuned (ft) LLMs.

Dataset → Model ↓	*RUWA-Gaze* Appeals BertScore	*RUWA-Gaze* Intents BertScore	*Politifact-Gaze* Appeals BertScore	*Politifact-Gaze* Intents BertScore	*PTC-Gaze* Appeals BertScore	*PTC-Gaze* Intents BertScore
GPT-4-Turbo$_{0s}$	0.282	0.849	0.298	0.863	0.228	0.869
GPT-4-Turbo$_{1s}$	0.324	**0.879**	0.345	**0.875**	**0.331**	**0.881**
Llama-7B-Chat$_{ft}$ *(Data-Sparse)*	0.313	0.851	0.342	0.860	0.249	0.843
Llama-7B-Chat$_{ft}$ *(Data-Rich)*	**0.612**	0.861	**0.495**	0.864	-	-

Table 3: Model performance on appeal and intent analysis sub-tasks under different training data settings. We report the performance of zero-shot ($0s$) and fine-tuned (ft) LLMs.

sub-datasets provided by *PropaGaze* in data-rich environments. This is consistent with the findings of Alhindi et al. (2024). Thus, with sufficient training data, we can implement a pipeline that first localizes and identifies propaganda techniques using MGNN, followed by appeals and intent analysis based on MGNN's output. This approach could potentially enhance the overall quality of the model's output for the entire propaganda analysis task.

4.4 Is Propaganda Analysis Transferable Across Domains?

In the real world, propaganda spans various domains, including military and war, politics, economics, science, environmental issues, and more. Although the specific use of propaganda may differ across these domains, we are particularly interested in determining whether the general patterns of propaganda are transferable between domains. Additionally, high-quality human-annotated data is scarce, prompting us to investigate whether leveraging data from other domains can improve propaganda analysis in a target domain.

As outlined in § 3, our dataset consists of three subsets: *RUWA-Gaze* (military and war), *Politifact-Gaze* (politics), and *PTC-Gaze* (general news). To explore cross-domain transferability, we perform additional training on each target sub-dataset using data from the other two sub-datasets after the in-domain training. For instance, in a data-sparse scenario, if *RUWA-Gaze* is the target, cross-domain training on *Politifact-Gaze* involves first training the model on the sparse *RUWA-Gaze* data, followed by further training with sparse *Politifact-Gaze* data. In a data-rich scenario, the model is trained on the full in-domain *RUWA-Gaze* data, then further trained on the entire *Politifact-Gaze* dataset. The results are presented in Tables 4 and 5.

In data-sparse settings, we observe that models benefit substantially from incorporating cross-domain data. As shown in Table 4, when evaluated on *RUWA-Gaze*, models trained on additional data from *Politifact-Gaze* and *PTC-Gaze* achieve higher performance than those trained solely on sparse in-domain data. Specifically, LLaMA-7B-Chat fine-tuned with additional *Politifact-Gaze* data achieves the highest text span IoU of 0.271, while MGNN trained with additional *Politifact-Gaze* data reaches the highest technique F1 score of 0.281. This pattern is consistent across other sub-datasets and holds true for appeal analysis as well, as shown in Table 5. This is expected, as models trained in data-sparse conditions tend to benefit from cross-domain data due to the need for a larger pool of training examples. Access to additional data from related domains enables models to learn generalized patterns of propaganda usage more effectively, leading to improved performance even on tasks outside of their original training domain.

However, in data-rich scenarios, the benefit of

Additional Train Data ↓	Eval Dataset → Model ↓	*RUWA-Gaze* Span Avg. IoU	*RUWA-Gaze* Techniques Macro F1	*Politifact-Gaze* Span Avg. IoU	*Politifact-Gaze* Techniques Macro F1	*PTC-Gaze* Span Avg. IoU	*PTC-Gaze* Techniques Macro F1
RUWA-Gaze	MGNN$_{ft}$	0.089 | **0.545**	0.139 | 0.591	0.243 | **0.471**	0.251 | **0.475**	0.157 | 0.224	0.212 | 0.272
	Llama-7B-Chat$_{ft}$	0.230 | 0.506	0.210 | **0.607**	0.262 | 0.379	0.274 | 0.418	**0.215** | 0.243	**0.220** | 0.258
Politifact-Gaze	MGNN$_{ft}$	0.246 | 0.456	**0.281** | 0.593	0.160 | 0.449	0.159 | 0.461	0.203 | **0.252**	0.210 | **0.298**
	Llama-7B-Chat$_{ft}$	**0.271** | 0.443	0.265 | 0.582	0.253 | 0.409	**0.281** | 0.453	0.196 | 0.237	0.204 | 0.273
PTC-Gaze	MGNN$_{ft}$	0.189 | - - - -	0.226 | - - - -	0.224 | - - - -	0.237 | - - - -	0.140 | - - - -	0.206 | - - - -
	Llama-7B-Chat$_{ft}$	0.215 | - - - -	0.239 | - - - -	**0.265** | - - - -	0.261 | - - - -	0.179 | - - - -	0.191 | - - - -

Table 4: Model performance (**data-sparse** | **data-rich**) on the propaganda techniques identification sub-task under cross-domain training. The **best result** and runner-up result are highlighted per column for the data-sparse and data-rich settings, respectively. Diagonal cells show in-domain training only, without cross-domain training, and are included for reference.

Eval Dataset → Additional Train Data ↓	*RUWA-Gaze* Appeals BertScore	*RUWA-Gaze* Intents BertScore	*Politifact-Gaze* Appeals BertScore	*Politifact-Gaze* Intents BertScore	*PTC-Gaze* Appeals BertScore	*PTC-Gaze* Intents BertScore
RUWA-Gaze	0.313 | **0.612**	0.851 | **0.861**	**0.362** | 0.452	0.858 | **0.865**	**0.293** | **0.352**	0.839 | 0.841
Politifact-Gaze	**0.373** | 0.584	**0.855** | 0.860	0.342 | **0.495**	0.860 | 0.864	0.267| 0.310	**0.845** | **0.847**
PTC-Gaze	0.366 | - - - -	0.848 | - - - -	0.350 | - - - -	**0.863** | - - - -	0.249 | - - - -	0.843 | - - - -

Table 5: Fine-tuned Llama-7B-Chat model performance (**data-sparse** | **data-rich**) on the appeals and intent analysis sub-tasks under cross-domain training. The **best result** and runner-up result are highlighted per column for the data-sparse and data-rich settings, respectively. Diagonal cells show in-domain training only, without cross-domain training, and are included for reference.

cross-domain training diminishes. For example, as shown in Table 4, models trained on additional *Politifact-Gaze* data underperform those trained solely on in-domain data when evaluated on *RUWA-Gaze*. Similarly, when evaluated on *Politifact-Gaze*, adding *RUWA-Gaze* data sometimes leads to performance improvements, but the gains are relatively small. This holds for appeal analysis as well, as we can see in Table 5. These results suggest that **when there is sufficient training data, the quality of the data has a greater impact on performance than its quantity**. We further observe that training on both *RUWA-Gaze* and *Politifact-Gaze* improves the performance on the human-annotated *PTC-Gaze* across all sub-tasks. While this is partly due to the data-sparse nature of *PTC-Gaze*, making extra training samples valuable, it also highlights that our synthetic data effectively complements the limited human-annotated data.

5 Discussion

5.1 Discrepancy between Human-Annotated and Synthetic Datasets

We acknowledge the discrepancy between the synthetic sub-datasets and the human-annotated sub-dataset in *PropaGaze*. As shown in Table 1, the average number of propaganda techniques per article in *PTC-Gaze* is 12.77, which is about 3.7 times higher than in the synthetic *RUWA-Gaze* and *Politifact-Gaze*. This occurs due to the way we generate the synthetic data, where we inject three propaganda techniques per article, with GPT-4-Turbo sometimes reusing techniques. However, we believe this is less of an issue, as *PTC-Gaze* articles are on average 3.3 times longer than those in the other sub-datasets. Moreover, since we treat the injected techniques as silver labels, we have not yet checked whether other sentences in the articles also use propaganda techniques. See the Limitations section for more details. Finally, we note the inherent difference in writing styles between synthetic and human-written articles, which is a common challenge with synthetic datasets.

