Discovery, Articulation and Analysis of Communication Framings

Anonymous ACL submission

Abstract

In social media discourse, the framing of a topic dictates the way opinions are formed. This paper presents two multi-agent frameworks: one for discovering *what* topic aspects are framed, and another for identifying *how* they are framed. This enables an analysis of a topic's discourse that reveals what framings are dominant while also providing an interpretation of the framings. Experiments across four different topics show promising results for discovering, articulating and analyzing framings, providing a novel understanding of controversial topics. [1]

1 Introduction

In social media discourse, a variety of issues are discussed and their *framing* dictates the way these issues are understood and how opinions are formed. Therefore, it is important to discover *what* issues are framed in various ways as well as *how* they are framed. This will allow us to understand *why* people form certain opinions about a given topic.

The notion of *framing* originates from communication theories (Entman, 2003; Reese et al., 2001; Scheufele, 2004; Chong and Druckman, 2012; Bolsen et al., 2014), where framing is believe to enable the portrayal of an issue from one perspective to the necessary exclusion of alternative perspectives, as noted in (Boydstun and Gross, 2013). For example, when discussing the issue of the economic impact of immigration, if it is framed as immigrants contributing to the economy through their labor and consumption, it informs pro-immigration attitudes, excluding the perspective of stealing jobs from citizens, which in turn may lead to anti-immigration sentiment. Because nowadays a large potion of the population participates and forms opinions based on social media discourse, it is imperative to develop methods capable of capturing which issues are framed on social media, and especially how they are framed.

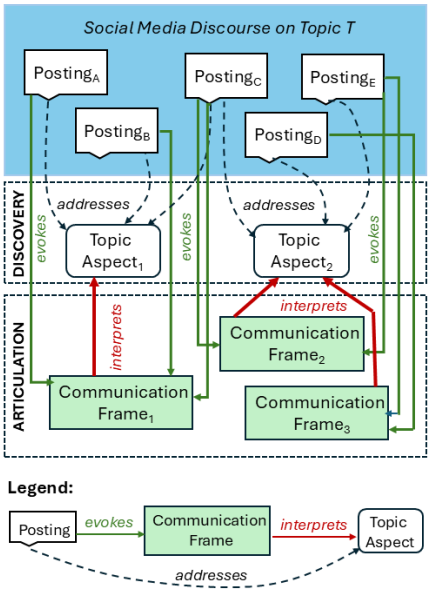

Figure 1: Social media postings addressing different topic aspects. The interpretation of the Topic Aspects (TAs) is provided by the articulation of the Communication Frames (CFs).

As defined by Entman (1993) *"to frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described."* Based on this definition, as illustrated in Figure 1, given the social media discourse on a topic T, revealing how the topic's issues are framed amounts to:
(a) the **discovery** of *what* Topic Aspects (TAs) are *addressed* in the various postings $\{P_A^T, P_B^T, P_C^T, P_D^T, P_E^T, ...\}$, as shown in Figure 1. We note that the TAs correspond to the issues selected for framing in discourse, according to the definition from Entman (1993). For example, for the topic

[1] Code:https://anonymous.4open.science/r/Discovery-Articulation-Analysis-CFs-2845/

T =*immigration*, we may have TA_1 = *Economic* and TA_2 = *Capacity and Resources*. Appendix A presents postings addressing TA_1.
(b) the **articulation** of Communication Frames (CFs) *evoked* by postings discussing topic T. The CFs have the role of providing causal interpretation to TAs. As shown in Figure 1, a CF may interpret more than one TA, and more than one CF may provide interpretations to a specific TA. For example, both TA_1 = *Economic* and TA_2 = *Capacity and Resources* are interpreted by CF_1 = *Immigrants largely spend their earnings locally rather than sending all their money abroad, so their wages are recycled into the local economy.* For each TA addressed by a posting P_i, the CF evoked by P_i provides its causal interpretation. The legend from Figure 1 summarizes the relations established between TAs, CFs and postings.
(c) **analyzing** the topic framings by finding which TAs are addressed the most. With millions of social media postings produced daily, it is important to have an understanding of what TAs are most discussed in each topic discourse and to have an immediate interpretations of these TAs through the CFs that are evoked.

The problem of automatic TA discovery is not new, nor is the problem of CF articulation. But until now, TA discovery was cast as TA recognition(Card et al., 2016; Naderi and Hirst, 2017; Kwak et al., 2020), as it was assumed that for each topic we are already aware of *all* TAs that are typically framed in various ways when discussing that topic. For example, Media Frames (Boydstun and Gross, 2013) proposed 14 TAs for topics involving policy issues. These TAs emerged from a combination of inductive and deductive strategies painfully conducted by social scientists, as reported in (de Vreese, 2005; Matthes and Kohring, 2008; Walter and Ophir, 2019).

When considering the discovery of TAs for the topic of immigration, (Mendelsohn et al., 2021) used not only the 14 TAs introduced by the Media Frames, but also 13 additional TAs which were specific to the topic of immigration, reported in studies focusing on immigration discourse (Mahtani, 2016; Hovden and Mjelde, 2019; Iyengar, 1991). This shows that even when we are aware of some of the TAs, we still should be able to *discover* additional TAs that help capturing all the ways a topic is framed. This observation allowed us to formulate the research question: **RQ1:** *How could we automatically discover all the TAs addressed in a set of postings, when we known only some TAs?*

To answer RQ1, we developed **ID-TA**, a multi-agentic LLM framework for Identifying and Discovering TAs addressed in postings discussing a certain topic. This multi-agent framework mimics the deductive and inductive reasoning performed by social scientists when trying to find and define the TAs that are framed for a topic. To do so, **ID-TA** is taking advantage of recent advances in reasoning with multi-agent LLMs through: (a) an *introspective* refinement interaction between two LLM agents (Sun et al., 2025); (b) *contrastive* reasoning (Wu et al., 2024; Chia et al., 2023a; Jacovi et al., 2021); and (c) multi-agent *debate* (Liang et al., 2024). In this way, **ID-TA** is not only identifying *known* TAs, but it also discovers new TAs and validates them.

Although the TAs new and known of a topic can indicate *which* topic framings are addressed in a discourse, they to not account for the interpretations of the framings, which indicate *why* certain TAs are addressed. This observation allowed us to formulate the research question: **RQ2:** *How could we automatically articulate the interpretations of the TAs discovered in a topic discourse?*

To answer RQ2, we developed **Art-FRAME**, a mutli-agentic LLM framework capable to (1) articulate FCs evoked by posting that addressed certain TAs, (2) filter out CFs that paraphrase already articulated CFs; (3) validate the CFs though an explainability test. Although **Art-FRAME** takes advantage of recently reported methods relying on LLM reasoning (Weinzierl and Harabagiu, 2025, 2024), it considers a new validation of CFs.

This paper makes three contributions: (1) it presents **ID-TA**, the first method capable to discover all topic aspects discussed in a discourse, known or new, across several topics; (2) it presents **Art-FRAME**, a method capable to articulate all valid CFs evoked in the discourse of several topics; and (3) it reports a novel analysis of controversial topics, focused on the way topics are framed.

The remainder of the paper is organized as follows. Section 2 is detailing the methodology that allows to discover framings and articulate them. Section 3 details the data used in experiments across discourse focused on five different topics. Section 5 describes the results of our experiments while Section 5 showcases framing analysis of a topic's discourse. Section 6 summarizes the conclusions.

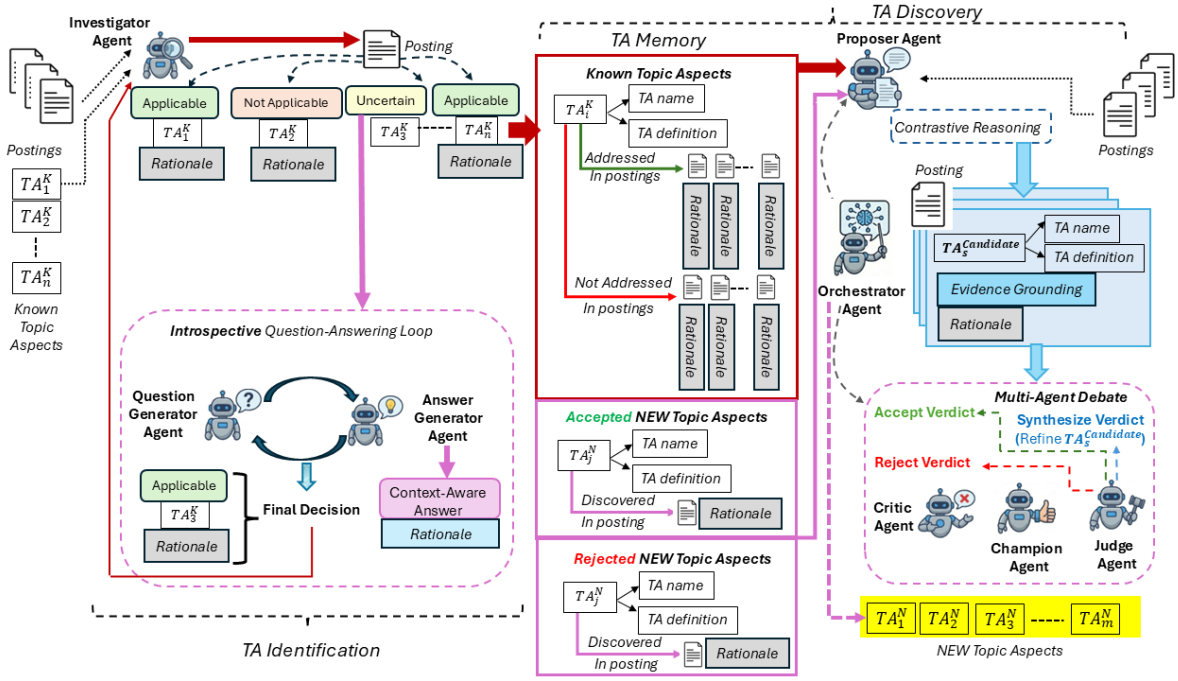

Figure 2: ID-TA: A Multi-Agent Module for Topic Aspect Discovery.

2 The Method

We propose a framework for processing framings in social media discourse consisting of two modules operating in sequential mode:
Module 1 is using the **ID-TA** multi-agent framework, described in Section 2.1
Module 2 is using the **Art-FRAME** multi-agent framework, described in Section 2.2

2.1 Module 1: ID-TA

As illustrated in Figure 2, before discovering new TAs from a set of postings discussing a topic T, **ID-TA** first addresses the problem of *identifying* known TAs in the postings. Appendix B.1 lists the known TAs considered for four different topics. **ID-TA** uses a *TA memory* that keeps track of which TAs are addressed in which postings. In addition, the TA memory informs the discovery of new TAs.
Identifying known TAs: An **Investigator Agent** is used to decide for each posting P_i which of the known TAs, $\{TA_1^K, TA_2^K, ..TA_n^K\}$ are addressed in P_i. The **Investigator Agent**, using the prompt detailed in Appendix C, can decide for any TA_x^K whether it is:
(a) *applicable*, because P_i is addressing TA_x^K; or
(b) *not applicable*, because P_i is not addressing TA_x^K; or
(c) *uncertain*, when it does not know if P_i is addressing TA_x^K.

When the **Investigator Agent** is uncertain about a TA_x^K, it triggers the *Introspective Question-Answering Loop* shown in Figure 2, which resolves ambiguity by simulating a dialogue between two specialized sub-agents: a **Question Generator Agent** and an **Answer Generator Agent**.

The **Question Generator Agent** formulates a specific *Yes / No* question designed to test the applicability of TA_x^K. The **Answer Generator Agent** carefully analyses the post P_i to answer the question. It then proceeds to provide a definitive *Yes* or *No* answer to the specific question, accompanied by a rationale explaining the decision. If the **Answer Generator Agent** cannot determine an answer for a question, it requests the **Question Generator Agent** to formulate a different or more granular question. This Q/A loop between the two agents continues until either a definitive answer is found of the threshold `max_loop` is reached.

The final decision of the **Investigator Agent** regarding the applicability of TA_x^K to P_i is resolved by inspecting the Question-Answer (Q/A) pairs. When the Investigator Agent finds sufficient evidence in the QA pairs that either the TA_x^K is applicable to P_i, or that TA_x^K is not applicable to P_i, it will change the uncertain decision to one of these decisions and it will also generate a *rationale* for the decision. If no evidence is found, it will decide that TA_x^K is not applicable to P_i. Detailed examples illustrating the introspective question-

answering loop are presented in Appendix D.

The TA Memory: Figure 2 shows how for each known TA_i^K, the memory keeps track of (a) the TA name (e.g. *Economy*); (b) its definition (e.g. *The costs, benefits, or monetary/financial implications*); (c) the list of the postings that addressed the TA; and eventually (d) the list of all postings that do not address the TA. The TA Memory is essential to the discovery of new TAs, and will be updated when new TAs are revealed.

Discovery of new TAs: is performed under the supervision of a central LLM **Orchestrator Agent**, which governs the life-cycle of new TA discovery through its three steps:

Step 1: For each posting P_j, the **Orchestrator Agent** invokes a **Proposer Agent** to generate a list of candidate new TAs $\{TA_i^{Candidate}\}$ by using *Contrastive Chain-of-Thought (CCoT)* reasoning (Chia et al., 2023b). Prior work has shown that such contrastive formulations improve reasoning quality and conceptual clarity (Jacovi et al., 2021; Chia et al., 2023b; Miller, 2019). Unlike simple Chain-of-Thought (CoT) reasoning (Wei et al., 2022), CCoT considers not only positive examples of demonstrations, but also negative examples. These examples are retrieved by the **Proposer Agent** from the TA Memory, where for each TA, examples of postings addressing it as well as examples of postings that do not address it are readily available. The CCoT prompt is detailed in Appendix C. For each $TA_i^{Candidate}$ generated by the CCoT prompt of the **Proposer Agent**, we also obtain: (a) the name of the $TA_i^{Candidate}$, e.g. *Victim Scapegoating*; (b) the definition of $TA_i^{Candidate}$, e.g. *Assigning unrelated social or policy failures to immigrants*; (c) evidence from P_j justifying that $TA_i^{Candidate}$ is being addressed; and (d) the rationale.

