CCAGENT: Learning Constructive Consensus for Multi-Agent LLMs in Real-World Environments

Anonymous ACL submission

Abstract

Real-world decision-making often involves complex deliberation among diverse stakeholders with conflicting values. However, existing LLM-based multi-agent frameworks struggle with two key challenges: (1) they lack real-world grounding, relying on synthetic tasks that fail to capture the complexity of real-world decision making, and (2) they are difficult to supervise effectively, since desirable behaviors like principled compromise, quality discussion, and open-mindedness are abstract and hard to quantify. We address both challenges with CCAGENT, a framework for training deliberative agents using contrastive supervision over natural language rationales and counterfactuals. First, we introduce two decision-making datasets grounded in real-world sources: city planning stakeholder interviews and U.S. Senator interviews and voting patterns. Second, we propose nine training objectives that reinforces socially aligned behaviors—such as consensus, compromise, and low dogmatism—without requiring scalar rewards or human preference labels. We also propose eight strategies for efficient multi-agent debate. Lastly, we introduce CCAGENT, a few-shot lightweight, automatic Direct Preference Optimization (DPO) method for efficient multi-agent debate. CCAGENT outperforms baselines achieving faster consensus with high quality discussions between agents. Our results demonstrate that DPO enables principled deliberation even in complex, disagreement-rich domains.

1 Introduction

Multi-agent large language model (LLM) systems have shown strong capabilities in reasoning, planning, and decision-making (Du et al., 2024; Liang et al., 2024b). Yet, most prior work remains confined to benchmark datasets or game-like environments (He et al., 2023; Du et al., 2023; Lin et al., 2023). These environments lack the complexity of real-world decision-making. In practice, domains like city planning or climate policy involve diverse stakeholders, conflicting values, partial information, and resource trade-offs, making consensus both necessary and difficult (Ni et al., 2024; Innes and Booher, 2000; Lindblom, 1959).

We argue that modeling such real-world deliberation is not only crucial for high-stakes applications, but also a stepping stone toward artificial general intelligence. Reaching consensus in open-ended, multi-agent environments requires adaptive reasoning, value alignment, and socially intelligent behavior, all core challenges in building general-purpose AI. This, in turn, also helps real-world decision-making with AI assistance.

While recent work explores multi-agent deliberation through voting (He et al., 2023), reflection (Li et al., 2024), or argumentation-style dialogue (Ruggeri et al., 2023; Du et al., 2023), most efforts optimize for outcome-level metrics like accuracy or faster convergence. These overlook process-level qualities: whether agents reason constructively, persuade others, and have a healthy discussion, or simply echo the majority. Single-agent self-debates (Yao et al., 2023; Creswell et al., 2022) lack diverse perspectives, while adversarial setups (Rescala et al., 2024; Zhu et al., 2023) focus narrowly on correctness. As a result, failure modes such as sycophancy (obsequious behavior of agents (Sharma et al., 2025), dogmatism, or shallow agreement remain underexplored.

To tackle both of these limitations, the lack of real-world deliberation datasets and the absence of process-level evaluation in existing frameworks, we introduce a new framework to evaluate and train multi-agent LLMs that deliberate more effectively in complex, high-stakes settings. Our contributions include the following:

- **Datasets.** We release two real-world multi-agent decision-making datasets: one from interviews with city planners, and another derived from U.S.

Senate voting records.

- **Reusable Strategies.** We develop debate strategies (moderation, nudging, alliances, etc.) that generalize across domains and improve convergence and reasoning quality.
- **Deliberation Metrics.** We propose novel process-level metrics (sycophancy, dogmatism, vote switching, and semantic convergence, etc.) to capture debate quality beyond final outcomes.
- **CCAGENT Method.** We introduce CCAGENT, a few-shot reinforcement-based Direct Preference Optimization (DPO) approach that trains agents to improve on deliberative behavior.

To our knowledge, this is the first work to introduce metrics for the quality of multi-agent LLM debate for real-world consensus. This is also the first work to introduce a lightweight, automatic few-shot DPO method for effective and efficient multi-agent LLM debates. Results show that agents trained with this framework reach consensus faster, switch votes more meaningfully, and reduce undesirable behaviors like blind conformity accross multiple types and numbers of agents and datasets.