5.2 Further Challenges of Propaganda Analysis

We identified that accurately pinpointing the occurrence of propaganda is a major challenge in propaganda analysis. As highlighted in the case study (Appendix F), LLMs often misclassify non-propagandistic sentences as propagandistic, leading to a high false positives rate. This issue may be partially attributed to hallucination or failing to account for subtle contextual differences. Although less frequent, similar errors occur with MGNN, indicating that the problem lies not only in the models themselves, but also in the training methodologies and the underlying algorithms. This underscores the need for improvements in both model development and in the training approaches to better distinguish propagandistic content from neutral text.

6 Related Work

Propaganda Detection Propaganda detection has long been a focus in both Natural Language Processing, with most work focusing on identifying propaganda usage and specific techniques. Various learning-based approaches have improved performance (Da San Martino et al., 2019; Yoosuf and Yang, 2019; Li et al., 2019; Vorakitphan et al., 2022) and interpretability (Yu et al., 2021) in detecting propaganda in news articles (Vlad et al., 2019; Da San Martino et al., 2019; Gupta et al., 2019; Yu et al., 2021) and tweets (Vijayaraghavan and Vosoughi, 2022; Khanday et al., 2021; Guarino et al., 2020). Recent efforts have also applied LLMs to this task (Sprenkamp et al., 2023; Jones, 2024). While these studies focus on identifying propaganda techniques, further research is needed to understand the appeals and intent behind them.

Following the escalation of the Russo-Ukrainian conflict in 2022, research has focused on analyzing propaganda campaigns, particularly from Russia. Chen and Ferrara (2023); Fung and Ji (2022) collected user content and opinions from social media platforms such as X and Weibo, while Golovchenko (2022) examined censorship of Ukrainian content on Russian platforms. Geissler et al. (2023) studied pro-Russian sentiment on social media and the role of bots, and Patrona (2022) explored intertextuality and rhetoric in political performances during the war. However, few studies developed frameworks to analyze the specific intent behind propagandistic efforts. Ai et al. (2024) examined two specific propaganda narrative intentions, but failed short of proposing a generalizable framework for propaganda analysis.

Propaganda Generation Compared to propaganda detection, research on propaganda generation is sparse. Zellers et al. (2019) explores generating propaganda to spread targeted disinformation, while Huang et al. (2023) focuses on incorporating emotional and non-emotional propaganda techniques into generated articles. Goldstein et al. (2024) find that GPT-3 can generate highly persuasive propaganda. Our data generation pipeline goes further by allowing a broader range of propaganda techniques to be inserted into generated articles to evoke specific intent, while allowing for more granular analysis of the appeals behind their use.

User Intent Detection Previous methods on intent detection concentrated primarily on understanding user queries in human-machine dialogue systems (Brenes et al., 2009; Liu et al., 2019; Weld et al., 2022). This research facilitated the construction of more robust search engines and virtual assistants. The similarity of this task to ours is that both tasks require strong natural language understanding. However, detecting user query intent is relatively superficial compared to the intent behind a propaganda tactic, which could be highly concealed and hard to recognize (Jowett and O'Donnell, 2012).

7 Conclusion and Future Work

We proposed a comprehensive approach to propaganda analysis that goes beyond simply identifying techniques and addresses the common challenge of obtaining high-quality human-annotated data. We further introduced *PropaInsight*, a conceptual framework for granular propaganda analysis that identifies propaganda techniques, arousal appeals, and underlying intent, grounded in foundational social science research. Moreover, we presented *PropaGaze*, a novel dataset for fine-grained propaganda analysis that includes both human-annotated and high-quality synthetic sub-datasets. Our experiments showed that models fine-tuned on *PropaGaze* outperform one-shot GPT-4-Turbo by a margin. *PropaGaze* proved highly beneficial in data-sparse and cross-domain scenarios, serving as an effective complement to limited human-annotated data.

Furthermore, *PropaInsight* has broader implications beyond propaganda analysis. It enhances tasks such as disinformation detection (Guess and Lyons, 2020; Ai et al., 2021; Huang et al., 2023), sentiment analysis (Ahmad et al., 2019), narrative framing (Colley, 2019; Andersen and Sandberg, 2020), media bias analysis (Nakov and Da San Martino, 2021; Zollmann, 2019), and social media monitoring (Chaudhari and Pawar, 2021), offering deeper insights into manipulative content and coordinated disinformation campaigns, making the framework applicable to a wide range of areas.In the future, we plan to expand *PropaGaze* into more diverse domains and genres, which will further broaden the scope of propaganda analysis. We will also explore how *PropaInsight* can improve downstream applications and contribute to a deeper understanding of propaganda.

Limitations

We reflect on the limitations of our work below:

1. Although our dataset, *PropaGaze*, is novel and reliable, its size is relatively small due to the computational costs associated with GPT-4 and the high expense of human annotation. Consequently, we are uncertain about the dataset's ability to generalize across a broad range of domains when models are fine-tuned exclusively on it.
2. While we aimed to include diverse domains and construct a cross-domain dataset, the vast range of real-world scenarios exceeds what we could capture. The extent of the domain gap where propaganda thrives remains unclear, and therefore, the cross-domain performance we tested across our paper might not generalize well under varied conditions.
3. Despite our careful calibration of the proposed propaganda framework, the real-world responses such as reader engagement and ultimate impact can vary significantly. Personalized appeals may emerge, influencing the effectiveness of propaganda; however, our study did not account for these individual differences. We did not take this into consideration and leave this part for future work.
4. Although we use a partially controlled pipeline to generate synthetic data and have basic human reviewers skim the content to ensure its overall quality, a more fine-grained review is necessary. Specifically, we need to assess whether the sampled propaganda techniques are contextually appropriate within each article. Additionally, while we treat the injected techniques as silver labels for our experiments, we do not check whether other sentences in the article, beyond those explicitly marked, also employ propaganda techniques. This means that our synthetic sub-datasets have high precision in labeled techniques but have not been evaluated for recall. It is likely that in reshaping the articles, additional sentences may also use propaganda techniques not explicitly labeled. Further evaluation, potentially involving more comprehensive human annotation, is needed for a more granular assessment of the dataset's quality.
5. Our research is currently limited to English, which may restrict the generalizability of our findings to other languages. Future work will focus on extending the approach to cross-lingual settings to address this limitation.
6. We acknowledge that while our framework attempts to model intrinsic aspects of propaganda, the experimental setup simplifies the characterization of propaganda intentions. Specifically, the process of generating and validating intentions relies on annotator feedback rather than leveraging domain-specific intent modeling, which may limit the system's ability to fully capture nuanced propaganda strategies.

Based on this, we propose several promising future directions to further push for the success of combating misinformation.

1. A larger dataset developed using our propaganda framework could be constructed to further evaluate how synthetic data enhances the misinformation detection capabilities of large language models.
2. Collect data from various domains where significant domain gaps typically exist, and investigate whether cross-domain data substantially influence the accuracy of models' understanding and detection capabilities.
3. Consider personalized responses (Sun et al., 2023a,b, 2024) to propaganda. Conduct an in-depth analysis of how propaganda articles truly affect their readers and explore how these effects differ from the author's original perspective. In terms of solutions, a mixture of experts can probably lead to a better result. It is also worth considering solutions with LLM-based agents (Guo et al., 2024), which typically include multi-round of planning (Liu et al., 2023) and interactions (Wang et al., 2024b) or code execution (Wang et al., 2023; Yang et al., 2024; Wang et al., 2024a) to work out a reasoning based solution.
4. Enhancing and shaping the knowledge of LLMs may help to cultivate a better understanding of propaganda articles from different perspectives. It is promising to combine perspective-based datasets to state-of-the-art knowledge control approaches (Han et al., 2024; Liu et al., 2024; Deng et al., 2024) to alter the perspective and standpoints of LLMs.
5. It would also be valuable to explore incorporating this deeper understanding of propaganda techniques, appeals, and intent to enhance situa-

tion understanding and improve the comprehensiveness of situation report generation (Reddy et al., 2024).