Step 2: The **Orchestrator Agent** checks if any of the $TA_i^{Candidate}$ are already known. For this reason, semantic similarity between the definitions of $TA_i^{Candidate}$ and any (known or new) $TA_j^{K/N}$ encoded in the TA memory is computed. We have considered this similarity between definitions $S_{i.j}^{Def}(TA_i^{Candidate}, TA_j^{K/N})$ to be measured by: (a) Sentence-BERT (SBERT) (Reimers and Gurevych, 2019) (b) BertScore (Zhang* et al., 2020); and (c) the Cross-Encoder introduced by (Nogueira and Cho, 2020). To decide which similarity measure to use, we have also relied on human comparisons of the pairs of definitions of TAs, judged with a three-point Likert scale: Duplicate (0), Overlap (0.5), or Distinct (1.0). When cross-validating the human judgments with the results of each considered similarity metric. we have found that SBERT produced the best scores. If $S_{i,j}^{Def} \geq S_K$, where S_K is a similarity threshold, then the **Orchestrator Agent** considers $TA_i^{Candidate}$ to be a duplicate of a TA already encoded in the TA Memory. Otherwise, the **Orchestrator Agent** proceeds to its third step.

Step 3: The **Orchestrator Agent** invokes a multi-agent debate framework where the *discriminant validity* and *applicability* of a TA_{cand} are debated. Three different agents, namely a **Champion Agent**, as **Critic Agent** and a **Judge Agent** participate in the debate. During the debate over TA_i^{cand}, associated with a posting P_j, the agents roles are the following:

◇ **The Champion Agent** defends TA_i^{cand}, arguing for (a) its *applicability* based on its grounding in posting P_j; and (b) the *discriminant validity*, based on TA_i^{cand} distinctness from all the known TAs addressed by P_j.

◇ **The Critic Agent** acts as a strict gatekeeper for the necessity of TA_i^{cand}. This agent operates on the principle of parsimony: new TAs should strictly not be created if existing ones suffice. It attempts to invalidate TA_i^{cand} by scrutinizing the **Champion Agent**'s arguments for (a) redundancy (when demonstrating TA_{cand} is similar to other TAs addressed in P_j); and (b) hallucination (demonstrating TA_{cand} relies on overreaching assumptions).

◇ **The Judge Agent** serves as the sole decision-maker and enforces a *dynamic termination policy*. Its verdict may take one of three forms: *Accept*, *Reject*, or *Correction*. When the arguments presented by any of the other two agents are decisive, the **Judge Agent** issues an early verdict to avoid unnecessary computation. Crucially, the **Judge Agent** also possesses *correction capabilities*: if the **Champion Agent** identifies a misalignment between the name or definitions of the TA encoded in the TA Memory and the candidate TA, the **Judge Agent** intervenes to refine the name or the definition of TA_i^{cand}. A detailed example of the multi-agent debate where corrections are performed is shown in Appendix E. The Appendix also shows the prompts used by the **Champion, Critic** and **Judge Agents**.

The advantage of using three different agents in this debate stems from the observation that in step 1, when the Proposer Agent utilizes contrastive reasoning to hypothesize a TA_i^{cand} for a posting

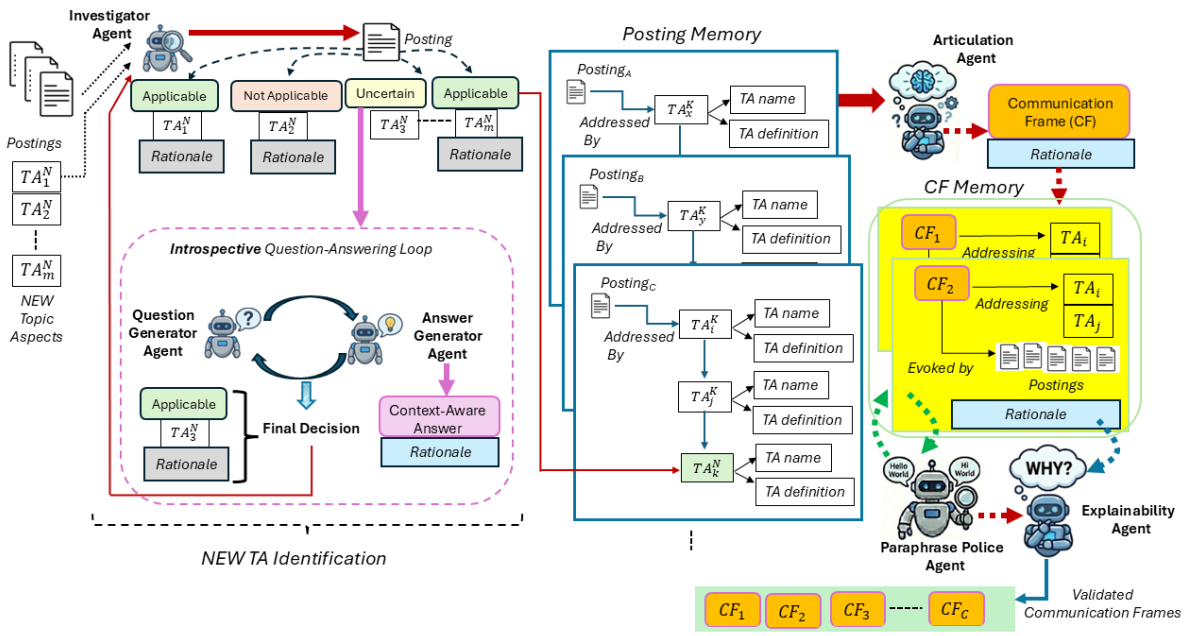

Figure 3: **Art-Frame:** A Multi-Agent Module for Articulating Communication Frames.

P_j, it uses a single generative model susceptible to Confirmation Bias (Turpin et al., 2023; Sharma et al., 2025). To mitigate this, we treat the Proposer Agent's output (P_{cand}) strictly as a Hypothesis, not a new TA. To validate each such hypothesis the Adversarial Multi-Agent Debate between three agents is employed. This form of debate is based on a paradigm rooted in Minsky's *Society of Mind* theory (Minsky, 1986), which posits that intelligence emerges from the interaction of conflicting subprocesses. The Multi-Agent Debate is rigorously testing the hypothesis to ensure it is not a hallucination or a subtle variant of a existing TAs (Turpin et al., 2023; Liang et al., 2024; Xiong et al., 2023).

When the **Judge Agent** accepts TA_i^{cand}, a new entry in the TA Memory is created to encode it as a new TA. As shown in Figure 2, along with the TA name and definition, the posting P_j where it was discovered is also stored, as well as the rationale for its discovery. Similarly, if TA_i^{cand} is rejected, a similar encoding is created, indicating that TA_i^{cand} is not a new TA. After all postings are processed, **ID-TA** outputs the list of newly discovered TAs, as illustrated in Figure 2, linked to their structured information available from the TA Memory.

2.2 Module 2: Art-FRAME

As illustrated in Figure 3, in order to articulate the CFs evoked in each posting P_i, we need to know which TAs, known or new, are addressed in P_i. While **ID-TA** identifies the known TAs in each posting, we need also in **Art-FRAME** to identify which new TAs are addressed in each posting, in the same way known TAs are identified in each posting by **ID-TA**. This allows **Art-FRAME** to assemble a *Posting Memory*, which keeps track of all TAs, known or new, addressed in each P_i. The Posting Memory informs an **Articulation Agent** that uses Chain-of-Thought (CoT) prompting (Wei et al., 2023) to generate, for each post P_i, a set of articulated Communication Frames $\{CF_i^k\}$. From the Posting Memory, the **Articulation Agent** is able to retrieve $ST_i = \{TA_i^j\}$, all TAs addressed by a posting P_i. ST_i along with P_i inform the prompting of the **Articulation Agent**. A $Rationale_i^j$ is generated for each articulated CF_i^k. All CFs generated in this way are evoked by P_i and provide a causal interpretation of the TAs from ST_i. Figure 3 shows how a *CF Memory* is created, to list all $\{TA_i^p\}$ (known or new) addressed by each articulated CF, along with all the postings that evoke it and the rationale for the CF.

The CF Memory, as illustrated in Figure 3, is used by two agents: A **Paraphrase-Police Agent**, whose role is to find paraphrasing CFs; and an **Explainability Agent**, whose role is to validate the articulated CFs. The **Paraphrase-Police Agent** identifies paraphrases between CFs that address the same TAs, using the same few-shot CoT prompting reported in (Ailneni and Harabagiu, 2025). When two CFs are paraphrases, one of them is randomly

selected and deleted from the CF Memory.

After the **Paraphrase-Police Agent** finishes to consolidate the CF Memory, the **Explainability Agent** starts operating. Drawing on prior research that CFs must be articulated clearly to be effective (Weinzierl and Harabagiu, 2024), we employ an *Articulation Reward Model* R_{art} as a quality filter. Implemented as a discriminative encoder, R_{art} is fine-tuned using binary a cross-entropy loss on a dataset comprising both valid and invalid CFs. During inference, the model predicts a scalar validity score $s \in [0, 1]$ for each CF. A strict acceptance threshold τ is used, ensuring that only CFs with high articulation quality are considered.

In addition, we assume that valid CF_i generated by the **Proposer Agent** should pass a *reconstruction test* performed by the **Explainability Agent**. Given rationale R_i provided for each CF_i in the CF Memory, as well as the TAs for which CF_i provides an interpretation, the **Explainability Agent** is prompted to re-generate the CF_i^R. The CoT prompt is detailed in Appendix C. If CF_i and CF_i^R are found as being paraphrases by the **Paraphrase-Police Agent**, then the reconstruction test is passed. When the reconstruction test is not passed, CF_i is not considered valid. As shown in Figure 3, **Art-FRAME** outputs a list of valid CFs evoked by all postings.

3 The Data

We use the data that was released publicly in (Weinzierl and Harabagiu, 2025), which covers four different discourse topics:

1. Dataset for Immigration (IMM): It consists of 2,730 posts from Twitter / X annotated with 14 issue-generic TAs introduced by the Policy Frames Codebook (Boydstun et al., 2018; Card et al., 2015), 11 issue-specific TAs (Benson, 2013; Hovden and Mjelde, 2019), and 2 narrative-focused TAs (Iyengar, 1990) surrounding the topic of immigration. In this study, we focus only on the 14 issue-generic TAs.

2. Dataset for COVID-19 Vaccine Hesitancy (VH): consists of the test split of the Twitter / X postings released by COVAXFRAMES (Wei). This dataset contains 2,815 postings annotated with 7 TAs and 113 CFs. These TAs correspond to seven dimensions of vaccine hesitancy that influence an individual's likelihood of vaccination (Geiger et al., 2022).

3.Dataset for Climate Change (CC) & 4. Dataset for Abortion (AB): consist of 933 Twitter / X postings about Abortion (AB) and 564 postings about Climate Change (CC), originating from the SemEval-2016 Task 6 dataset (Mohammad et al., 2016). using the release from TweetEval (Barbieri et al., 2020). The known TAs and their definitions for both topics are provided in Appendix B.1.

4 Experimental Results

Implementation Details We used GPT-5-mini both for **ID-TA** and **Art-FRAME** as the backbone LLM for all agents. In accordance with the GPT-5 model family, all agents operate with a fixed temperature of 1.0. For **ID-TA**, the introspective question-answering loop is limited to (`max_loop` = 2). The multi-agent debate terminates either by the decision of the **Judge Agent** or after (`max_run` = 3). In **Art-FRAME**, τ is set to 0.8, ensuring that only CFs with high articulation quality are validated.

Quantitative Results: The evaluation results for **ID-TA** allowed us to quantify (1) the identification of *known* TAs as well as (2) the discovery of *new* TAs across all four topics we considered. To evaluate the identified known TAs, we used the annotations from the **IMM** dataset, the only dataset with annotated known TAs and measured the Precision $P = 0.76$, Recall $R = 0.86$ and $F_1 = 0.80$ obtained by **ID-TA**. We note that on the **IMM** dataset, (Mendelsohn et al., 2021) reported the best F1-score of 0.657 for known TA identification, which was outperformed by **ID-TA**.

ID-TA also discovers new TAs. For the evaluation of the discovery of new TAs, two baselines were considered:

◇ *(Baseline 1): TA Discovery through postings clustering (TAD_{HAC}):* Assuming that all postings that address the same new TA belong to the same cluster, we have used Hierarchical Agglomerative Clustering (HAC) (Ward, 1963) with the single-link strategy, measuring the similarity between postings with the Sentence-BERT model (Reimers and Gurevych, 2019). The name and the definition of a new TA, corresponding to one of the obtained clusters, was generated by an LLM (GPT-5-mini) with the prompt detailed in Appendix C.

◇ *(Baseline 2) Zero-Shot Prompting (ZS-Prompt)* for TA Discovery: A Chain-of-Thought (CoT) prompting, in a zero-shot manner, of an LLM (GPT-5-mini) was used, conditioned on known TAs. The number of new TAs produced by each baseline are

compared with the number of TAs discovered by **ID-TA** and listed in Table 1.

Method	IMM	CC	AB	VH
TAD_{HAC}	21	18	19	15
ZS-Prompt	34	27	29	22
ID-TA	**13**	**10**	**11**	**9**

Table 1: Number of new TAs discovered across four topics: Immigration (IMM), Climate Change (CC), Abortion (AB), and Vaccine Hesitancy (VH).

The evaluation of the multi-agent debate taking place in **ID-TA** measures the proportion of candidate TAs that are accepted, rejected, or corrected, illustrated in Table 2. Across all topics, a substantial proportion of candidate TAs proposed by the **Proposer Agent** are rejected by the Judge Agent, indicating that CCoT prompting alone is not sufficient for proposing new TAs. Among accepted TAs, a non-trivial fraction are corrected by the **Judge Agent**, indicating that adversarial debate often warrants refinement of the name or definition of new TAs.

Topic	Accepted (%)	Corrected (%)	Rejected (%)
IMM	43	9	48
CC	40	10	50
AB	42	8	50
VH	39	7	54

Table 2: Distribution of decisions of the **Judge Agent** during the multi-agent debate from ID-TA.

We also quantify the ability of a new TA to elicit CFs we compute the number of valid CFs articulated by **Art-FRAME** for each new TA, those discovered by **ID-TA** as well as those produced by the two baselines. Table 3 reports the average number of articulated CFs elicited per new TA for each topic. While baseline methods often discover a larger number of TAs, these TAs tend to support fewer causal interpretations. In contrast, TAs discovered by **ID-TA** consistently yield a higher number of validated CFs per TA, indicating that **ID-TA** discovers new TAs that are more interpretable.

Method	IMM	CC	AB	VH
TAD_{HAC}	5.6	4.0	2.7	6.8
ZS-Prompt	4.2	3.1	2.0	5.0
ID-TA	**9.0**	**6.6**	**3.8**	**11.4**

Table 3: Average number of Communication Frames (CFs) articulated per new TA across topics.