6. We plan to incorporate intent mining approaches that include domain-specific characterizations of propaganda intentions. By integrating techniques such as hierarchical intent modeling, pretrained language models fine-tuned for intent detection, and emotion-informed analysis, we aim to strengthen our framework's ability to detect and classify propaganda intentions with greater specificity and robustness.

Ethical Considerations

Our paper introduces a pipeline capable of generating news articles with strong intent and the potential for propaganda use. While our primary goal is to leverage this synthetic data to combat misinformation, it is important to acknowledge that the same technology could be misused to produce high-quality deceptive news content that could mislead public opinion. Consequently, the use of this pipeline must be approached with caution and safeguarded to prevent exploitation by malicious actors.

Acknowledgement

This research was done with funding from the Defense Advanced Research Projects Agency (DARPA) under Contracts No. HR001120C0123 and HR0011-24-3-0325. The views, opinions, and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government. We also thank Kitware.Inc. and Rapidata.Inc. for their help in the data annotation process.

References

Siti Rohaidah Ahmad, Muhammad Zakwan Muhammad Rodzi, Nurlaila Syafira Shapiei, Nurhafizah Moziyana Mohd Yusop, and Suhaila Ismail. 2019. A review of feature selection and sentiment analysis technique in issues of propaganda. *International Journal of Advanced Computer Science and Applications*, 10(11).

Lin Ai, Run Chen, Ziwei Gong, Julia Guo, Shayan Hooshmand, Zixiaofan Yang, and Julia Hirschberg. 2021. *Exploring New Methods for Identifying False Information and the Intent Behind It on Social Media: COVID-19 Tweets*. ICWSM, virtual.

Lin Ai, Sameer Gupta, Shreya Oak, Zheng Hui, Zizhou Liu, and Julia Hirschberg. 2024. Tweetintent@ crisis: A dataset revealing narratives of both sides in the russia-ukraine crisis. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, pages 1872–1887.

Tariq Alhindi, Smaranda Muresan, and Preslav Nakov. 2024. Large language models are few-shot training example generators: A case study in fallacy recognition. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 12323–12334, Bangkok, Thailand and virtual meeting. Association for Computational Linguistics.

Jan Christoffer Andersen and Sveinung Sandberg. 2020. Islamic state propaganda: Between social movement framing and subcultural provocation. *Terrorism and Political Violence*, 32(7):1506–1526.

David J Brenes, Daniel Gayo-Avello, and Kilian Pérez-González. 2009. Survey and evaluation of query intent detection methods. In *Proceedings of the 2009 Workshop on Web Search Click Data*, pages 1–7.

Deptii Devendra Chaudhari and Ambika Vishal Pawar. 2021. Propaganda analysis in social media: a bibliometric review. *Information Discovery and Delivery*, 49(1):57–70.

Emily Chen and Emilio Ferrara. 2023. Tweets in time of conflict: A public dataset tracking the twitter discourse on the war between ukraine and russia. In *Proceedings of the International AAAI Conference on Web and social media*, volume 17, pages 1006–1013.

Thomas Colley. 2019. Strategic narratives and war propaganda. *The SAGE Handbook of propaganda*, pages 38–54.

Giovanni Da San Martino, Yu Seunghak, Alberto Barrón-Cedeno, Rostislav Petrov, Preslav Nakov, et al. 2019. Fine-grained analysis of propaganda in news article. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pages 5636–5646. Association for Computational Linguistics.

Giovanni Da San Martino, Shaden Shaar, Yf Zhang, Yu Sh, Alberto Barrón-Cedeno, Preslav Nakov, et al. 2020. Prta: A system to support the analysis of propaganda techniques in the news. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 287–293. ASSOC COMPUTATIONAL LINGUISTICS-ACL.

Jingcheng Deng, Zihao Wei, Liang Pang, Hanxing Ding, Huawei Shen, and Xueqi Cheng. 2024. Unke: Unstructured knowledge editing in large language models. *Preprint*, arXiv:2405.15349.

Shizhe Diao, Rui Pan, Hanze Dong, Ka Shun Shum, Jipeng Zhang, Wei Xiong, and Tong Zhang. 2023.

Lmflow: An extensible toolkit for finetuning and inference of large foundation models. *arXiv preprint arXiv:2306.12420*.

Jacques Ellul. 2021. *Propaganda: The formation of men's attitudes*. Vintage.

Yi R. Fung and Heng Ji. 2022. A weibo dataset for the 2022 russo-ukrainian crisis. *Preprint*, arXiv:2203.05967.

Dominique Geissler, Dominik Bär, Nicolas Pröllochs, and Stefan Feuerriegel. 2023. Russian propaganda on social media during the 2022 invasion of ukraine. *EPJ Data Science*, 12(1):35.

Josh A Goldstein, Jason Chao, Shelby Grossman, Alex Stamos, and Michael Tomz. 2024. How persuasive is ai-generated propaganda? *PNAS nexus*, 3(2):pgae034.

Yevgeniy Golovchenko. 2022. Fighting propaganda with censorship: A study of the ukrainian ban on russian social media. *The Journal of Politics*, 84(2):639–654.

Stefano Guarino, Noemi Trino, Alessandro Celestini, Alessandro Chessa, and Gianni Riotta. 2020. Characterizing networks of propaganda on twitter: a case study. *Applied Network Science*, 5(1):1–22.

Andrew M Guess and Benjamin A Lyons. 2020. Misinformation, disinformation, and online propaganda. *Social media and democracy: The state of the field, prospects for reform*, 10.

Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang, Shichao Pei, Nitesh V Chawla, Olaf Wiest, and Xiangliang Zhang. 2024. Large language model based multi-agents: A survey of progress and challenges. *arXiv preprint arXiv:2402.01680*.

Pankaj Gupta, Khushbu Saxena, Usama Yaseen, Thomas Runkler, and Hinrich Schütze. 2019. Neural architectures for fine-grained propaganda detection in news. *arXiv preprint arXiv:1909.06162*.

Chi Han, Jialiang Xu, Manling Li, Yi Fung, Chenkai Sun, Nan Jiang, Tarek Abdelzaher, and Heng Ji. 2024. Word embeddings are steers for language models. *Preprint*, arXiv:2305.12798.

Freddy Heppell, Kalina Bontcheva, and Carolina Scarton. 2023. Analysing state-backed propaganda websites: a new dataset and linguistic study. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 5729–5741, Singapore. Association for Computational Linguistics.

Renee Hobbs. 2020. *Mind over media: Propaganda education for a digital age*. WW Norton & Company.

Kung-Hsiang Huang, Kathleen McKeown, Preslav Nakov, Yejin Choi, and Heng Ji. 2023. Faking fake news for real fake news detection: Propaganda-loaded training data generation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14571–14589, Toronto, Canada. Association for Computational Linguistics.

Daniel Gordon Jones. 2024. Detecting propaganda in news articles using large language models. *Eng OA*, 2(1):01–12.

Garth S Jowett and Victoria O'donnell. 2018. *Propaganda & persuasion*. Sage publications.

Garth S Jowett and Victoria O'Donnell. 2012. What is propaganda, and how does it differ from persuasion. *Propaganda & persuasion*, pages 1–48.

Nina Khairova, Bogdan Ivasiuk, Fabrizio Lo Scudo, Carmela Comito, and Andrea Galassi. 2023. A first attempt to detect misinformation in russia-ukraine war news through text similarity. *Proceedings of the 4th Conference on Language, Data and Knowledge (LDK)*, pages 559–564.

Akib Mohi Ud Din Khanday, Qamar Rayees Khan, and Syed Tanzeel Rabani. 2021. Detecting textual propaganda using machine learning techniques. *Baghdad Science Journal*, 18(1):0199–0199.

Jinfen Li, Zhihao Ye, and Lu Xiao. 2019. Detection of propaganda using logistic regression. In *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, pages 119–124.

Jiao Liu, Yanling Li, and Min Lin. 2019. Review of intent detection methods in the human-machine dialogue system. In *Journal of physics: conference series*, volume 1267, page 012059. IOP Publishing.

Jiateng Liu, Sha Li, Zhenhailong Wang, Manling Li, and Heng Ji. 2023. A language-first approach for procedure planning. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 1941–1954, Toronto, Canada. Association for Computational Linguistics.

Jiateng Liu, Pengfei Yu, Yuji Zhang, Sha Li, Zixuan Zhang, and Heng Ji. 2024. Evedit: Event-based knowledge editing with deductive editing boundaries. *Preprint*, arXiv:2402.11324.

Kristin M Lord and Katya Vogt. 2021. Strengthen media literacy to win the fight against misinformation.

G Martino, Alberto Barrón-Cedeno, Henning Wachsmuth, Rostislav Petrov, and Preslav Nakov. 2020a. Semeval-2020 task 11: Detection of propaganda techniques in news articles. *arXiv preprint arXiv:2009.02696*.

Giovanni Da San Martino, Stefano Cresci, Alberto Barrón-Cedeño, Seunghak Yu, Roberto Di Pietro, and Preslav Nakov. 2020b. A survey on computational propaganda detection. *arXiv preprint arXiv:2007.08024*.

Clyde Raymond Miller. 1937. *How to detect propaganda*.

Preslav Nakov and Giovanni Da San Martino. 2021. Fake news, disinformation, propaganda, and media bias. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 4862–4865.

Richard Alan Nelson. 1997. A chronology and glossary of propaganda in the united states (book review). *Journalism and Mass Communication Quarterly*, 74(3):645.

OpenAI. 2023. Gpt-4 technical report. *Preprint*, arXiv:2303.08774.

Nicholas J O'Shaughnessy. 2004. *Politics and propaganda: Weapons of mass seduction*. Manchester University Press.

Marianna Patrona. 2022. Snapshots from an information war: Propaganda, intertextuality, and audience design in the russia–ukraine conflict. *Violence: An International Journal*, 3(2):253–280.

Revanth Gangi Reddy, Daniel Lee, Yi R. Fung, Khanh Duy Nguyen, Qi Zeng, Manling Li, Ziqi Wang, Clare Voss, and Heng Ji. 2024. Smartbook: Ai-assisted situation report generation for intelligence analysts. *Preprint*, arXiv:2303.14337.

Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2020. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big data*, 8(3):171–188.

Umutcan Şimşek and Dieter Fensel. 2018. Intent generation for goal-oriented dialogue systems based on schema. org annotations. *arXiv preprint arXiv:1807.01292*.

Kilian Sprenkamp, Daniel Gordon Jones, and Liudmila Zavolokina. 2023. Large language models for propaganda detection. *arXiv preprint arXiv:2310.06422*.

Jason Stanley. 2015. *How propaganda works*. Princeton University Press.

Chenkai Sun, Jinning Li, Hou Pong Chan, ChengXiang Zhai, and Heng Ji. 2023a. Measuring the effect of influential messages on varying personas. In *Proc. The 61st Annual Meeting of the Association for Computational Linguistics (ACL2023)*.

Chenkai Sun, Jinning Li, Yi Fung, Hou Pong Chan, Tarek Abdelzaher, ChengXiang Zhai, and Heng Ji. 2023b. Decoding the silent majority: Inducing belief augmented social graph with large language model for response forecasting. *Preprint*, arXiv:2310.13297.

Chenkai Sun, Ke Yang, Revanth Gangi Reddy, Yi R. Fung, Hou Pong Chan, ChengXiang Zhai, and Heng Ji. 2024. Persona-db: Efficient large language model personalization for response prediction with collaborative data refinement. In *arxiv*.

Robyn Torok. 2015. Symbiotic radicalisation strategies: Propaganda tools and neuro linguistic programming.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models. *Preprint*, arXiv:2302.13971.

Prashanth Vijayaraghavan and Soroush Vosoughi. 2022. Tweetspin: Fine-grained propaganda detection in social media using multi-view representations. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3433–3448.

George-Alexandru Vlad, Mircea-Adrian Tanase, Cristian Onose, and Dumitru-Clementin Cercel. 2019. Sentence-level propaganda detection in news articles with transfer learning and bert-bilstm-capsule model. In *Proceedings of the second workshop on natural language processing for internet freedom: censorship, disinformation, and propaganda*, pages 148–154.

Vorakit Vorakitphan, Elena Cabrio, and Serena Villata. 2022. Protect: A pipeline for propaganda detection and classification. In *CLiC-it 2021-Italian Conference on Computational Linguistics*, pages 352–358.

Julia N Wagner. 2022. *Open Intent Generation Through Unsupervised Semantic Clustering of Task-Oriented Dialog*. Ph.D. thesis, Massachusetts Institute of Technology.

Douglas Walton. 1997. What is propaganda, and what exactly is wrong with it. *Public Affairs Quarterly*, 11(4):383–413.

Xingyao Wang, Yangyi Chen, Lifan Yuan, Yizhe Zhang, Yunzhu Li, Hao Peng, and Heng Ji. 2024a. Executable code actions elicit better llm agents. In *Proc. The Forty-first International Conference on Machine Learning (ICML2024)*.

Xingyao Wang, Hao Peng, Reyhaneh Jabbarvand, and Heng Ji. 2023. Leti: Learning to generate from textual interactions. In *arxiv*.

Xingyao Wang, Zihan Wang, Jiateng Liu, Yangyi Chen, Lifan Yuan, Hao Peng, and Heng Ji. 2024b. Mint: Evaluating llms in multi-turn interaction with tools and language feedback. *Preprint*, arXiv:2309.10691.

Henry Weld, Xiaoqi Huang, Siqu Long, Josiah Poon, and Soyeon Caren Han. 2022. A survey of joint intent detection and slot filling models in natural language understanding. *ACM Computing Surveys*, 55(8):1–38.

Ke Yang, Jiateng Liu, John Wu, Chaoqi Yang, Yi R. Fung, Sha Li, Zixuan Huang, Xu Cao, Xingyao Wang, Yiquan Wang, Heng Ji, and Chengxiang Zhai.

2024. If llm is the wizard, then code is the wand: A survey on how code empowers large language models to serve as intelligent agents. *Preprint*, arXiv:2401.00812.

Junjie Ye, Xuanting Chen, Nuo Xu, Can Zu, Zekai Shao, Shichun Liu, Yuhan Cui, Zeyang Zhou, Chao Gong, Yang Shen, et al. 2023. A comprehensive capability analysis of gpt-3 and gpt-3.5 series models. *arXiv preprint arXiv:2303.10420*.

Shehel Yoosuf and Yin Yang. 2019. Fine-grained propaganda detection with fine-tuned bert. In *Proceedings of the second workshop on natural language processing for internet freedom: censorship, disinformation, and propaganda*, pages 87–91.

Seunghak Yu, Giovanni Da San Martino, Mitra Mohtarami, James Glass, and Preslav Nakov. 2021. Interpretable propaganda detection in news articles. *arXiv preprint arXiv:2108.12802*.

Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. 2019. Defending against neural fake news. *Advances in neural information processing systems*, 32.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.

Florian Zollmann. 2019. Bringing propaganda back into news media studies. *Critical Sociology*, 45(3):329–345.

A Details of a Propaganda Frame

We list the closed set of propaganda techniques that are used in the paper in Table 6. We also included the full template that we used to describe appeals and intent. Note that (1) The set of propaganda techniques included here can be freely extended with any other techniques. (2) We made the templates for Appeals and intent with a valid rationale, as detailed in Section§ 2. However, we are not claiming that this is the optimal template design among all other possible designs. We believe that prompt engineering and further human assessment are necessary for discovering the optimal template for this task. And we leave this part for future work.

Propaganda Technique Set	
Loaded Language	*Name Calling / labeling*
Repetition	*Obfuscation*
Doubt	*Straw man*
Flag-waving	*Causal oversimplification*
Slogans	*Black-and-white fallacy*
Appeal to authority	*thought-terminating cliche*
Whataboutism	*Reductio and Hitlerum*
Smears	*Glittering Generalities*
Appeals Raised by Propaganda Usage	
1. Make the readers feel [Emotion] about [Sth that is related]	
2. Make the readers realize that [Sth that is related]	
3. Make the readers ignore that [Sth that is related]	
Ultimate Intent Generation	
[Anything that is applicable for explaining the intent]	

Table 6: The complete formulation for each component during propaganda analysis. The parts marked by '[]' indicates the allowance for free generation.