Criterion 1: **Definition Clarity (DC)** of new TAs: Human judges decide whether the TA definition is fluent and easy to understand and rate on a 5-point Likert scale, ranging from 1 (incomprehensible) to 5 (perfectly clear).
Criterion 2: **Definition-Name Alignment (DNA)**: Judges are asked to judge the extent to which the definition explains, justifies, and faithfully reflects the meaning implied by the TA name. Judgments use a 5-point Likert scale, with a score of 1 indicates that the definition and name are semantically mismatched; while a score of 5 indicates that the definition offers a comprehensive explanation of the TA name.
Criterion 3: **Posting Coverage (PC)**: Judges evaluate whether postings addressing a certain TA genuinely discuss it. Judges are shown the TA name and definition, along with five postings randomly sampled from the list of postings addressing that TA. For each posting, judges answer *Yes* if the posting clearly addresses the TA and *No*, otherwise. The PC score for a TA is computed as the proportion of postings receiving a *Yes* judgment over all postings addressing the same TA.
Criterion 4 **Semantic Distinctiveness (SD)**: measures how clearly TAs differ in meaning. Sentence Bert (Reimers and Gurevych, 2019) is used to measure the similarity between the definitions of pairs of TAs. After selecting the most similar pairs of TAs, judges are asked to consider how distinctive the TAs are, based on their definitions, using a three-point scale: *Duplicate* (0), *Overlap* (0.5), or *Distinct* (1.0).

Table 4: Evaluation Criteria for Discovered New TAs.

Model	DC	DNA	PC	SD
ID-TA	**4.30**	**4.29**	**0.94**	**0.88**
TAD_{HAC}	3.10	3.35	0.48	0.65
ZS-Prompt	3.78	4.05	0.54	0.59

Table 5: Qualitative results for the discovery of New TAs for the topic of Immigration (IMM)

Qualitative Results are produced for (1) the discovered TAs; and (2) the articulated CFs.

Quality of the discovered TAs are revealed by human evaluation using four criteria, listed in Table 4, based on several Likert scales (Robinson, 2014). Table 5 reports the qualitative evaluation results for the IMM topic, whereas Appendix G.1 lists the results for all four topics. The results indicate that the TAs discovered by **ID-TA** produce evaluations which consistently outperform the results of the baselines across of the measures of all four criteria. The largest gains are observed in Post Addressing (PA) and Semantic Distinctiveness (SD), indicating that TAs discovered by **ID-TA** are not only well grounded in the underlying posts but also semantically non-redundant. Improvements in Definition Clarity (DC) and Definition–Name Alignment (DNA) further suggest that **ID-TA** produces interpretable and well-formed TA definitions.

Quality of the articulated CFs are revealed by

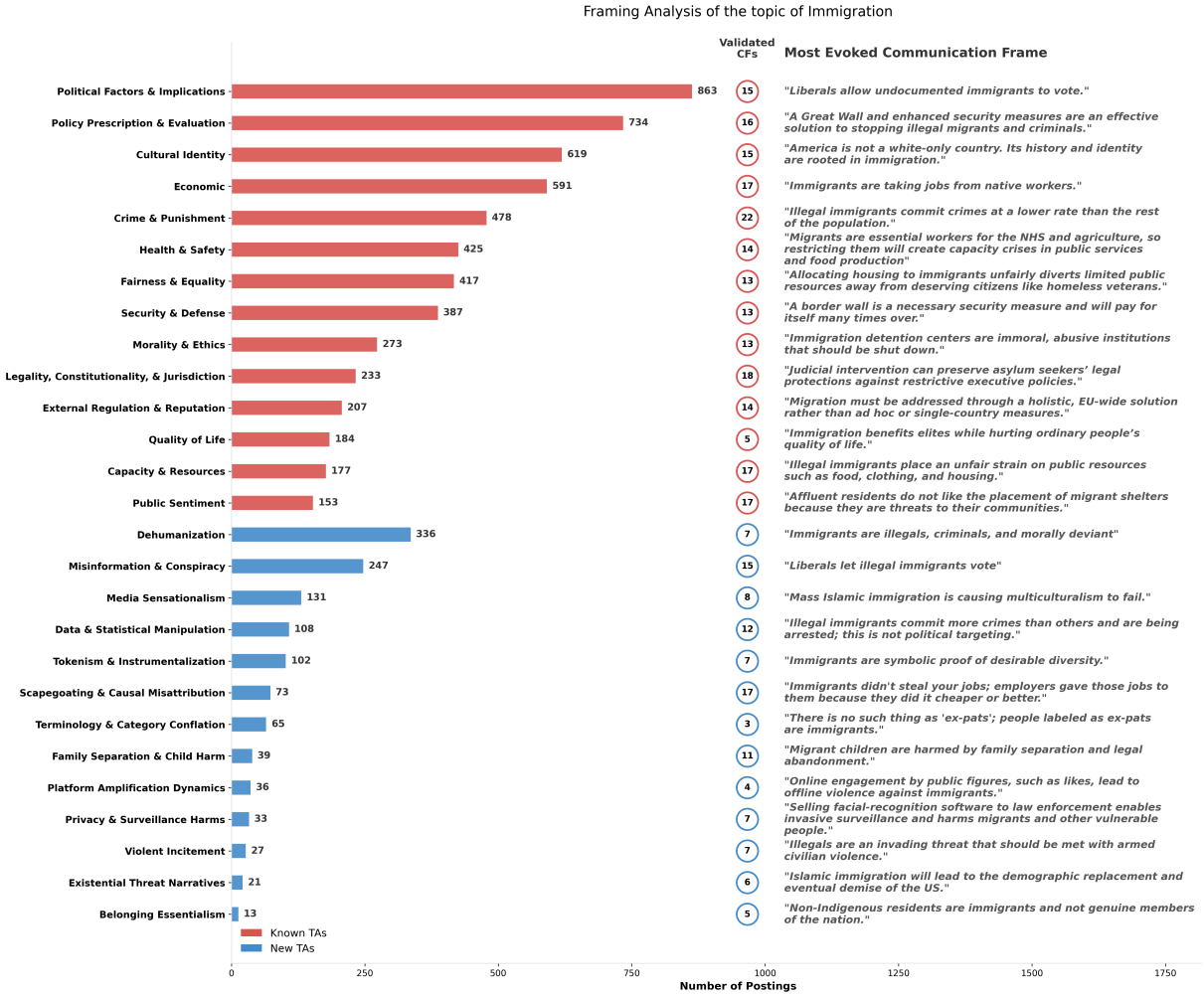

Figure 4: Framing Analysis of the topic of Immigration.

two properties: (a) the *soundness* of the rationale provided when articulating a CF; (b) the *clarity* of the CF articulation. Two linguists were tasked to judge the soundness and clarity of CFs, with N_S being the number of CFs deemed sound, and N_C the number of CFs deemed clear. With N_T the total number of CFs produced by **Art-FRAME**, *the quality of reasoning (Z)* involved in articulating CFs is computed as $Z = N_S/N_T$ while *the quality of the articulation (A)* of CFs is $A = N_C/N_T$. The results of $Z \geq .97$ for and $A \geq .95$ across all topics indicates high quality of the articulated CFs (see Appendix G.2). Also high Inter-Annotator Agreement was observed (see Appendix H).

5 Framing Analysis

Figure 4 illustrates the framing analysis for the immigration (IMM) topic, showing (a) the number of postings that address both the known and the new TAs; (b) the number of CFs articulated for each TA; and (c) examples of the most evoked frame for each TA. Appendix I details the analysis for the other three topics.For each topic, we can notice that the frame analysis allows us to understand which TAs are most discussed, enabling our understanding of *what* is aspects of a topic are framed the most. Moreover, the analysis shows the importance of discovering new TAs, as many of them are much more discussed that some of the known TAs. In addition, the CFs explain *why* some TAs are discussed the most and *how* they are framed.

6 Conclusion

This paper is describing **ID-TA**, an multi-agent framework for discovering topic aspects in social media discourse. It also details **Art-FRAME**, a second multi-agent framework capable to articulate the way topic aspects are framed. The operation in tandem of these two LLM architectures provides a novel way of understanding a topic's discourse.

7 Ethical Statement

We respected the privacy and honored the confidentiality of the authors who wrote the social media posts in the datasets from (Wei) and (Mendelsohn et al., 2021). We received approval from the Institutional Review Board at ANONYMIZED for working with these Twitter/X social media datasets, and no attempt was made to identify or profile individual users. All experimental settings, configurations, and procedures are clearly documented in this paper, the supplemental material, and the linked GitHub repository to support transparency and reproducibility. Human evaluations were conducted to assess the quality of the discovered Topic Aspects and the articulated Communication Frames, and to ensure the responsible and reliable use of human judgment, we measured and reported inter-annotator agreement (IAA). We do not identify any significant risks arising from this research. The goal of this work is to improve understanding of how public issues are discussed and framed in social media discourse, with a broader commitment to advancing research at the intersection of natural language processing and communication studies for the public good.

8 Limitations

This work has several limitations that point to directions for future research. First, our experiments are conducted exclusively on social media posts from a single platform, Twitter/X, and rely solely on textual content. In practice, many social media posts also incorporate images, videos, hyperlinks, and other forms of multimedia. In future work, we plan to extend our method to support multimodal data, which would enable its application to other social media platforms such as Instagram and Facebook, where discourse is often expressed through a combination of text, images, and videos.

Second, the proposed framework assumes the availability of a predefined set of known Topic Aspects for each topic, which serves as the starting point for TA identification and discovery. As described in Section 2.1, the Proposer Agent relies on these known TAs to perform Contrastive Chain-of-Thought (CCoT) reasoning, and they are also required to ground the multi-agent debate. While ID-TA is designed to discover new TAs beyond this initial set, its performance is influenced by the coverage and quality of the known TA inventory. In domains where such prior knowledge is unavailable, additional mechanisms may be needed to more effectively bootstrap the discovery process. Finally, although the multi-agent architecture is designed to mitigate common LLM failure modes such as hallucination and redundancy, the framework remains dependent on the reasoning capabilities of large language models. As a result, errors in interpretation or subtle model biases may propagate through the discovery and articulation pipeline. Moreover, the use of multi-agent debate and introspective question–answering loops increases computational cost and inference latency, which may limit scalability in large-scale or real-time settings.

References

Greg D. Adams. 1997. Abortion: Evidence of an issue evolution. *American Journal of Political Science*, 41(3):718–737.

Rakshitha Rao Ailneni and Sanda M. Harabagiu. 2025. Automatically discovering how misogyny is framed on social media. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 12189–12208, Albuquerque, New Mexico. Association for Computational Linguistics.

Francesco Barbieri, Jose Camacho-Collados, Luis Espinosa Anke, and Leonardo Neves. 2020. TweetEval: Unified benchmark and comparative evaluation for tweet classification. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1644–1650, Online. Association for Computational Linguistics.

Steven E. Barkan. 2014. Gender and abortion attitudes: Religiosity as a suppressor variable. *Public Opinion Quarterly*, 78(4):940–950.

Rodney Benson. 2013. *Shaping Immigration News: A French-American Comparison*. Communication, Society and Politics. Cambridge University Press.

Toby Bolsen and James N. Druckman. 2015. Counteracting the politicization of science. *Journal of Communication*, 65(5):745–769.

Toby Bolsen, James N. Druckman, and Fay Lomax Cook. 2014. How Frames Can Undermine Support for Scientific Adaptations: Politicization and the Status-Quo Bias. *Public Opinion Quarterly*, 78(1):1–26.

Amber E. Boydstun, Dallas Card, Justin Gross, Paul Resnick, and Noah A. Smith. 2018. Tracking the development of media frames within and across policy issues. *Carnegie Mellon University*.

Amber E. Boydstun and Justin H. Gross. 2013. Identifying media frames and frame dynamics within and across policy issues.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 438–444, Beijing, China. Association for Computational Linguistics.

Dallas Card, Justin Gross, Amber Boydstun, and Noah A. Smith. 2016. Analyzing framing through the casts of characters in the news. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1410–1420, Austin, Texas. Association for Computational Linguistics.

Yew Ken Chia, Guizhen Chen, Anh Tuan Luu, Soujanya Poria, and Lidong Bing. 2023a. Contrastive chain-of-thought prompting. *ArXiv*, abs/2311.09277.

Yew Ken Chia, Guizhen Chen, Luu Anh Tuan, Soujanya Poria, and Lidong Bing. 2023b. Contrastive chain-of-thought prompting. *Preprint*, arXiv:2311.09277.

Dennis Chong and James N. Druckman. 2012. Counter-framing effects. *The Journal of Politics*, 75(1):1–16.

Elizabeth Adell Cook, Ted G. Jelen, and Clyde Wilcox. 1993. Measuring public attitudes on abortion: Methodological and substantive considerations. *Perspectives on Sexual and Reproductive Health*, 25(3):118–.

Claes H. de Vreese. 2005. News framing: Theory and typology. *Information Design Journal*, 13:51–62.

James N. Druckman and Mary C. McGrath. 2019. The evidence for motivated reasoning in climate change preference formation. *Nature Climate Change*, 9(2):111–119.

Robert M. Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of Communications*.

Robert M. Entman. 2003. Cascading activation: Contesting the white house's frame after 9/11. *Political Communication*, 20:415 – 432.

Myra Marx Ferree. 2003. Resonance and radicalism: Feminist framing in the abortion debates of the united states and germany. *American Journal of Sociology*, 109(2):304–344.

Diana Greene Foster. 2020. *The Turnaway Study: Ten Years, a Thousand Women, and the Consequences of Having–or Being Denied–an Abortion*.

Mattis Geiger, Franziska Rees, Lau Lilleholt, Ana P. Santana, Ingo Zettler, Oliver Wilhelm, Cornelia Betsch, and Robert Böhm. 2022. Measuring the 7Cs of vaccination readiness. *European Journal of Psychological Assessment*, 38(4):261–269.

Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2021. Deberta: Decoding-enhanced bert with disentangled attention. In *International Conference on Learning Representations*.

Jan Fredrik Hovden and Hilmar Mjelde. 2019. Increasingly controversial, cultural, and political: The immigration debate in scandinavian newspapers 1970–2016. *Javnost - The Public*, 26(2):138–157.

Shanto Iyengar. 1990. Framing responsibility for political issues: The case of poverty. *Political Behavior*, 12(1):19–40.

Shanto Iyengar. 1991. Is anyone responsible? how television frames political issues. *Contemporary Sociology*, 21:467.

Alon Jacovi, Swabha Swayamdipta, Shauli Ravfogel, Yanai Elazar, Yejin Choi, and Yoav Goldberg. 2021. Contrastive explanations for model interpretability. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1597–1611, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Ted G. Jelen. 2014. The subjective bases of abortion attitudes: A cross-national comparison of religious traditions. *Politics and Religion*, 7(3):550–567.

Kirsti M. Jylhä and Kahl Hellmer. 2020. Right-wing populism and climate change denial: The roles of exclusionary and anti-egalitarian preferences, conservative ideology, and antiestablishment attitudes. *Analyses of Social Issues and Public Policy*, 20(1):315–335.

Anna Kheyfets, Shubhecchha Dhaurali, Paige Feyock, Farinaz Khan, April Lockley, Brenna Miller, and Ndidiamaka Amutah-Onukagha. 2023. The impact of hostile abortion legislation on the united states maternal mortality crisis: A call for increased abortion education. *Frontiers in Public Health*, 11:1291668.

Klaus Krippendorff. 2011. Computing krippendorff's alpha-reliability.

Haewoon Kwak, Jisun An, and Yong-Yeol Ahn. 2020. A systematic media frame analysis of 1.5 million new york times articles from 2000 to 2017. *Proceedings of the 12th ACM Conference on Web Science*.

Anthony Leiserowitz. 2006. Climate change risk perception and policy preferences: The role of affect, imagery, and values. *Climatic Change*, 77(1-2):45–72.

Anthony Leiserowitz and Nicholas Smith. 2017. Affective imagery, risk perceptions, and climate change communication.

Stephan Lewandowsky, Michael E. Mann, Nicholas J. L. Brown, and Harris Friedman. 2016. Science and the public: Debate, denial, and skepticism. *Journal of Social and Political Psychology*, 4(2):537–553.

Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang, Yan Wang, Rui Wang, Yujiu Yang, Shuming Shi, and Zhaopeng Tu. 2024. Encouraging divergent thinking in large language models through multi-agent debate. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 17889–17904, Miami, Florida, USA. Association for Computational Linguistics.

Scott Liebertz and Jaclyn Bunch. 2021. Backfiring frames: Abortion politics, religion, and attitude resistance. *Politics and Religion*, 14(3):403–430.

Matthew Lockwood. 2018. Right-wing populism and the climate change agenda: Exploring the linkages. *Environmental Politics*, 27(4):712–732.

Kristin Luker. 1984. *Abortion and the Politics of Motherhood*.

Minelle K. Mahtani. 2016. Shaping immigration news: a french-american comparison. *Journal of Ethnic and Migration Studies*, 42:680 – 683.

Jörg Matthes and Matthias Kohring. 2008. The content analysis of media frames: Toward improving reliability and validity. journal of communication. *Journal of Communication*, 58:258–279.

Dawn McCaffrey and Jennifer Keys. 2000. Competitive framing processes in the abortion debate: Polarization-vilification, frame saving, and frame debunking. *Sociological Quarterly*, 41(1):41–61.

Holly J. McCammon. 2022. A war of words over abortion: The legal-framing contest over the undue burden standard. *Justice System Journal*, 43(4):623–644.

Aaron M. McCright and Riley E. Dunlap. 2011. The politicization of climate change and polarization in the american public's views of global warming, 2001–2010. *Sociological Quarterly*, 52(2):155–194.

Julia Mendelsohn, Ceren Budak, and David Jurgens. 2021. Modeling framing in immigration discourse on social media. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2219–2263, Online. Association for Computational Linguistics.

Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38.

Marvin Minsky. 1986. *The Society of Mind*. Simon & Schuster.

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

Nona Naderi and Graeme Hirst. 2017. Classifying frames at the sentence level in news articles. In *Recent Advances in Natural Language Processing*.

Abigail Newell. 2019. Pro-life and pro-woman: Complicating the anti-abortion narrative in america.

Matthew C. Nisbet. 2009. Communicating climate change: Why frames matter for public engagement. *Environment: Science and Policy for Sustainable Development*, 51(2):12–23.

Rodrigo Nogueira and Kyunghyun Cho. 2020. Passage re-ranking with bert. *Preprint*, arXiv:1901.04085.

Stephen D. Reese, Oscar H. Gandy, and August E. (Eds.) Grant. 2001. *Framing Public Life: Perspectives on Media and Our Understanding of the Social World*. Routledge.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. 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 3982–3992, Hong Kong, China. Association for Computational Linguistics.

John Robinson. 2014. *Likert Scale*, pages 3620–3621. Springer Netherlands, Dordrecht.

Loretta J. Ross and Rickie Solinger. 2017. *Reproductive Justice: An Introduction*.

Bertram Scheufele. 2004. Framing-effects approach: A theoretical and methodological critique. *Communications*, 29(4):401–428.

Howard Schuman, Stanley Presser, and Jacob Ludwig. 1981. Context effects on survey responses to questions about abortion. *Public Opinion Quarterly*, 45(2):216–223.

Mrinank Sharma, Meg Tong, Tomasz Korbak, David Duvenaud, Amanda Askell, Samuel R. Bowman, Newton Cheng, Esin Durmus, Zac Hatfield-Dodds, Scott R. Johnston, Shauna Kravec, Timothy Maxwell, Sam McCandlish, Kamal Ndousse, Oliver Rausch, Nicholas Schiefer, Da Yan, Miranda Zhang, and Ethan Perez. 2025. Towards understanding sycophancy in language models. *Preprint*, arXiv:2310.13548.

Reva B. Siegel. 2007. The new politics of abortion: An equality analysis of woman-protective abortion restrictions. *University of Illinois Law Review*, 2007(3):991–1053.

Daniel Skinner. 2012. The politics of medical necessity in american abortion debates. *Politics & Gender*, 8(1):1–24.

Alexa Spence, Wouter Poortinga, and Nick Pidgeon. 2012. The psychological distance of climate change. *Risk Analysis*, 32(6):957–972.

Li Sun, Liu He, Shuyue Jia, Yangfan He, and Chenyu You. 2025. DocAgent: An agentic framework for multi-modal long-context document understanding. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 17712–17727, Suzhou, China. Association for Computational Linguistics.

Miles Turpin, Julian Michael, Ethan Perez, and Samuel R. Bowman. 2023. Language models don't always say what they think: unfaithful explanations in chain-of-thought prompting. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA. Curran Associates Inc.

Sander van der Linden. 2015. The social-psychological determinants of climate change risk perceptions: Towards a comprehensive model. *Journal of Environmental Psychology*, 41:112–124.

Sander van der Linden, Anthony Leiserowitz, Seth Rosenthal, and Edward Maibach. 2017. Inoculating the public against misinformation about climate change. *Global Challenges*, 1(2).

Dror Walter and Yotam Ophir. 2019. News frame analysis: An inductive mixed-method computational approach. *Communication Methods and Measures*, 13:248 – 266.

Joe H. Ward. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58:236–244.

Elke U. Weber. 2010. What shapes perceptions of climate change? *Wiley Interdisciplinary Reviews: Climate Change*, 1(3):332–342.

Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, Ed H. Chi, Tatsunori Hashimoto, Oriol Vinyals, Percy Liang, Jeff Dean, and William Fedus. 2022. Emergent abilities of large language models. *ArXiv*, abs/2206.07682.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2023. Chain-of-thought prompting elicits reasoning in large language models. *Preprint*, arXiv:2201.11903.

Maxwell Weinzierl and Sanda Harabagiu. 2024. Discovering and articulating frames of communication from social media using chain-of-thought reasoning. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1617–1631, St. Julian's, Malta. Association for Computational Linguistics.

Maxwell Weinzierl and Sanda M. Harabagiu. 2025. Investigating controversy framing across topics on social media. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 6795–6814, Suzhou, China. Association for Computational Linguistics.

Shirley Wu, Shiyu Zhao, Qian Huang, Kexin Huang, Michihiro Yasunaga, Vassilis N. Ioannidis, Karthik Subbian, Jure Leskovec, and James Zou. 2024. Avatar: Optimizing llm agents for tool usage via contrastive reasoning. *Advances in Neural Information Processing Systems 37*.

Kai Xiong, Xiao Ding, Yixin Cao, Ting Liu, and Bing Qin. 2023. Examining inter-consistency of large language models collaboration: An in-depth analysis via debate. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 7572–7590, Singapore. Association for Computational Linguistics.

Tianyi Zhang*, Varsha Kishore*, Felix Wu*, Kilian Q. Weinberger, and Yoav Artzi. 2020. Bertscore: Evaluating text generation with bert. In *International Conference on Learning Representations*.

A Example of Postings, Topic Aspects and Communication Frames for Immigration

Figure 5 illustrates an example of a Topic Aspect, namely TA_1 corresponding to the topic of immigration (IMM). The Figure also shows several postings discussing the topic and addressing TA_1 . The name and the definition of TA_1 are also provided.

The definition of TA_1 is also shown in Figure 5. Each TA has a *Name* and a *Definition*. As shown in Figure 5, the name of TA_1 is *"Economic"*. The definitions of TA_1 explains its name. This is different from Communication Frames (CFs) which provide a causal interpretation to a TA. We note that each TA typically has multiple CFs that provide different interpretations.

When identifying TAs in postings, the definition of each TA informs the entailment of the TA from the posting. However, when new TAs need to be discovered, their name and definitions needs to be generated.

The first two postings illustrated in Figure 5 evoke communication frame CF_1, while the next two postings evoke communication frame CF_1. Each of these communication frame provide a different causal interpretation of TA_1.

As shown in Figure 5, the first two postings evoke CF_1 which articulates a causal interpretation of TA_1 which (1) does not use any words appearing in the first two postings, is not a summary of those postings, but rather it expressed the impact of immigrants on economy as spending their earnings

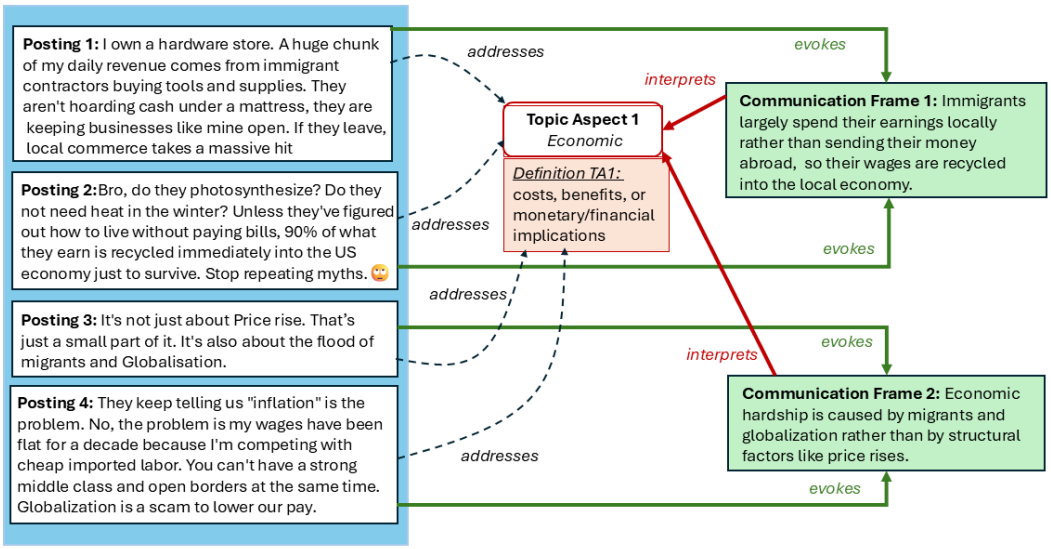

Figure 5: Example of an immigration Topic Aspect (TA) addressed by multiple postings and the Communication Frames (CFs) evoked by the postings.

locally instead of sending their money abroad, to their families.

The communication frame CF_2 illustrated in Figure 5 provides a different interpretation to TA_1 than the one offered by CF_1, because it articulates the reasons for the economic hardship faced by some folks today, namely (1) the immigrants; and (2) globalization. It also states that price rises have a smaller effect. As with CF_1, as shown in Figure 5, CF_2 is not a paraphrase or s summery of the postings that evoke it.

In the analysis presented in this paper, we have found that TA_1 is interpreted by 17 different CFs. We have also found that this TA was one of the first four most addressed TAs for the topic of immigration. It was addressed by 591 different postings. However, just knowing that all the 591 postings address the $TA_1 = Economic$ aspect of immigration, does not inform us about the specific ways in which we need to interpret and understand TA_1 in order to comprehend and form an opinion on immigration, based on the discourse from social media. The communication frames CF_1 and CF_2 exemplify some of the 17 possible interpretations of TA_1.

If one is aware of CF_1, because they were exposed only to the first two postings illustrated in Figure 5, they might have a favorable opinion of immigration, In contrast, if someone else would have been exposed only to the last two postings illustrated in Figure 5, they would have inferred the communication frame CF_2, and form an unfavorable opinion about immigration.

Knowing which communication frames are most evoked for each topic aspect can explain the dominant opinions about a topic discussed on social media. For example, if for topic aspect TA_1 shown in Figure 5 the dominant communication frame would be CF_2, we could have an explanation for anti-immigrant opinions.

#. TA Name	TA Definition
1. *Confidence*	Trust in the security and effectiveness of vaccinations, the health authorities, and the health officials who recommend and develop vaccines.
2. *Complacency*	Complacency and laziness to get vaccinated due to low perceived risk of infections.
3. *Constraints*	Structural or psychological hurdles that make vaccination difficult or costly.
4. *Calculation*	Degree to which personal costs and benefits of vaccination are weighted.
5. *Collective Responsibility*	Willingness to protect others and to eliminate infectious diseases.
6. *Compliance*	Support for societal monitoring and sanctioning of people who are not vaccinated.
7. *Conspiracy*	Conspiracy thinking and belief in fake news related to vaccination.

Table 6: Names and Definitions of Known Topic Aspects for the topic of Vaccine Hesitancy (VH).

B Topic Aspects and Definitions

B.1 Known Topic Aspects

Table 6 lists the known TAs and their definitions for the topic of COVID-19 Vaccine Hesitance

(VH), provided in (Weinzierl and Harabagiu, 2024). These TAs were informed by the seven factors reported in (Geiger et al., 2022) to increase or decrease an individual's likelihood of getting vaccinated. The names and the definitions of these known TAs were provided in (Geiger et al., 2022).

Table 7 lists all the names and definitions for the known TAs for the topic of Immigration (IMM). These TAs represent the dimensions of framings introduced in (Card et al., 2015). Both the sames and the definitions of these TAs have been reported in (Card et al., 2015).

For the topic of Climate Change (CC), Table 8 lists all the known TAs and their definitions. These TAs have been first described in (Weinzierl and Harabagiu, 2025). The TA names and their definitions originate, as reported in (Weinzierl and Harabagiu, 2025) is a multitude of academic papers addressing the topic of climate change, namely (Nisbet, 2009; Lewandowsky et al., 2016; van der Linden et al., 2017) for the first known TA; (Druckman and McGrath, 2019; Bolsen and Druckman, 2015) for the second known TA; (McCright and Dunlap, 2011; van der Linden, 2015) for the third known TA, (Lockwood, 2018; Jylhä and Hellmer, 2020) for the fourth known TA; (Leiserowitz, 2006; Leiserowitz and Smith, 2017) for the fifth known TA and (Weber, 2010; Spence et al., 2012) for the sixth known TA. This shows that a extensive inspection of academic literature is needed when requiring awareness of known TAs for some topics.