B Data Generation Prompt Templates

Step1: News Summarization

> *System Message: You are a helpful assistant.*
>
> News: {news}
>
> Provide an objective summary of the news article, ensuring to present information in neutral terms. Avoid using language or phrases that show bias towards either party involved.

Step 2: Intent Creation

> *System Message: You are a helpful assistant.*
>
> Article: {article}
>
> Identify all parties mentioned in the article. Select one party randomly and create an intent narrative to potentially reshape the article.

Step 4: Techniques Insertion

> *System Message: You are a skilled journalist, proficient in composing short brief news pieces.*
>
> Article: {article}
>
> Rewrite the article into a short news piece to {intent}. Convey the intent narrative effectively by applying the following rhetorical tactics, once or more as needed. The revision must be concise, with a clear emphasis on using these tactics to communicate the intended message. Avoid generating non-factual information.
>
> 1. {appeal tactic 1}
> Example: {appeal tactic 1 example}
>
> 2. {appeal tactic 2}
> Example: {appeal tactic 2 example}
>
> 3. {appeal tactic 3}
> Example: {appeal tactic 3 example}

Step 4: Appeals Generation

> *System Message: You are a helpful assistant that identifies how the writer of a news article wants the readers of the article to feel after reading some sentences.*
>
> In this task, the input will be a news article, then some sentence in the article will be provided and you need to identify how the specific sentence raises appeals among the readers, the propaganda tactics used in these sentences will also be proivded as a hint. Also remeber that your response should be aware of the main goal of the whole article. For each sentence, you only need to output a sentence describing the feelings in one of the following two templates:
>
> Make the readers feel [Positive & Negative Emotions] about [Something that is related]
> Make the readers realize/Ignore [Something that is related]
>
> Here is an example indicating the input and output format:
>
> Input: News article: {article}

> Sentence: {first sentence}
> Tactic: {the tactic that is used in the sentence}
>
> Sentence: {second sentence}
> Tactic: {the tactic that is used in the sentence}
>
> ...
>
> Output:
>
> [1] Your response for the first sentence: Make the readers feel [Positive & Negative Emotions] about [Something that is related]
>
> [2] Your response for the second sentence: Make the readers realize/Ignore [Something that is related]
>
> ...
>
> Now let's begin!
>
> Now given the following news article:

C Templates and Prompts We Used for Propaganda Analysis

This section describes each generation and prompt used in this paper. While these prompts could be enhanced through prompt engineering and additional human evaluation, we use them here as proof of concept.

Template for Composing the Predicted Elements into a Descriptive Sentence We use the following template to compose the predicted elements into a descriptive sentence as the final output for the propaganda analysis task:

> This article uses propaganda to further its author's ultimate intent of *{The ultimate intent that is predicted by the model}*. Specifically, the author uses *{The first identified propaganda technique}* in the sentence: "*{The first sentence that is identified to use propaganda}*" to make the readers *{The first appeal that is raised among the readers}*. The author also uses *{The second identified propaganda technique}* in the sentence: "*{The second sentence that is identified to use propaganda}*" to make the readers *{The second appeal that is raised among the readers}...*

Prompt Template for the Language Models to Analyze Propaganda in a Zero-shot Manner We use the following prompt to encourage language models to correctly predict the elements with in a propaganda frame, this prompt also enables simple parsing to obtain the results.

> News article: {The news article that needs to be analyzed}
>
> Given the news article above, you should detect the major intent of the article. The intent is conveyed by using certain tactics and raise appeals in some text spans. You are also going to output all the text spans and the corresponding tactics and appeals.
>
> The tactics that maybe used are listed here: loaded language, flag waving, slogans, repetition, straw man, red herring, whataboutism, obfuscation, causal oversimplification, false dilemma, thought terminating cliche, appeal to authority, bandwagon, glittering generalities, name calling, doubt, smears, reducito ad hitlerum
>
> You should also formulate the generated appeals in the following format, choose one of the following template to fill in the appeals:
>
> Make the readers feel [Some Emotion] about [Something that is related] Make the readers realize about [Something that is related] Make the readers ignore [Something that is related]
>
> Your should firstly output the ultimate intent, then sequentially output all the text spans within the original article that contains tacic and appeals related to this intent and the corresponding tactics and appeals. You should only output one appeal for each text span. Here is an example:
>
> {Ultimate intent} The intent detected {Ultimate intent}
>
> {Text Span} Text Span 1 {Text Span} {Tactic} Tactic 1 {Tactic} {Appeal} Appeal 1 {Appeal}
>
> {Text Span} Text Span 2 {Text Span} {Tactic} Tactic 2 {Tactic} {Appeal} Appeal 1 {Appeal}

...

Now, output your answer with the given News article!

Template for Instruction Tuning with Llama2-Chat-7B on Tactics

User: Is the sentence below using propaganda techniques? Answer with [Yes] [Propaganda Technique] or [No] [None], candidate techniques are: Providing the candidate techniques Sentence: {The sentence that needs to be identified. }

Assistant: {The Templated Answers}

Template for Instruction Tuning with Llama2-Chat-7B on Appeals

What is the appeal that the author tries to arouse in the following sentence ? Answer with 'Make the readers ...', Sentence: {The sentence that needs to be identified. }

Assistant: {The Templated Answers}

Template for Instruction Tuning with Llama2-Chat-7B on intent What is the intent that the author tries to convey in the following article ? Answer with a paragraph of intent. Article: {The article that needs detection. }

Assistant: {The Answers}

D Details of the *PropaGaze* Dataset

As introduced in Section § 3, the *PropaGaze* dataset comprises three subsets: *RUWA-Gaze*, *Politifact-Gaze*, and *PTC-Gaze*. More details and data examples are provided in this section.

RUWA-Gaze The *RUWA-Gaze* dataset is constructed focusing on the Ukraine-Russia War. The original news dataset was from (Khairova et al., 2023). After human verification on the construction steps, we keep 497 articles, with each article having an average of 3.46 times propaganda usage. We provide an example piece of data from the constructed *RUWA-Gaze* below, the article has three times of propaganda usage:

Generated News Article: In light of recent intelligence reports from Ukraine detailing a covert Russian operation to recover classified remains from a sunken cruiser in Crimea, the international call for transparency has never been louder. The Moskva missile cruiser, a symbol of strength and resilience, met its fate in the Black Sea on April 14, igniting a flurry of claims and counterclaims between Ukraine and Russia regarding the circumstances of its sinking. Ukraine has steadfastly maintained that the cruiser was struck by two of its Neptune missiles, a claim seemingly bolstered by the Pentagon's confirmation, while Russia vehemently denies such an event. In this critical moment of geopolitical tension, the need for concrete evidence from Ukraine to substantiate its claims cannot be overstated. As we stand united in our pursuit of truth and justice, it becomes imperative for us to rally around the call for transparency, ensuring that every claim made is backed by irrefutable proof. This is not just about a sunken ship; it's about maintaining the delicate balance of peace and preventing any further escalation that could lead our brothers and sisters into an unwarranted conflict. Some may argue, why this focus on Ukraine's need to present evidence when there are other pressing issues at hand that demand our attention. However, this moment offers a unique opportunity to address underlying issues of trust and accountability in a world already beset by misinformation and conflict. By embracing a posture of openness, Ukraine can lead by example, joining the chorus of nations that have chosen the path of transparency and responsibility. Such a move would not only vindicate Ukraine's claims but also strengthen international confidence in its commitment to integrity and truth. As the situation develops, the world watches closely. The presentation of solid evidence will be a pivotal step in resolving the current standoff, soothing tensions, and charting a course towards resolution and understanding between nations.

The sentence that uses propaganda 1: As we stand united in our pursuit of truth and justice, it becomes imperative for us to rally around the call for transparency, ensuring that every claim made is backed by irrefutable proof.

Technique annotation 1: Flag-waving

Appeal annotation 1: Make the readers feel positive about the demonstration of unity and transparency.

The sentence that uses propaganda 2: Some may argue, why this focus on Ukraine's need to present evidence when there are other pressing issues at hand that demand our attention.

Technique annotation 2: Whataboutism

Appeal annotation 2: Make the readers ignore the distraction of other issues and focus on Ukraine's need to present evidence.

The sentence that uses propaganda 3: By embracing a posture of openness, Ukraine can lead by example, joining the chorus of nations that have chosen the path of transparency and responsibility.

Technique annotation 3: Bandwagon

Appeal annotation 3: Make the readers realize the opportunity Ukraine has to set a strong example of transparency and responsibility.

Intent annotation: This article urges Ukraine to provide concrete evidence to support their claim that the Moskva missile cruiser was hit by Ukrainian-made Neptune missiles, in order to maintain transparency and prevent further escalation of tensions with Russia.

Politifact-Gaze Similar to the *RUWA-Gaze* dataset, this subset is also constructed with our partially controlled generation pipeline. This dataset is based on the PolitiFact partition of the FakeNewsNet dataset (Shu et al., 2020) with a focus on political status across countries. We keep 593 generated articles with with each article having an average of 3.47 times propaganda usage. We provide an example piece of data from the constructed *Politifact-Gaze* below, the article has four times of propaganda usage:

Generated News Article: Alabama Attorney General John Simmons has taken a bold stand for truth and justice by filing charges against Mary Lynne Davies, the accuser of Roy Moore, who now faces allegations of falsification, a glaring first-degree misdemeanor. In a shocking twist, Davies, a Democrat, has been exposed as a fabricator of claims against Moore, specifically relating to an alleged incident that supposedly occurred when she was 14 and he was in his early 30s. Legal experts confirm the charges could result in a year of imprisonment and $10,000 in fines for Davies. This sordid tale of deceit came to a head after the yearbook inscription that Moore purportedly wrote in the 1970s was definitively debunked as a forgery. Davies, in a desperate bid for attention, concocted a story so heinous, it has now backfired, forcing her into hiding following her release on $500 bail. Despite the unraveling of her narrative, her attorney maintains, without evidence, that she is a victim, not a perpetrator of falsehood. The root of this entire scandal lies solely in the vindictive actions of an individual who sought to manipulate public opinion for political gain2̌014disregarding the serious harm inflicted on Roy Moore's reputation and life. Moore's attorney has extended an olive branch, stating that Moore harbors no ill will and remains committed to championing women's rights, a testament to his character and integrity. In a move that restores faith in the justice system, Attorney General Simmons is not only holding Davies accountable but is also investigating other dubious claims against Moore. This encompasses looking into malicious allegations made by employees of a mall, who labeled him with reprehensible titles without a shred of proof. Through the exposing of these falsehoods and the pursuit of accountability, Roy Moore stands vindicated. Legal and moral authorities alike have rallied to his defense, recognizing the travesty that nearly cost a steadfast advocate for American values and family principles his reputation and career. The pursuit of justice for Moore sends a strong message against the weaponization of false accusations in political warfare.

The sentence that uses propaganda 1: Legal experts confirm the charges could result in a year of imprisonment and $10,000 in fines for Davies.

Technique annotation 1: Appeal to authority

Appeal annotation 1: Make the readers feel anxious and sympathetic towards Davies about her potential legal consequences.

The sentence that uses propaganda 2: This sordid tale of deceit came to a head after the yearbook inscription that Moore purportedly wrote in the 1970s was definitively debunked as a forgery.

Technique annotation 2: Loaded Language

Appeal annotation 2: Make the readers feel disgusted about the forgery that Moore was accused of.

The sentence that uses propaganda 3: The root of this entire scandal lies solely in the vindictive actions of an individual who sought to manipulate public opinion for political gain2̌014disregarding the serious harm inflicted on Roy Moore's reputation and life.

Technique annotation 3: Causal Oversimplification

Appeal annotation 3: Make the readers feel anger and resentment about the sole individual who manipulated public opinion.

The sentence that uses propaganda 4: Legal and moral authorities alike have rallied to his defense, recognizing the travesty that nearly cost a steadfast advocate for American values and family principles his reputation and career

Technique annotation 4: Appeal to authority

Appeal annotation 4: Make the readers feel relief and satisfaction about the support Moore is receiving from legal and moral authorities.

Intent annotation: This article urges Ukraine to provide concrete evidence to support their claim that the Moskva missile cruiser was hit by Ukrainian-made Neptune missiles, in order to maintain transparency and prevent further escalation of tensions with Russia.

***PTC-Gaze*:** The *PTC-Gaze* subset is constructed based on the propaganda techniques corpus (Martino et al., 2020a), the propaganda articles within the PTC dataset are from real-world news articles and the usage of propaganda together with propaganda techniques annotation is done by human annotators. To simulate the propaganda frame usage in the real world, we further hire human annotators from kitware to annotate each propaganda usage with further information of appeals, and conclude the article with intent. To alleviate the annotation burden, we firstly let GPT-4 models to generate a synthetic annotation, and then let human annotators to just the generated parts of this synthetic annotation and further rewrite into their own annotations. We collected 79 long articles with each article has an average number of 12.77 times of propaganda usage. We notice that in the real-world propaganda corpus, the times of propaganda usage for each article can be far exceeding that of synthetic data. We attribute this difference to a domain gap existing between synthetic articles and real articles. We give an example of the annotated article from *PTC-Gaze* as below:

Real-World News Article: Ex-Sailor Pardoned By Trump Says He's SUING Obama And Comey A former Navy sailor, who is one of five people to receive a pardon from President Donald Trump, is planning to file a lawsuit against Obama administration officials. Kristian Saucier, who served a year in federal prison for taking photos of classified sections of the submarine on which he worked, says he was subject to unequal protection by the law. Saucier said that he realizes he had erred in taking the photos, which he said he wanted to show only to his family to show them where he worked. He has also lashed out at Obama officials, saying that his prosecution was politically motivated, prompted by sensitivity about classified information amid the scandal involving Clinton's emails. According to Fox News, Saucier argues that the same officials who sought out punishment to Saucier for his actions chose to be lenient with Hillary Clinton in her use of a private email server and mishandling of classified information. Saucier's lawyer, Ronald Daigle, told Fox News on Monday that the lawsuit, which he expects to file soon in Manhattan, will name the U.S. Department of Justice, former FBI Director James Comey and former President Barack Obama as defendants, among others. Saucier, who lives in Vermont, pleaded guilty in 2016 to taking photos inside the USS Alexandria while it was stationed in Groton, Connecticut, in 2009. He said he only wanted service mementos, but federal prosecutors argued he was a disgruntled sailor who had put national security at risk by taking photos showing the submarine's propulsion system and reactor compartment and then obstructed justice by destroying a laptop and camera. Fox News They interpreted the law in my case to say it was criminal, Saucier told Fox News, referring to prosecuting authorities in his case, but they didn't prosecute Hillary Clinton. Hillary is still walking free. Two guys on my ship did the same thing and weren't treated as criminals. We want them to correct the wrong. Daigle said that a notice about the pending lawsuit was sent to the Department of Justice and others included in it in December. There is usually a six-month period that must elapse before the lawsuit actually is actually filed. My case was usually something handled by military courts, he said. They used me as an example because of [the backlash over] Hillary Clinton, he continued, alleging his life was ruined for political reasons. With a pardon, there's no magic wand that gets waved and makes everything right, Saucier said, But I try to stay positive and look forward. Saucier has had cars repossessed and is in debt due to the loss of income after having a felony on his record. The government actively destroyed his life and made it all but impossible for his family to get back on track. But Hillary Clinton is running around free, to this day. And that is what Saucier is so burnt about, with good reason.

The sentence that uses propaganda 1: Fox News: They interpreted the law in my case to say it was criminal, Saucier told Fox News, referring to prosecuting authorities in his case, but they didn't prosecute Hillary Clinton.

Technique annotation 1: Whataboutism

Appeal annotation 1: Make the readers feel indignant about the contrasting legal treatments toward Saucier and Hillary Clinton.

The sentence that uses propaganda 2: Two guys on my ship did the same thing and weren't treated as criminals.

Technique annotation 2: Whataboutism

Appeal annotation 2: Make the readers feel unjust about the inequality in punitive measures for similar actions.

The sentence that uses propaganda 3: They used me as an example because of [the backlash over] Hillary Clinton, he continued, alleging his life was ruined for political reasons.

Technique annotation 3: Causal oversimplification

Appeal annotation 3: Make the readers feel sympathetic towards Saucier's inopportune life circumstances allegedly resulting from political motivations.