For the topic of Abortion (AB), we have relied on the known TAs and their definitions first reported in (Weinzierl and Harabagiu, 2025). The names and definitions of these 17 TAs are listed in Table 9. These TAs also resulted from the inspection of a large number of academic articles.

Table 9 lists of the TAs and their definitions for the topic of Abortion (AB), available from (Weinzierl and Harabagiu, 2025).

#. **TA Name**	**TA Definition**
1. *Economic*	The costs, benefits, or monetary/financial implications of immigration.
2. *Capacity & Resources*	The availability or lack of time, physical, human, or financial resources.
3. *Morality & Ethics*	Perspectives compelled by religion or secular sense of ethics or social responsibility.
4. *Fairness & Equality*	The (in)equality with which laws, punishments, rewards, resources are distributed.
5. *Legality, Constitutionality & Jurisdiction*	Court cases and existing laws that regulate policies; constitutional legal processes such as seeking asylum or obtaining citizenship; jurisdiction.
6. *Crime & Punishment*	The violation of policies in practice and the consequences of those violations.
7. *Security & Defense*	Any threat to a person, group, or nation and defenses taken to avoid that threat.
8. *Health & Safety*	Health and safety outcomes of a policy issue, discussions of health care.
9. *Quality of Life*	Effects on people's wealth, mobility, daily routines, community life, happiness.
10. *Cultural Identity*	Social norms, trends, values, and customs; integration/assimilation efforts.
11. *Public Sentiment*	General social attitudes, protests, polling, interest groups, public passage of laws.
12. *Political Factors & Implications*	Focus on politicians, political parties, governing bodies, political campaigns and debates; discussions of elections and voting.
13. *Policy Prescription & Evaluation*	Discussions of existing or proposed policies and their effectiveness.
14. *External Regulation & Reputation*	Relations between nations or states/provinces; agreements between governments; perceptions of one nation/state by another.

Table 7: Names and Definitions of Known Topic Aspects for the topic of Immigration (IMM).

#. **TA Name**	**TA Definition**
1. *Scientific Uncertainty & Climate Denial*	Framing climate science as uncertain or unsettled, casting doubt on the scientific consensus, or portraying climate change as a hoax. This reduces trust in experts and delays action. (Nisbet, 2009; Lewandowsky et al., 2016; van der Linden et al., 2017)
2. *Ideological Polarization & Identity*	Presenting climate change as a partisan issue aligned with a specific political ideology, leading to social division and resistance based on group identity rather than scientific facts. (Druckman and McGrath, 2019; Bolsen and Druckman, 2015)
3. *Economic Trade-offs (Environment vs. Economy)*	Framing climate policies as harmful to economic growth, job security, and affordability, emphasizing short-term economic costs over long-term benefits or the costs of inaction. (McCright and Dunlap, 2011; van der Linden, 2015)
4. *Elitist vs. Public Needs (Populist Framing)*	Portraying climate action as an agenda pushed by global elites, bureaucrats, and scientists, disconnected from the struggles of ordinary citizens. This frame suggests policies are unfair or imposed without grassroots support. (Lockwood, 2018; Jylhä and Hellmer, 2020)
5. *Catastrophic "Doomsday" Narratives & Fatalism*	Overemphasizing apocalyptic consequences of climate change without solutions, leading to public fear, helplessness, and disengagement due to perceived inevitability of disaster. (Leiserowitz, 2006; Leiserowitz and Smith, 2017)
6. *Psychological Distance & Abstraction*	Framing climate change as a distant, abstract issue affecting future generations, remote places, or complex scientific processes, making it seem less relevant to people's immediate lives and concerns. (Weber, 2010; Spence et al., 2012)

Table 8: Names and Definitions of Known Topic Aspects for Climate Change (CC).

#. TA Name	TA Definition
1. *Women's Rights*	Emphasizes abortion as a fundamental right and an aspect of gender equality, framing the decision as a personal choice that the state should not interfere with. (Newell, 2019; Luker, 1984; Siegel, 2007)
2. *Fetal Rights*	Frames abortion as the taking of a human life, prioritizing the moral and legal rights of the fetus, often invoking the concept of personhood from conception. (Newell, 2019; Luker, 1984; Siegel, 2007)
3. *Sanctity of Life*	Portrays abortion as morally wrong under virtually all circumstances, often rooted in religious or ethical arguments regarding the sacredness of human life. (Barkan, 2014; Jelen, 2014; Luker, 1984)
4. *Personal Morality*	Emphasizes that abortion is a complex moral decision best left to individual conscience rather than dictated by the state. (Barkan, 2014; Jelen, 2014; Luker, 1984)
5. *Legality & Constitutional Rights*	Focuses on abortion as a constitutional right tied to privacy and liberty, debating legal standards such as undue burden and personhood. (Skinner, 2012; McCammon, 2022; Siegel, 2007)
6. *Legislative & States' Rights*	Argues that abortion should be regulated by state legislatures rather than courts, emphasizing democratic decision-making and states' authority. (Adams, 1997; Cook et al., 1993; Liebertz and Bunch, 2021)
7. *Health & Safety*	Positions abortion as a necessary medical procedure that is safe when legally performed, while highlighting the risks of illegal or unsafe procedures. (Kheyfets et al., 2023; Miller, 2019)
8. *Abortion Harms Women*	Claims that abortion poses physical and psychological risks to women, using disputed evidence to support restrictions. (Siegel, 2007; Skinner, 2012)
9. *Medical Necessity*	Debates whether abortion is ever truly necessary for maternal health, with pro-choice advocates highlighting exceptions and pro-life advocates arguing modern medicine diminishes the need. (Siegel, 2007; Skinner, 2012)
10. *Economic Consequences*	Abortion is economically essential for women's financial stability and career prospects, especially affecting low-income and marginalized groups. (Miller, 2019; Foster, 2020)
11. *Abortion as Exploitation*	Argues that abortion disproportionately harms disadvantaged communities, sometimes framed as a form of racial or economic oppression. (Ross and Solinger, 2017; McCaffrey and Keys, 2000)
12. *Gender Equality*	Asserts that access to abortion is essential for ensuring equal opportunities in education, employment, and public life. (Ross and Solinger, 2017; Siegel, 2007; Ferree, 2003)
13. *Motherhood & Traditional Roles*	Abortion is undermining the value of motherhood, suggesting that women deserve more support to choose parenthood over termination. (Ross and Solinger, 2017; Siegel, 2007; Ferree, 2003)
14. *Reproductive Justice*	Expands the debate to include racial, economic, and social factors, emphasizing the right to parent in safe and supportive environments. (Ross and Solinger, 2017)
15. *Religious Freedom*	Argues that abortion laws should either reflect religious beliefs about life or avoid imposing a particular religious viewpoint, thus protecting individual religious freedom. (Jelen, 2014; Barkan, 2014)
16. *Public Opinion & Political Polarization*	Considers how partisan divides and identity politics reinforce entrenched positions in the abortion debate, contributing to polarization. (McCaffrey and Keys, 2000; Adams, 1997; Liebertz and Bunch, 2021)
17. *Language & Terminology*	Highlights how strategic word choices (e.g., "pro-choice" vs. "pro-life", "fetus" vs. "unborn baby") frame the debate and influence public perception. (Cook et al., 1993; Schuman et al., 1981)

Table 9: Names and Definitions of Known Topic Aspects for Abortion (AB).

B.2 New Topic Aspects

#. TA Name	TA Definition
1. *Dehumanizing Immigrants*	Language that strips migrants of dignity or personhood using slurs, animalizing or invasion metaphors, blanket criminalization, or calls to exclude.
2. *Misinformation & Conspiracy*	False, unverified, or conspiratorial claims about immigration (facts, numbers, laws, motives) that mislead audiences, including fabricated statistics, replacement-style plots, and foreign or coordinated disinformation.
3. *Data & Statistical Manipulation*	Cherry-picking, opaque methods, misleading graphics, and selective metrics that distort immigration trends, impacts, or causal claims under a veneer of quantitative authority.
4. *Media Sensationalism & Framing*	Selective headlines, emotive imagery, anecdotes, and performative coverage that dramatize or skew immigration stories, steering attention and debate away from balanced evidence and context.
5. *Platform Amplification Dynamics*	Algorithmic boosts, viral mechanics, and coordinated or elite signaling that disproportionately surface and normalize immigration frames (fueling echo chambers, harassment, and rapid spread of misleading content).
6. *Terminology & Category Conflation*	Loaded or imprecise labels and the collapsing of distinct categories (asylum seekers, visa overstays, trafficking, refugees) into a single 'immigration' bucket, which misleads audiences about rules, actors, and remedies.
7. *Existential Threat Narratives*	Portraying immigration as an existential threat via replacement, invasion, or extreme historical analogies that recasts migrants as civilizational enemies and justifies discrimination.
8. *Scapegoating & Causal Misattribution*	Blaming immigrants for complex social or economic problems and asserting simple, unsupported causal stories while ignoring structural drivers or counterevidence.
9. *Privacy & Surveillance Harms*	Collecting, exposing, or commercializing migrants' personal/biometric data and monitoring their activities in ways that enable doxxing, intimidation, chilling effects, or misuse by authorities and third parties.
10. *Family Separation & Child Harm*	Rhetoric and practices that normalize or justify separating children from caregivers or otherwise harming migrant children (deterrence, opaque custody, exploitation), sidelining child-specific rights and welfare.
11. *Tokenism & Instrumentalization*	Using migrants as props, tokens, or utilitarian symbols through curated stories, art/marketing, or superficial consultation thereby reducing people to optics, leverage, or single representatives of diverse communities.
12. *Belonging Denial & Gatekeeping*	Framings that police who counts as a "real" member of the nation by treating some groups as perpetual foreigners, privileging certain origins, or ethnicizing national identity in ways that exclude or delegitimize migrants and their descendants.
13. *Violent Incitement*	Rhetoric that praises, threatens, or mobilizes violence against migrants including dehumanizing calls for harm or vigilantism thereby escalating discourse into direct intimidation and risk.

Table 10: Names and Definitions of New Topic Aspects for Immigration (IMM).

#. TA Name	TA Definition
1. *Mechanistic Misinformation*	Misstates how COVID-19 vaccines or the virus work—e.g., DNA alteration, 'spike protein toxin,' ADE, LNP biodistribution, or shedding—often cloaked in technobabble to imply hidden harms.
2. *Contextless Statistics*	Presents vaccine numbers without essential context—denominators, endpoints, timeframes, severity—or uses mismatched metrics, creating misleading risk or efficacy impressions.
3. *Zero-Risk Standards*	Demands absolute safety, perfect proof, or long-term certainty before vaccinating or easing policies, using uncertainty or absent data as decisive evidence against vaccination.
4. *Natural Immunity Supremacy*	Claims infection-derived immunity is inherently superior or sufficient and uses that assertion to argue vaccination is unnecessary, ignoring comparative risks, durability, and population impact.
5. *Treatment Substitution*	Promotes unproven or decontextualized therapeutics and remedies as replacements for vaccination, implying vaccines are unnecessary if such alternatives are used.
6. *Authority-as-Evidence*	Substitutes status cues—endorsements, titles, celebrities, group behavior, or policy statements—for scientific evidence to justify claims about vaccines.
7. *Human Experimentation Narrative*	Frames vaccination as unethical human experimentation or a mass clinical trial—often invoking the Nuremberg Code—to justify refusal or alarm.
8. *Rights & Coercion Framing*	Portrays mandates, verification, or incentives as authoritarian coercion or absolute rights violations, collapsing nuanced public-health and legal considerations into moral outrage.
9. *Partisan Weaponization*	Recasts vaccine issues as partisan identity battles or elite score-setting, substituting political motives for engagement with public-health evidence.

Table 11: Names and Definitions of New Topic Aspects for COVID-19 Vaccine Hesitancy (VH).

#. **TA Name**	**TA Definition**
1. *Weather–Climate Conflation*	Using single weather events or short-term local observations to infer, prove, or disprove long-term climate trends, ignoring variability and attribution science.
2. *Individual–Action Fixation*	Portraying climate change as primarily solvable via personal lifestyle or consumer choices (often marketized), diverting attention from systemic, corporate, and policy levers.
3. *Performative Signaling*	Substituting hashtags, slogans, events, and visibility for concrete policies, mechanisms, and accountability—often creating an illusion of engagement or progress and enabling delay.
4. *Single-Fix Solutionism*	Claiming a single policy, technology, or lever can solve climate change, glossing over system complexity, trade-offs, feasibility, and the need for complementary measures.
5. *Technocratic Fixation*	Casting climate action as an expert-managed technical or market service, sidelining politics, democracy, justice, and public deliberation about causes, responsibilities, and trade-offs.
6. *Authority-First Validation*	Validating or dismissing climate claims primarily based on who delivers them (celebrity, institution, or credentials), which misattributes expertise and overextends authority instead of focusing on evidence.
7. *Causal Oversimplification*	Explaining complex climate phenomena, impacts, or policy debates through a single cause, lone event, or decontextualized statistic, ignoring uncertainty, interactions, and multiple drivers.
8. *Misleading Metrics & Baselines*	Using numbers without clear sources, baselines, timeframes, or uncertainty, often cherry-picking records or targets to inflate certainty or mislead about magnitude or progress.
9. *Moralized Blame & Duty*	Framing climate action through guilt, shame, universalized culpability, or moral or religious duty, often using demonizing labels that crowd out analysis of mechanisms, trade-offs, and policy design.
10. *Mockery & Trivialization*	Using sarcasm, memes, or ridicule to belittle climate science, impacts, or solutions, substituting derision for substantive engagement.

Table 12: Definitions of New Topic Aspects for Climate Change (CC).