Intent Annotation: The news intends to inform the public about Kristian Saucier's plans to sue Obama administration officials.

Analysis for Annotation Quality We provide an analysis of the annotation quality of our *PTC-Gaze* dataset. We used Label Studio for design the annotation interface. We present the user interface design of the intent annotation and appeal annotation tasks in Figure 3 and Figure 4. Two professional annotators from Kitware.Inc is in charge of the annotation task. Annotators choose to utilize the candidate annotation generated by GPT-4 under 59.8% annotated intent data points and 75.1% annotated appeal data points. This demonstrates the high quality of GPT4-provided annotation in terms of appeals and intent, further enhanced our points in § 5.2.

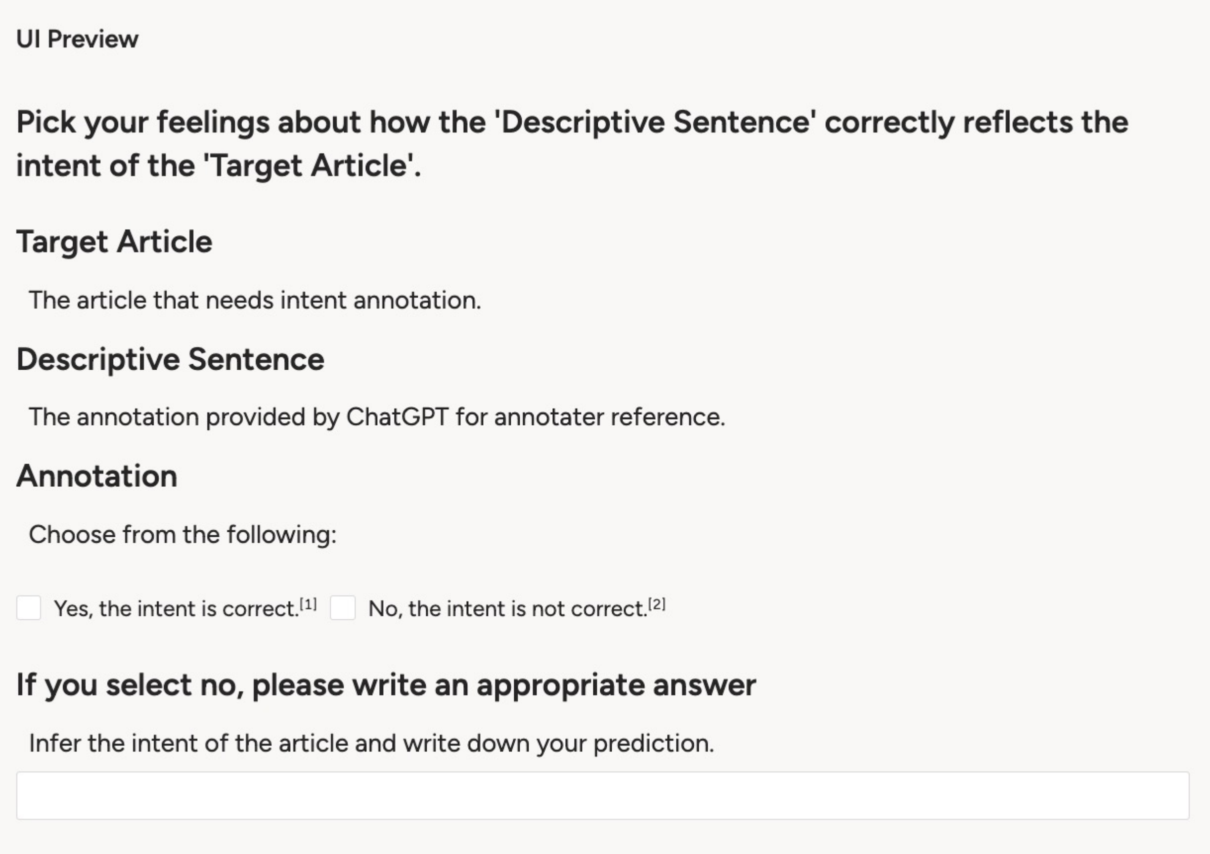
E Experimental Details

We provide experimental details for fine-tuning with Llama-Chat-7B and MGNN. For Llama-Chat-7B model, we used the LMFlow (Diao et al., 2023) framework for fine-tuning. We used four A100 GPUs for training, we set the learning rate to 0.00002 and batch size to 4. We tune the model for 3 epochs with our training data. During inference, we always set the inference temperature of the Llama-Chat-7B model to 1.0. For GPT-4-turbo, we used the default temperature for generation. In terms of tuning MGNN, we set batch size to 16 as MGNN takes a smaller memory space, and we set the learning rate to 0.00003. We then tune the model for 20 epochs.

F Case Study: Bottleneck of Propaganda Analysis

As discussed in § 5.2, we find that the bottleneck of propaganda analysis lies in identifying the correct propagandistic sentences. In this section, we give a case study on LLMs doing propaganda analysis to explain the cause further.

Input Example Data In a riveting instance of journalism that pierced through the veil of political spin, Fox News host Shepard Smith launched into a fervent condemnation of Donald Trump Jr.'s misleading explanations about his meeting with a Russian lawyer. During a segment that left audiences grappling with notions of truth and integrity, Smith vociferously questioned the incessant l̈iesänd d̈eceptionẗhat seem to shroud the Trump administration's dealings, implying the audience's own complacency in the face of such deceit unless they demand accountability. Smith's critique, grounded in an urgent plea for transparency, resonated strongly in an era where allegations of Russian collusion loom over the presidential election2̆014a matter of paramount significance under investigation by multiple congressional committees and a special counsel. Ẅhy all these lies? Why is it lie after lie after lie? If you're clean, come on clean,S̈mith implored on S̈hepard Smith Reporting,ëmphasizing the sheer implausibility of the evolving narrative woven by Trump Jr. and, by extension, the administration. In a moment that laid bare the discomforting truths surrounding this saga, Smith's acrimonious outburst was underscored by an on-air exchange with fellow anchor Chris Wallace, whose own speechlessness served as a testament to the gravity of Smith's statements. The interaction, a compelling dramatization of the inner turmoil gripping the nation, amplified the weight of Smith's words as he navigated the treacherous waters of political discourse. Yet, amid this pursuit of clarity and honesty, voices emerged calling for Smith's removal from the network2̆014a diversion that starkly contrasts the core issue at hand: the integrity of democratic institutions and the transparency of those in power. Smith's unapologetic defense of the press earlier in the year, where he rebuked claims against CNN as n̈ot fake news,f̈urther cements his role as a stalwart advocate for journalistic integrity in the face of political adversity. As the narrative of Donald Trump Jr.'s Russian rendezvous unfolds, Shepard Smith's impassioned critique serves as a poignant reminder of the media's critical role in dissecting



UI Preview

Pick your feelings about how the 'Descriptive Sentence' correctly reflects the intent of the 'Target Article'.

Target Article

The article that needs intent annotation.

Descriptive Sentence

The annotation provided by ChatGPT for annotater reference.

Annotation

Choose from the following:

☐ Yes, the intent is correct.[1] ☐ No, the intent is not correct.[2]

If you select no, please write an appropriate answer

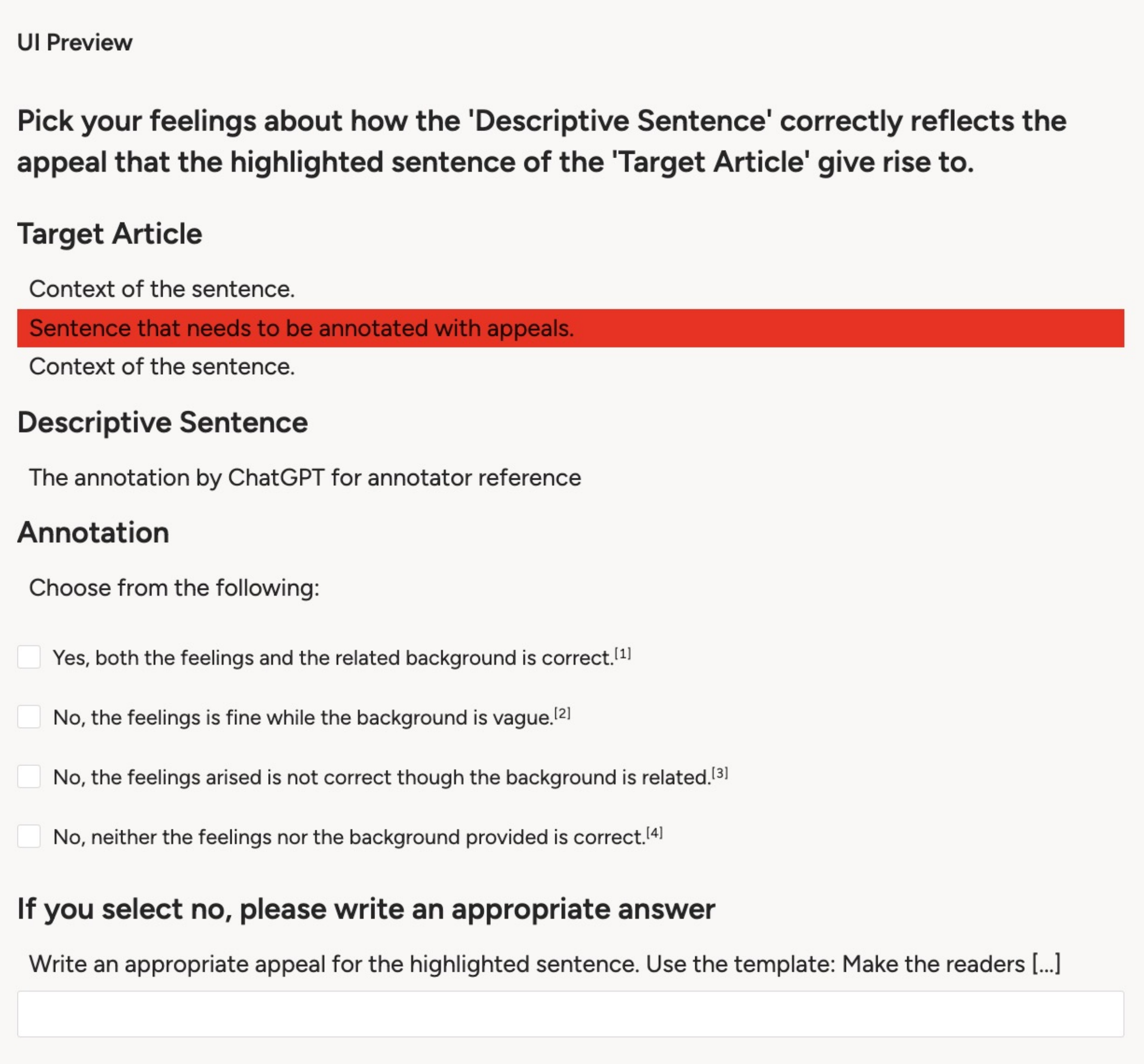
Infer the intent of the article and write down your prediction.

Figure 3: The user interface we used in Label Studio to annotate intent based on a given article.

the complex web of political narratives, urging the public to remain vigilant, question narratives, and demand nothing short of the truth.

Ground Truth Answers We list the ground truth propaganda identification and their related appeals and intent below:

> **Sentence:** During a segment that left audiences grappling with notions of truth and integrity, Smith vociferously questioned the incessant Ïiesänd d̈eceptionẗhat seem to shroud the Trump administration's dealings, implying the audience's own complacency in the face of such deceit unless they demand accountability.
>
> **Technique:** loaded language
>
> **Appeal:** Make the readers realize the serious implications of Smith's statements, as emphasized by Chris Wallace's speechlessness.
>
> **Sentence:** In a moment that laid bare the discomforting truths surrounding this saga, Smith's acrimonious outburst was underscored by an on-air exchange with fellow anchor Chris Wallace, whose own speechlessness served as a testament to the gravity of Smith's statements.
>
> **Technique:** appeal to authority
>
> **Appeal:** Make the readers realize the serious implications of Smith's statements, as emphasized by Chris Wallace's speechlessness.
>
> **Sentence:** Yet, amid this pursuit of clarity and honesty, voices emerged calling for Smith's removal from the network2̆014a diversion that starkly contrasts the core issue at hand: the integrity of democratic institutions and the transparency of those in power.
>
> **Technique:** red herring
>
> **Appeal:** Make the readers ignore calls for Smith's removal and focus instead on the integrity of democratic institutions and transparency of those in power.
>
> **Intent:** Highlight Fox News host Shepard Smith's criticism of Donald Trump



UI Preview

Pick your feelings about how the 'Descriptive Sentence' correctly reflects the appeal that the highlighted sentence of the 'Target Article' give rise to.

Target Article

Context of the sentence.

Sentence that needs to be annotated with appeals.

Context of the sentence.

Descriptive Sentence

The annotation by ChatGPT for annotator reference

Annotation

Choose from the following:

☐ Yes, both the feelings and the related background is correct.[1]

☐ No, the feelings is fine while the background is vague.[2]

☐ No, the feelings arised is not correct though the background is related.[3]

☐ No, neither the feelings nor the background provided is correct.[4]

If you select no, please write an appropriate answer

Write an appropriate appeal for the highlighted sentence. Use the template: Make the readers [...]

Figure 4: The user interface we used in Label Studio to annotate appeals based on a context. The highlighted part will be the sentence to be annotated, while other parts of ’Target Article’ provide related context.

Jr.’s Ïiesänd d̈eceptionr̈egarding his meeting with a Russian lawyer, questioning the ongoing narrative of dishonesty and secrecy within the Trump administration.

Predictions from LLMs

Sentence: In a riveting instance of journalism that pierced through the veil of political spin, Fox News host Shepard Smith launched into a fervent condemnation of Donald Trump Jr.’s misleading explanations about his meeting with a Russian lawyer.

Technique: loaded language

Appeal: Make the readers feel inspired by Shepard Smith’s courage to confront political deceit.

Sentence: Smith vociferously questioned the incessant Ïiesänd d̈eceptionẗhat seem to shroud the Trump administration’s dealings.

Technique: repetition

Appeal: Make the readers realize about the persistent dishonesty in the Trump administration.

Sentence: Implying the audience’s own complacency in the face of such deceit unless they demand accountability.

Technique: false dilemma

Appeal: Make the readers feel guilty for any lack of action against political deception.

Sentence: “Why all these lies? Why is it lie after lie after lie? If you’re clean, come on clean,”

Technique: repetition

Appeal: Make the readers feel outraged at the continuous lies by encouraging them to demand honesty.

Sentence: Smith’s critique, grounded in an urgent plea for transparency,

Technique: appeal to authority

Appeal: Make the readers trust in the necessity of transparency for integrity.

Sentence: the sheer implausibility of the evolving narrative woven by Trump Jr. and, by extension, the administration.

Technique: doubt

Appeal: Make the readers skeptical about any explanations from the Trump administration.

Sentence: voices emerged calling for Smith’s removal from the network2̌014a diversion that starkly contrasts the core issue at hand: the integrity of democratic institutions and the transparency of those in power.

Technique: red herring

Appeal: Make the readers ignore the calls for Smith’s removal and focus on the larger issues of integrity and transparency.

Sentence: " Smith’s unapologetic defense of the press earlier in the year, where he rebuked claims against CNN as n̈ot fake news,"̈

Technique: appeal to authority

Appeal: Make the readers feel supportive of Smith’s stance on defending press integrity.

Sentence: urging the public to remain vigilant, question narratives, and demand nothing short of the truth.

Technique: glittering generalities

Appeal: Make the readers realize about their role in ensuring political honesty and transparency.

Intent: The intent detected is to criticize the Trump administration’s lack of transparency and integrity, particularly focusing on Donald Trump Jr., while highlighting Shepard Smith’s role as a defender of truth and journalistic integrity.

Analysis for the Presented Result In the example above, LLMs are making reasonable predictions in terms of general intent, with Bertscore over 0.9, however, the model gives poor predictions on the sentences which are propagandistic. Though only three sentences are identified as propagandistic, the LLM predicts there is nine. Despite the techniques included covers the ground truth techniques, its low grounding rate causes these predictions to be useless.