#. **TA Name**	**TA Definition**
1. *Authority as Argument*	Invokes revered figures, celebrities, institutions, or scripture as decisive proof about abortion, substituting authority for reasoning or evidence.
2. *Population-Control Conspiracy*	Claims abortion policy constitutes a coordinated plot to exterminate or manage targeted populations without credible evidence.
3. *Violent Remedy or Incitement*	Normalizes, endorses, or recruits for violence or elimination as the solution to abortion disputes, displacing lawful and ethical debate.
4. *Shock and Doom Rhetoric*	Uses graphic depictions and apocalyptic claims of societal collapse to provoke fear and disgust rather than provide evidence or reasoned trade-offs.
5.*Adoption as Panacea*	Presents adoption as a simple, universal substitute for abortion, ignoring the medical, emotional, legal, and systemic realities involved.
6.*Contraception–Abortion Conflation*	Collapses or confuses contraception and abortion, or misstates contraception efficacy, to skew ethical responsibility and policy debate.
7.*Science/Marker Reductionism*	Treats single biological milestones or disputed scientific claims (e.g., heartbeat, fetal pain, DNA, birth) as dispositive of personhood or legality, collapsing ethical and legal complexity into a biomarker.
8. *Ad Hominem and Shaming*	Attacks character, identity, or dignity—via insults, stigma, caricature, or moral vilification—instead of addressing abortion arguments or evidence.
9. *Atrocity and Crime Analogies*	Portrays abortion or its supporters as akin to murder, genocide, or warfare to inflate moral stakes and foreclose nuance.
10. *Statistical Misuse*	Uses raw counts, unsourced figures, precise-looking ratios, or mismatched denominators as decisive proof about abortion without methodological context.
11. *Legal Distortion and Extreme-Case Focus*	Misstates abortion laws or spotlights rare, late-term scenarios to suggest extreme practices are typical or legally endorsed.

Table 13: Definitions of New Topic Aspects for Abortion (AB).

C Prompting Details

System Prompt for the Investigator Agent:

IDENTITY
You are the Investigator Agent. Your role is to determine, for each post, which of the given Topic Aspects (TAs) are addressed.
##INPUTS
You will receive a set of TAs along with their definitions. Optionally, you may also receive some clarifying questions about the applicability of the given TA(s) to the post, together with the corresponding answers.
##OBJECTIVES
For each TA, you must assign exactly one of the following labels:
(a) Applicable: The post clearly addresses the TA.
(b) Not applicable: The post clearly does not address the TA.
(c) Uncertain: It is unclear whether the post addresses the TA.
You must also provide a brief rationale justifying each decision.
INSTRUCTIONS
- Base decisions strictly on explicit textual evidence from the post.
- If you are **not confident** that a TA is addressed, assign the label **Uncertain**.
- If clarifying question–answer pairs are provided, you must take them into account and decide whether the TA is **Applicable** or **Not applicable**; in this case, you may **not** choose **Uncertain**.
- If a clarifying question yields an ambiguous or unavailable answer, assign **Not applicable** to the corresponding TA.

System Prompt for the Question Generator Agent:

IDENTITY
You are the Question Generator Agent. Your role is to generate targeted clarification questions that determine whether a specific Topic Aspect (TA) is addressed by a post.
##INPUTS
You will receive the ***uncertain*** TA (along with its definition) and the rationale from the investigator agent. You may also receive feedback from the answer generator agent.
##OBJECTIVES
Your objective is to generate **precise, minimal yes/no questions** that test whether the post satisfies the defining criteria of the TA.
##QUESTION GENERATION RULES
- Each question must be grounded in the TA definition.
- Questions must not assume intent, motivation, or unstated context.
- Questions must be answerable using the post text alone.

QUESTION REFINEMENT RULES
- If the Answer Generator responds with **CANNOT ANSWER**, you must generate a more specific follow-up question.
- Refinement must narrow the scope of the question and reduce reliance on implied context.

System Prompt for the Answer Generator Agent:

IDENTITY
You are the Answer Generator Agent. Your role is to answer clarification questions about a Topic Aspect applicability to a given post.
##INPUTS
A yes/no clarification question generated by the Question Generator.
##OBJECTIVES
Your objective is to produce one of the following responses:
(a) Yes
(b) No
(c) CANNOT ANSWER
You must also provide a brief rationale for each answer.
ANSWERING RULES
- Base decisions strictly on textual evidence from the post. You should not make overreaching assumptions when you answer.
- Use **CANNOT ANSWER** when the post provides insufficient information or the question is too broad to decisively answer *Yes* or *No*.
- Keep the rationale factual and concise.

Figure 6: System Prompts for the Investigator, Question Generator, and Answer Generator Agents.

System Prompt for the Proposer Agent:

IDENTITY
You are the Proposer Agent. Your role is to propose one or more candidate topic aspects (***$TA_i^{Candidate}$***) for each input post.
##INPUTS
- Known TAs along with their definitions.
- Examples of posts that **address known TAs**, accompanied by rationales explaining why the post addresses those TAs.
- Examples of posts that **do not address known TAs**, accompanied by rationales explaining why the post does not address those TAs.

##OBJECTIVES
Your objective is to propose one or more $TA_i^{Candidate}$ that are addressed by the input post but are not adequately captured by any existing TA. For each $TA_i^{Candidate}$, you must produce:
1. A clear and descriptive TA name.
2. A precise definition of the TA.
3. Explicit textual evidence from the post supporting the TA.
4. A rationale explaining why the known TAs are insufficient to capture this aspect.

INSTRUCTIONS
- Do not reuse or lightly rephrase existing TA names or definitions.

System Prompt for the Orchestrator Agent:

##IDENTITY
You are the LLM Orchestrator Agent. Your role is to coordinate and supervise the discovery of new Topic Aspects (TAs) from a collection of social media postings.
##INPUTS
- The current **TA Memory**, which contains both known and previously discovered Topic Aspects (TAs), including their names, definitions, associated postings, and rationales.
- A **semantic similarity function** based on Sentence-BERT (SBERT) that computes cosine similarity between the definition of a candidate Topic Aspect $TA_i^{Candidate}$ and the definitions of any existing Topic Aspects $TA_j^{K/N}$ (known or new) stored in the TA Memory.
- A **similarity threshold** S_k .

##OBJECTIVES
Your objective is to govern the lifecycle of new Topic Aspect discovery by:
1. Initiating the proposal of candidate new TAs for a posting.
2. Determining whether a proposed TA is redundant with existing TAs using the semantic similarity function and the threshold S_k.
3. Initiating a multi-agent debate to validate genuinely novel TAs through adversarial evaluation.
4. Updating the TA Memory based on the outcome of the debate.

INSTRUCTIONS
- For each proposed candidate TA, compute its semantic similarity with all existing TAs in the TA Memory using the provided similarity function.
- If the similarity score exceeds the threshold S_k for any existing TA, treat the candidate as redundant and do not forward it to the multi-agent debate.

Figure 7: System Prompts for the Proposer and the Orchestrator Agent.

System Prompt for the Champion Agent:

IDENTITY
You are the Champion Agent in a debate. You will receive a candidate topic aspect (***$TA_i^{Candidate}$***) generated by the Proposer Agent. You are a rigorous advocate for the $TA_i^{Candidate}$.
##INPUTS
You will receive the known TAs with definitions as well as the $TA_i^{Candidate}$, its definition, evidence grounding and the rationale.
##OBJECTIVES
Your sole objective is to defend the $TA_i^{Candidate}$ based on:

a) **Applicability:** show that the $TA_i^{Candidate}$ is truly addressed by the post. Identify specific spans in the post that justify the presence of the $TA_i^{Candidate}$.
b) **b) Discriminant Validity:** show that the $TA_i^{Candidate}$ is distinct from all the known TAs. Your primary goal is to demonstrate that $TA_i^{Candidate}$ is not merely a synonym for an existing TA but can elicit distinct framings.

INSTRUCTIONS
- Do NOT Change the $TA_i^{Candidate}$: You must defend the Name and Definition exactly as given. Do not rewrite them.
- Do NOT Hallucinate: Do not infer intent or meaning that is not supported by the post.
- Avoid Tautology: Do not simply repeat the definition. Explain *why* it fits.

System Prompt for the Critic Agent:

IDENTITY
You are the Critic Agent in a debate. You will receive a candidate topic aspect (***$TA_i^{Candidate}$***) generated by the Proposer Agent. You are a rigorous skeptic whose role is to challenge the validity of the $TA_i^{Candidate}$.
##INPUTS
You will receive the known TAs with definitions as well as the $TA_i^{Candidate}$ and its definition, evidence grounding and the rationale.
##OBJECTIVES
Your sole objective is to reject the $TA_i^{Candidate}$ based on:

a) **Applicability:** show that the $TA_i^{Candidate}$ is not genuinely by the post. Identify missing, weak, or ambiguous textual evidence. Highlight cases where the justification relies on speculation, implicit assumptions, or inferred intent rather than explicit content.
b) **Discriminant Validity: show that** the $TA_i^{Candidate}$ is redundant and insufficiently distinct from known TAs.
c) ## INSTRUCTIONS

- Do NOT Change the $TA_i^{Candidate}$. You must defend the Name and Definition exactly as given. Do not rewrite them.
- Do NOT Hallucinate: Base all critiques strictly on the content of the post and provided definitions.

System Prompt for the Judge Agent:

IDENTITY
You are the Judge Agent. You do not take sides. Your only loyalty is to the rigorous validation of a Candidate Topic Aspect (***$TA_i^{Candidate}$***).
##INPUTS
You will with receive the known TAs with definitions and the arguments of the Champion and the Critic Agent.
CAPABILITIES
1. Dynamic Stop: You decide when the debate ends. If the arguments are clear, issue a VERDICT.
2. Refinement: You are not limited to "Yes/No". If the Champion presented valid arguments, but the $TA_i^{Candidate}$ is named poorly or defined too broadly, you must regenerate it.

DECISION CRITERIA
1. Discriminant Validity: Is the $TA_i^{Candidate}$ strictly distinct from the known TAs?
2. Applicability: Is the $TA_i^{Candidate}$ actually addressed by the post?
   - If the Champion relies on overreaching assumptions, you must REJECT.

INSTRUCTIONS
- Listen to the Champion and Critic.
- VERDICT OPTIONS:
  - **ACCEPT:** $TA_i^{Candidate}$ is perfect as is.
  - **REJECT:** $TA_i^{Candidate}$ is redundant, hallucinated, or trivial.
  - **CORRECT:** The Champion presented valid arguments, but the $TA_i^{Candidate}$ needs a better name or stricter definition to avoid the Critic's objections.

Figure 8: System Prompts for the Multi-Agent Debate.

System Prompt for the Articulation Agent:

IDENTITY
You are the Articulation Agent. Your role is to articulate **Communication Frames (CFs)** evoked by a post.
CF DEFINITON
##INPUTS
Along with the input post, you will receive the full set of Topic Aspects (TAs) addressed by the post (*known and new*).
##OBJECTIVES
A single post may articulate multiple Communication Frames. Each CF may interpret a subset or all the TAs addressed by the post. For each **articulated CF**, you must provide:
(a) Interpreted TAs: A list of TAs, selected from the TAs addressed by the post, that the CF interprets.
(b) Rationale: A concise explanation of why the CF constitutes an interpretive explanation of the listed TA(s).
INSTRUCTIONS
- You must capture the CF *exactly* as the user intends, even if it is controversial, biased, conspiratorial, or factually incorrect.
- State the CF as a direct assertion of truth from the perspective of the author, avoiding meta-language like *'portrayed as', 'framed as',* or *'suggests that'.*
- Do not force every sentence into a frame; only articulate a frame if the text presents a distinct reasoning, causal link, or strong perspective.

System Prompt for the Paraphrase-Police Agent:

IDENTITY
You are the Paraphrase-Police Agent, responsible for identifying **paraphrased** Communication Frames (CFs).
##INPUTS
An aggregated list of CFs articulated for a given Topic Aspect (TA), where each CF is articulated at the post level by the articulation agent.
##OBJECTIVES
Your objective is to identify CFs that are paraphrases of one another and group them accordingly. Two CFs are considered paraphrases if they convey the same underlying meaning despite differences in wording or phrasing. In other words, if an individual agrees with CF *X*, they would also agree with CF *Y,* and vice versa. CFs that do not share this semantic equivalence should be placed in separate groups.
##INSTRUCTIONS
- Focus strictly on **semantic meaning**, not surface wording or stylistic variation.
- If a CF has no paraphrase among the provided list, it should form a group by itself.

System Prompt for the Explainability Agent:

IDENTITY
You are the Explainability Agent, responsible for validating a Communication Frame (CF) through a reconstruction test. Your role is to determine whether a CF can be faithfully articulated using only its rationale and the Topic Aspects (TAs) it is intended to interpret.
CF DEFINITON
##INPUTS
You will be provided with the original rationale generated alongside a CF, which explains why that CF interprets the given Topic Aspect(s).
##OBJECTIVES
Your objective is to reconstruct the CF using only the provided rationale and the TA definitions, **without access to the original posting or the original CF text.**
##INSTRUCTIONS
- Regenerate the CF in a manner that is consistent with the rationale and the intended TA(s).

Figure 9: System Prompts for the Art-FRAME Module.

System Prompt

You are given a cluster of social media postings that are assumed to address the same **unknown Topic Aspect (TA)**. The postings in the cluster were grouped based on semantic similarity and are expected to reflect a coherent underlying aspect of discussion.
TASK
Your task is to generate a **single new Topic Aspect (TA)** that best captures the shared theme expressed across all postings in the cluster.
OUTPUT REQUIREMENTS
1. **TA Name:** A concise, descriptive name that captures the central theme shared by the cluster.
2. **TA Definition:** A precise and generalizable definition describing the common aspect addressed by the postings.

INSTRUCTIONS
- Base the TA name and definition **only** on the content of the provided postings.
- Do **not** reference individual users, platforms, or metadata.
- Do **not** introduce multiple Topic Aspects; generate exactly **one** TA per cluster.
- Avoid overly specific phrasing tied to a single post; focus on the shared semantic content across the cluster.
- The TA definition should abstract away from individual posts while remaining faithful to their shared meaning.

Figure 10: System prompt used for the TAD_{HAC} baseline to generate a TA name and definition from a cluster of semantically similar postings obtained via Hierarchical Agglomerative Clustering.

D Example of Introspective Question-Answering Loop in ID-TA

Post text: New Federation report on how restrictive immigration policies are creating a "new walled order". This criminalisation of compassion is extremely worrying. Read more here: #Migration #ProtectHumanity.
Applicable TAs: Morality & Ethics, Crime & Punishment, Policy Prescription & Evaluation
Uncertain TAs: Security & Defense (TA1), Quality of life (TA2)
Round 1 (TA1):
Question Generator:
Clarifying Question 1 : *Does the report frame the 'new walled order' primarily as a national security measure?*
Answer Generator:
Answer: No
Rationale: The text uses the phrase 'new walled order' but frames it as a 'criminalisation of compassion' and links it to 'ProtectHumanity'. It does not explicitly frame it as a necessary security measure.
Round 1 (TA2):
Question Generator:
Clarifying Question 1 : *Does the post describe concrete impacts on migrants' daily lives and services?*
Answer Generator:
Answer: No
Rationale: The text describes impacts using abstract, metaphorical terms ("new walled order," "criminalization of compassion") rather than describing concrete daily life impacts (such as access to food, shelter, or healthcare services).
Final Investigator Decision:
Security & Defense: Not Applicable
Quality of life: Not Applicable

Figure 11: Example of a Introspective Question-Answering Loop illustrating a case where both uncertain TAs are found to be not applicable.

Figures 11 and 12 illustrate the operation of the introspective question–answering loop triggered when the Investigator Agent is uncertain about the applicability of one or more known Topic Aspects (TAs) to a given posting. In such cases, the Investigator Agent invokes a Question Generator Agent and an Answer Generator Agent to resolve ambiguity through targeted Yes/No questions grounded exclusively in the textual evidence of the post.
Figure 11 presents a case in which two known Topic Aspects, Security & Defense (TA1) and Quality of Life (TA2), are initially marked as uncertain by the Investigator Agent. For each TA, the Question Generator Agent formulates a targeted Yes/No question to test the defining criteria of the TA against the content of the post. For TA1, the Question Generator Agent asks whether the phrase "new walled order" is used as a national security measure. The Answer Generator Agent responds No, explaining that the phrase is employed to argue against criminalizing compassion rather than to invoke border protection or security enforcement. The Investigator Agent considers both the negative answer and its rationale and concludes that TA1 is not applicable. A similar introspective exchange is carried out for TA2, after which the Investigator Agent determines that neither TA is supported by sufficient textual evidence, resulting in both being marked as not applicable.
Figure 12 illustrates a scenario in which two uncertain TAs are evaluated for the same posting, resulting in Crime & Punishment (TA1) being deemed applicable, while External Regulation & Reputation (TA2) is determined to be not applicable. For TA2, the Question Generator Agent initially poses

Post text: And meanwhile.... he is sending unwanted migrants in unmarked police cars to the Spanish border.... how's that for hypocrisy!
Applicable TAs: Security & Defense
Uncertain TAs: Crime & Punishment (TA1), External Regulation & Reputation (TA2)
Round 1 (TA1):
Question Generator:
Clarifying Question 1: *Does the text explicitly mention law enforcement agents, legal proceedings, or the violation of standard procedures?*
Answer Generator:
Answer: Yes
Rationale: The text explicitly identifies the agents as operating "police cars" (law enforcement machinery). Furthermore, the modifier "unmarked" combined with "sending unwanted migrants" provides textual evidence of a "violation of policy in practice"—specifically, the use of surreptitious, non-standard police methods to bypass formal deportation or transfer protocols.
Round 1 (TA2):
Question Generator:
Clarifying Question 1 : *Does the post describe an interaction between nations regarding migration policy?*
Answer Generator:
Answer: CANNOT ANSWER
Rationale: The text mentions the "Spanish border" (a location implying interaction) but frames the action as a unilateral move by an individual ("he is sending"). It is unclear if this counts as a "nation-to-nation" interaction or a personal violation. The context is insufficient to answer "Yes" or "No" to such a broad policy question.
Round 2 (TA2):
Question Generator:
Clarifying Question 2 (Refined): *Does the text explicitly mention a diplomatic agreement, a treaty, or the official reputation of one state in the eyes of another?*
Answer Generator:
Answer: No
Rationale: : The text focuses entirely on the *hypocrisy* of the individual sender ("he") and the specific method of transport ("unmarked police cars"). There is no mention of a treaty being violated, a diplomatic agreement being cited, or how the Spanish state views this action.
Final Investigator Decision:
Crime & Punishment: Applicable
External Regulation & Reputation: Not Applicable

Figure 12: Example of an introspective question–answering loop illustrating a case in which one TA is determined to be applicable, while another is determined to be not applicable.

Candidate Topic Aspect 1: **Child Protection**

Debate: Round1

Champion: The post explicitly mentions a "7-Year-Old." This confirms the subject is a minor. This is distinct from "Health & Safety". "Health" frames the issue as a medical failure (e.g., lack of medicine). "Child Protection" frames the issue as a violation of the special status of children. The grievance is that a child was detained *at all*, not just that they got sick. This is a unique topic aspect not covered by reference topics aspects.

Critic: I object. This is Redundant. The text describes a death. That is a biological even. We already have "Health & Safety" for the death and "Legality" for the detention. The proposal adds no new semantic information. A "7-year-old" is a human; their rights are covered under standard Human Rights frames. Creating a new topic aspect for every demographic (Children, Elderly, Women) creates infinite, useless micro-categories.

Debate: Round 2

Champion: You are ignoring the specialized vocabulary of the text. The user chose to highlight the age ("7-Year-Old") to invoke **Child-Specific Rights**, not general human rights. If we classify this merely as "Health," we lose the "Family Separation" context implied by the custody status. The new topic aspect is necessary to capture this specific political grievance. I stand by "Child Protection."

Critic: "Implied" context is not evidence. The text does not say "Family Separation." It says, "Dies in Custody." That is "Health" + "Security." Your proposal "Child Protection" is technically applicable but **taxonomically unnecessary**. It blurs the lines of our clear existing definitions. I recommend strict Rejection.

Debate Termination

Judge: Champion's point correctly identified that "Health" fails to capture the moral weight of detaining a minor. The "Child-Specific Rights" argument is valid. Critic's point **c**orrectly identified that "Child Protection" is too generic and risks overlap. The Champion is right about the *content*, but the Critic is right about the *structure*.

Verdict: CORRECT

Final Topic Aspect: Family Separation & Child Harm

Definition: Rhetoric and practices that normalize or justify separating children from caregivers or otherwise harming migrant children
(deterrence, opaque custody, exploitation), sidelining child-specific rights and welfare.

Post: 7-Year-Old Migrant Girl Dies in US Border Patrol Custody
Reference Topic Aspects: Legality, Constitutionality & Jurisdiction, Crime & Punishment, Health and Safety, Security and Defense
Proposed Candidates: Child Protection, Government Accountability & Transparency

Candidate Topic Aspect 2**: Government Accountability & Transparency**

Debate: Round1

Champion: : The text does not just say "A girl died." It explicitly specifies *"US Border Patrol Custody."* By attributing the death to a specific state agency, the user is utilizing a **Blame Attribution** frame. This is a core component of "Government Accountability"—identifying institutional failure. This is distinct from "Security". "Security" describes the border infrastructure. "Accountability" describes the *failure* of that infrastructure and the implicit demand for answers. Reference taxonomy contains no topic aspect that captures this institutional critique.

Critic: I object. This is Hallucination and Redundancy. You are confusing a *fact* with a *frame*. The mention of "Border Patrol" is a necessary factual detail of the event, not a political demand. The text contains **zero accountability markers**: no "investigate," no "resign," no "why." The event is fully covered by the intersection of "Health & Safety" (the death) and "Legality" (the custody status). Adding "Accountability" based on the mere mention of an agency name would force us to label *every* government press release as an "Accountability" problem. That is catastrophic taxonomy bloat.

Debate Termination

Judge: Champion's argument relies entirely on the *implication* that naming an agency equals blaming them. Critic's argument relies on strict evidence. The text lacks the specific lexical markers required to constitute a "Demand for Accountability."

Verdict: REJECT

Figure 13: Example of a Multi-Agent Debate.

a clarifying Yes/No question probing whether the post describes an interaction between nations regarding migration policy. In the first round, the Answer Generator Agent is unable to provide a definitive answer and explicitly requests that the question be refined. Upon reformulation, the Answer Generator Agent responds *No*. After inspecting the accumulated Question–Answer pairs and their accompanying rationales, the Investigator Agent concludes that TA2 lacks sufficient grounding in the post and marks it as not applicable.

E Example of Multi-Agent Debate in ID-TA

In Figure 13, the post has been associated with known TAs Health & Safety, Legality, Constitutionality & Jurisdiction, Security & Defense, and Crime & Punishment by the Investigator. The Proposer then introduces two new candidate TAs for the same post: *Child Protection* and *Government Accountability*. For the first TA_{cand}, *Child Protection*, the Champion successfully argues for its necessity by highlighting that *child-specific rights* are different from general human rights. Additionally, while Health & Safety captures the biological fact of death, it fails to encode the moral and rights-based salience introduced by the subject being a minor. The Champion emphasizes that the explicit mention of a "7-year-old" invokes a category of vulnerability and protection that is not adequately subsumed by Health & Safety or Security & Defense. In response, the Judge accepts the Champion's core insight (Child Rights are distinct from Health) but accepts the Critic's argument that the label "Child Protection" is too broad. Conse-

quently, the Judge *CORRECTS* the input TA to a new, robust TA, *Family Separation & Child Harm* explicitly capturing the nuance argued in the debate. In contrast, for the second TA_{cand}, *Government Accountability*, the Critic successfully attacks the proposal by noting that the text mentions "Border Patrol" only as a factual location of the event. The Critic argues that the text lacks lexical accountability markers (e.g., "resign," "investigate"), and creating a new label based solely on an agency's name creates taxonomic bloat. The Judge concurs with this assessment, concluding that the proposal rests on overreaching assumptions rather than evidence, and issues a *REJECT* verdict. This example demonstrates how the debate framework both enables the refinement of genuinely novel TAs and prevents the proliferation of redundant ones.

F Training details of the Articulation Reward Model in Art-FRAME

F.1 Data Collection Strategy

Since no off-the-shelf dataset exists for *valid* vs. *invalid* framing, we constructed a dataset consisting of 1,165 valid CFs that are clearly articulated and 1000 invalid CFs that are poorly articulated. The valid CFs are taken from existing datasets on framings (Weinzierl and Harabagiu, 2025; Ailneni and Harabagiu, 2025), out of which 83 are based on COVID-19 Vaccine hesitancy, 249 are based on Immigration, 75 are based on Abortion, 50 are based on Climate Change and 708 CFs are based on Misogyny. We synthetically generated invalid CFs to mimic common failure modes of zero-shot LLMs such as: 1) Statements describing the post rather than the topic aspect such as *"The user uses a metaphor to describe the border situation."*, 2) Incomplete statements such as *Rising crime rates and lack of police funding.*, etc.

F.2 Training Configuration

We fine-tuned the DeBERTa-v3-large (He et al., 2021) model using binary cross-entropy loss. Training was conducted on a single NVIDIA T4 GPU for 3 epochs to prevent overfitting, utilizing a batch size of 16, a maximum sequence length of 128 tokens, and a learning rate of 2e-5 with linear decay. On a held-out validation set of 200 pairs, the model achieved an F1-score of 0.94, demonstrating robust capability in distinguishing between valid and invalid CFs.

G Additional Qualitative Results

G.1 Quality of the discovered TAs

Method	DC	DNA	PC	SD
Climate Change (CC)				
ID-TA	**4.21**	**4.18**	**0.91**	**0.90**
TAD_{HAC}	3.15	3.40	0.48	0.68
ZS-Prompt	3.76	3.98	0.60	0.47
Abortion (AB)				
ID-TA	**4.39**	**4.67**	**0.89**	**0.93**
TAD_{HAC}	3.35	3.60	0.52	0.63
ZS-Prompt	3.86	4.10	0.65	0.60
Vaccine Hesitancy (VH)				
ID-TA	**4.45**	**4.58**	**0.92**	**0.92**
TAD_{HAC}	3.30	3.55	0.41	0.70
ZS-Prompt	3.92	4.08	0.69	0.72

Table 14: Qualitative results for the discovery of New TAs for the topic of Climate Change (CC), Abortion (AB), and Vaccine Hesitancy (VH).

Qualitative results for the discovery of New TAs for the topic of Climate Change (CC), Abortion (AB), and Vaccine Hesitancy (VH) are summarized in Table 14. Across all three topics, ID-TA consistently outperforms both baselines, achieving higher scores in Definition Clarity (DC) and Definition–Name Alignment (DNA). These results indicate that the newly discovered TAs produced by ID-TA are clearly articulated and remain semantically faithful to their assigned names. In contrast, both TAD_{HAC} and $ZS-Prompt$ exhibit noticeably lower DC and DNA scores, suggesting definitions that are either less precise and less aligned with the name.

The advantages of ID-TA are even more pronounced for Posting Coverage (PC) and Semantic Distinctiveness (SD). Across all three topics, ID-TA yields PC and SD values close to 0.9, demonstrating that postings grouped under each new TA genuinely address that TA and that the discovered TAs are meaningfully distinct from one another.

From a topic-wise perspective, Vaccine Hesitancy stands out for strong overall definition quality and high post coverage. Abortion, in contrast, achieves the highest scores in definition–name alignment and semantic distinctiveness, indicating particularly well-defined and non-overlapping Topic Aspects.

G.2 Quality of the articulated CFs

Qualitative results for the articulation of Communication Frames (CFs) across topics Immigration

(IMM), Climate Change (CC), Abortion (AB), and Vaccine Hesitancy (VH) are summarized in Table 15.

Method	Soundness (Z)	Clarity (A)
Immigration (IMM)		
ID-TA	**0.98**	**0.96**
TAD_{HAC}	0.70	0.73
ZS-Prompt	0.81	0.83
Climate Change (CC)		
ID-TA	**0.97**	**0.95**
TAD_{HAC}	0.71	0.75
ZS-Prompt	0.79	0.79
Abortion (AB)		
ID-TA	**0.99**	**0.97**
TAD_{HAC}	0.72	0.74
ZS-Prompt	0.84	0.87
Vaccine Hesitancy (VH)		
ID-TA	**0.98**	**0.98**
TAD_{HAC}	0.69	0.73
ZS-Prompt	0.82	0.83

Table 15: Qualitative results for the articulation of Communication Frames (CFs) across topics.

Qualitative results for the articulation of Communication Frames (CFs) across topics Immigration (IMM), Climate Change (CC), Abortion (AB), and Vaccine Hesitancy (VH) are summarized in Table 15. To rigorously assess the generated frames, we engaged human linguists to evaluate two critical dimensions: **Soundness** (Z) and **Clarity** (A).
As shown in Table 15, our **ID-TA** framework achieves superior performance across all domains, consistently scoring above 0.97 in Soundness and 0.95 in Clarity. This demonstrates that identifying high-quality TAs is a prerequisite for articulating logically sound CFs; because ID-TA provides a robust foundation, the resulting frames are both coherent and grounded. In contrast, the baselines struggle to match this fidelity. The *ZS-Prompt* method produces reasonably fluent text (Clarity ≈ 0.83) but often suffers from lower Soundness, reflecting a tendency for zero-shot models to generate plausible-sounding but generic interpretations. The TAD_{HAC} baseline performs the worst across all metrics (Soundness ≈ 0.70), indicating that clustering-based approaches often produce noisy TAs that are difficult to articulate into meaningful frames.

H Inter Annotation Agreement

To ensure the reliability of our human evaluation, we calculated the Inter-Annotator Agreement (IAA) for both the qualitative evaluation of discovered Topic Aspects (TAs) and the quality assessment of articulated Communication Frames (CFs).

H.1 IAA for Discovered Topic Aspects

The inter-annotator agreement for TA discovery was calculated on the complete set of 43 novel TAs discovered by **ID-TA** across the four topics: 13 for Immigration, 10 for Climate Change, 11 for Abortion, and 9 for Vaccine Hesitancy.
Three graduate-level annotators independently evaluated these TAs across four criteria. We report Krippendorff's α (Krippendorff, 2011) to account for the different measurement scales:
(a) Definition Clarity (DC): Evaluated on a 5-point Likert scale. We observed a Krippendorff's α of **0.83**, indicating substantial agreement on the fluency and comprehensibility of the definitions.
(b) Definition-Name Alignment (DNA): Evaluated on a 5-point Likert scale. The agreement score was **0.85**, showing strong consensus on how well the definitions matched the generated names.
(c) Post Coverage (PC): Evaluated using a binary decision (Yes/No) on 5 randomly sampled posts per TA. We obtained a Krippendorff's α of **0.88**, demonstrating high reliability in grounding the TAs to source text.
(d) Semantic Distinctiveness (SD): Evaluated on a 3-point ordinal scale (Duplicate, Overlap, Distinct) for the top-2 most similar pairs. The agreement score was **0.81**, confirming that annotators consistently distinguished unique topics from redundant ones.

H.2 IAA for Articulated Communication Frames

For evaluating CFs, we randomly sampled 100 frames from the set of valid CFs generated by **Art-FRAME**, with samples drawn proportionally from each topic.
Two linguists assessed this sample for *Soundness* and *Clarity*. We report **Cohen's** κ for this binary classification task. Soundness measures whether the rationale underlying a CF is logically coherent and well grounded, yielding a Cohen's κ of **0.81**, while clarity assesses whether the CF is expressed fluently and unambiguously, with a Cohen's κ of **0.87**. Together, these scores indicate strong annotator agreement, reflecting a high level of consensus on both the logical validity and linguistic quality of the frames generated by Art-FRAME.

I Additional Analysis

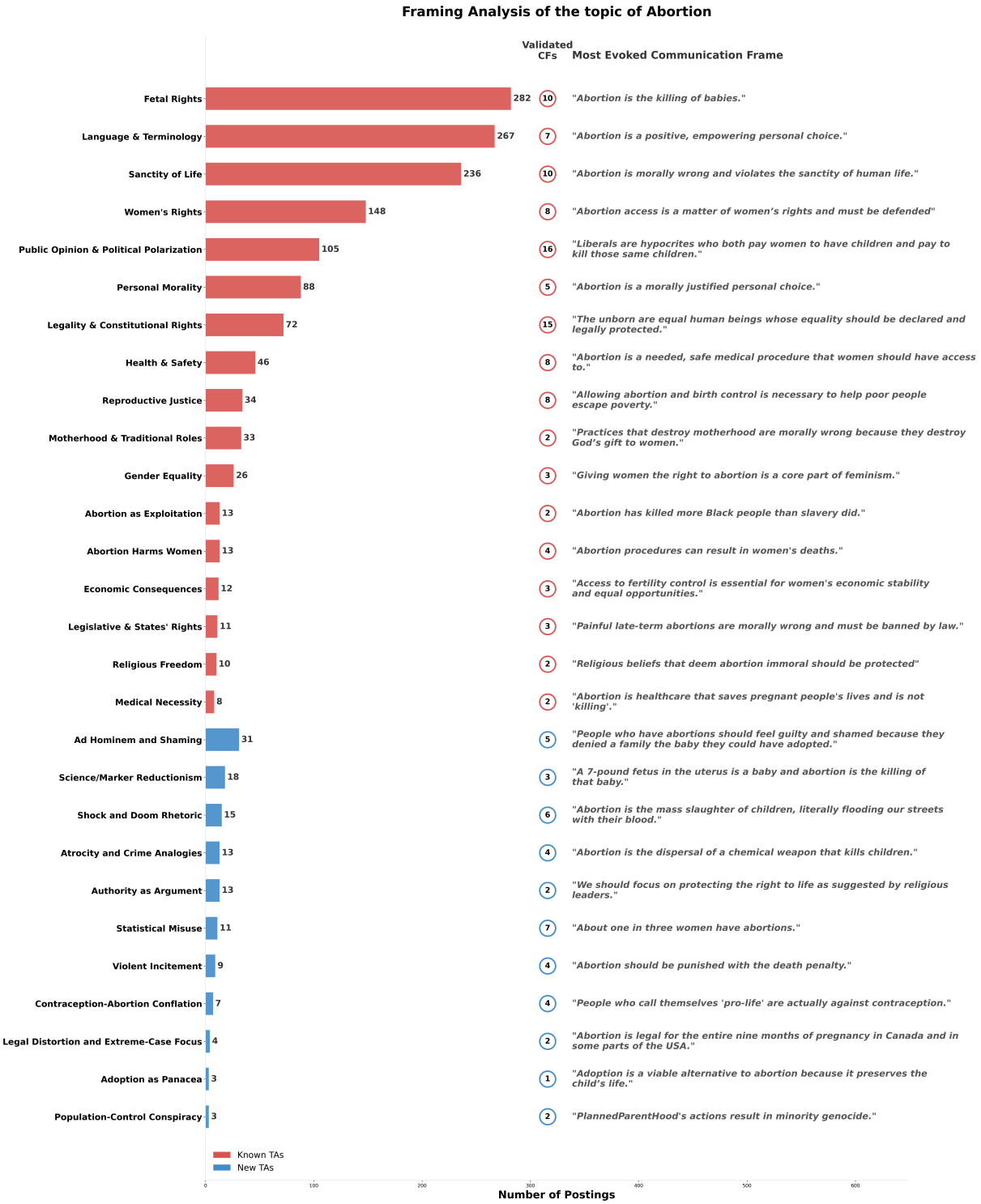

Figure 14: Framing Analysis of the topic of Abortion.

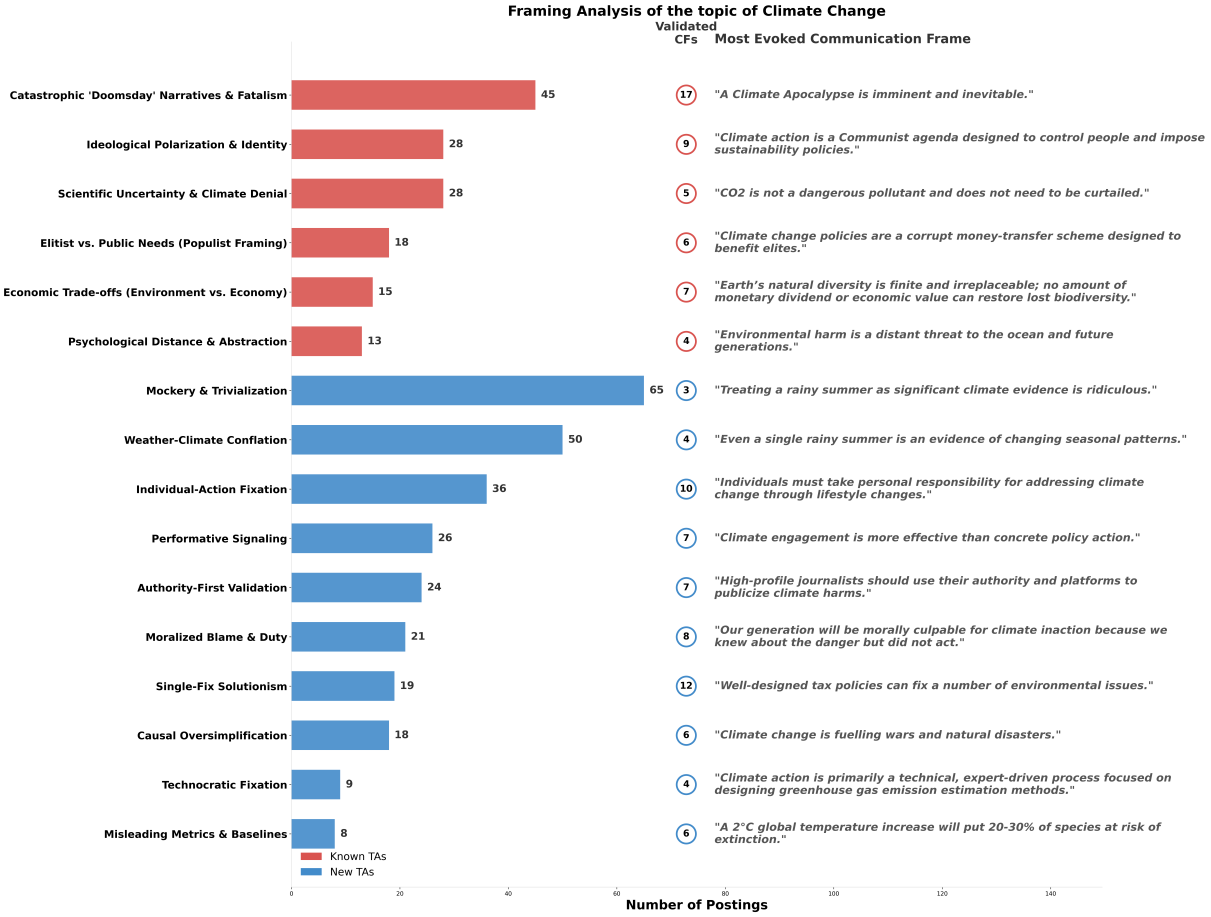

Figure 15: Framing Analysis of the topic of Climate Change.

Figure 14 illustrates the framing analysis for the Abortion topic. The most addressed new TA is *Ad Hominem and Shaming*, which appears in a substantial number of postings, indicating that moral judgment and social condemnation form a prominent dimension of abortion discourse that extends beyond the existing TA set. In contrast, *Statistical Misuse* is the new TA associated with the largest number of unique CFs, despite appearing in fewer postings overall. This pattern suggests that numerical claims and statistics are leveraged through a wide range of framing strategies.

Figure 15 illustrates the framing analysis for the Climate Change topic, showing (a) the number of postings that address both the known and the new TAs; (b) the number of CFs articulated for each TA; and (c) examples of the most evoked frame for each TA. For this topic, the most addressed new TA is *Mockery and Trivialization*, which exceeds several known TAs in posting frequency, indicating that dismissive and minimizing rhetoric forms a substantial component of climate discourse that is not captured by existing TAs. In contrast, the new TA associated with the largest number of unique CFs is *Single-Fix Solutionism*. Although it appears in fewer postings than *Mockery and Trivialization*, it elicits a richer set of articulated CFs, reflecting diverse narratives that promote narrowly scoped or overly simplistic solutions to climate change.

Finally, the framing analysis for Vaccine Hesitancy, shown in Figure 16, reveals *Authority-as-Evidence* as the most addressed new TA, indicating frequent reliance on expert credentials or selective appeals to scientific authority that fall outside the scope of existing TAs. In addition, *Mechanistic Misinformation* emerges as the new TA with the largest number of distinct CFs.

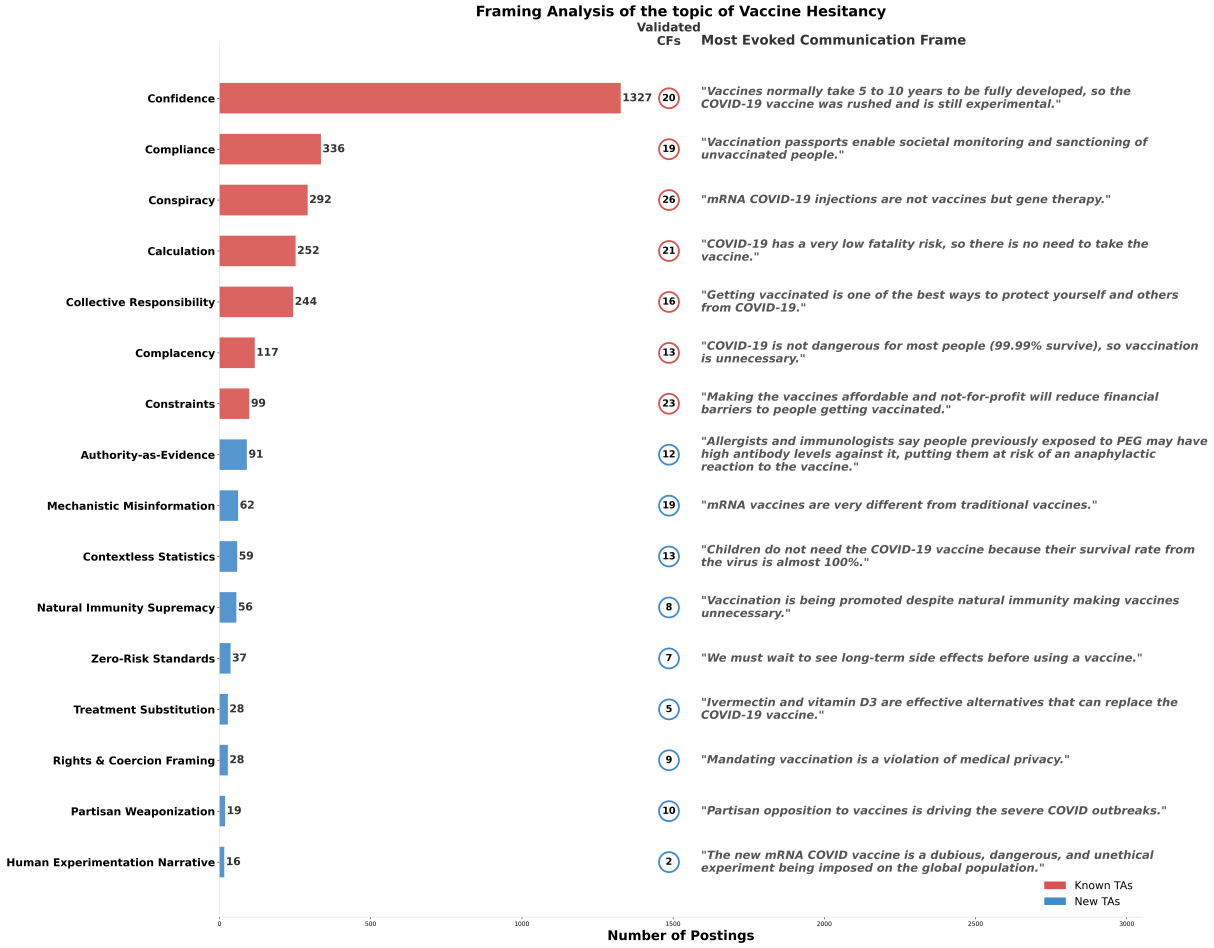

Figure 16: Framing Analysis of the topic of Vaccine Hesitancy.