2 Related Work

Multi-Agent LLM Simulations Recent work has explored multi-agent simulations for reasoning and coordination. Broadly, this research falls into three directions: (1) structured workflow agents for tasks like software engineering and SOP writing (Hong et al., 2024; Xu et al., 2024), (2) strategic reasoning using game-theoretic setups like negotiation or Nash equilibria (Mao et al., 2025; Hua et al., 2024), and (3) social simulations in narrative or gameplay environments such as Werewolf or Avalon (Lin et al., 2023; Du et al., 2023; Zhang et al., 2024; Park et al., 2023b). Other work has extended multi-agent simulation to domains like psychology (Arzani et al., 2022), finance (Qian et al., 2024), and policy-making (Wang et al., 2024).

However, these systems often lack the complexities of real-world deliberation, including partial information, conflicting stakeholder values, and societal trade-offs. Our work addresses this gap by introducing natural constraints (e.g., limited resources and diverse agent perspectives) and focusing on structured debate settings, a common mode of public reasoning in domains like politics and urban planning.

Consensus and Compromise in LLMs Consensus has been used as a proxy for correctness in many multi-agent frameworks (Lee et al., 2024; Wu et al., 2023), but it is often unsuitable in real-world domains where value disagreements persist. Arrow's impossibility theorem shows that no voting system can perfectly reconcile all fairness criteria (MASKIN et al., 2014). Consequently, researchers have explored majority voting (Yin et al., 2024) and negotiation-based approaches (Liang et al., 2024a; Chiang et al., 2023) to foster agreement.

We extend this line of work by modeling compromise in realistic decision settings and introducing it as a first-class metric of deliberation success, especially when consensus cannot be easily achieved. Furthermore, these works focus on fast consensus and accurate consensus as a sole metric- without taking into consideration the quality of the debate. We analyze the quality of the underlying discussion with our nine metrics.

Reinforcement Learning for Deliberative Behavior Reinforcement learning has been used to align LLM outputs with desirable behaviors, either via reward models (DeepSeek-AI et al., 2025) or safety constraints (Lindström et al., 2024; Zhan et al., 2025). Recent advances in in-context learning from feedback (e.g., verbal reinforcement (Liang et al., 2022; Scheurer et al., 2023)) and Direct Preference Optimization (DPO) (Rafailov et al., 2024) show promise in shaping LLM behavior without full fine-tuning.

We adapt Direct Preference Optimization (DPO) for few-shot reinforcement learning, enabling lightweight, automatic training without full model fine-tuning. Instead of relying solely on labeled examples or scalar rewards, we automatically generate rationales and counterfactuals as rich supervision signals. These guide agents not just on what to prefer, but why — encouraging behaviors like compromise while discouraging sycophancy and dogmatism. This enables the development of socially aligned agents that learn to deliberate effectively in real-world, multi-agent settings.

Appendix B and Table 7 shows detailed Related Works comparison to other works.

3 Dataset Creation

This section describes the construction of two datasets designed for evaluating real-world multi-agent decision-making: one focused on urban planning in Miami and the other on U.S. Senate-level

political debates. Additional details about this can be found in the Appendix D and Appendix F Algorithm 2.

City Planning We construct a city planning dataset based on 51 interview transcripts with stakeholders from the City of Miami, categorized into four groups: NGO, Private, Public, and Academic (Pathak et al., 2020). This interview data was obtained in accordance with appropriate sharing agreements. Data availability statement can be found in Appendix O. To adapt this data to our debate task, we manually extract question-answer pairs, use LLMs to correct transcription errors, and categorize each interview by issue area. Using these themes, we generate 25 Likert-style debate prompts with five response options, and identify ground-truth agent responses and public preferences for each statement.

Politics We build a political agent dataset using U.S. Senator quotes and stances from the OnTheIssues website. We focus on twelve key policy issues and structure the dataset in two parts: one for ranking issue priorities, and one for multiple-choice question answering. For the latter, we adapt VoteMatch quiz items, linking each to supporting quotes and position summaries as ground truth responses.

Dataset Statistics Demographic statistics of interviewees are shown in Appendix A in Tables 4 and 6, with dataset-wide statistics in Table 8.

Test Question Generation For both datasets, test-time questions were selected to maximize disagreement potential. In City Planning, we used GPT-4-turbo to generate candidate questions and manually selected 50 that triggered substantive divergence. For Politics, we selected 20 VoteMatch questions and added five more generated via disagreement-focused prompting. Public opinion data from Brenan (2024) was used to analyze majority preferences for ground truth for politics, and from Wikipedia contributors (2025) was used to estimate majority preferences for city planning in Miami.

4 Methodology

4.1 Agent Setup

We adopt a few-shot learning setup, following Park et al. (2023b). Unlike their approach, which uses full interview transcripts, we only included five question–answer pairs per agent. To evaluate agents in this setup, we ran two tests:

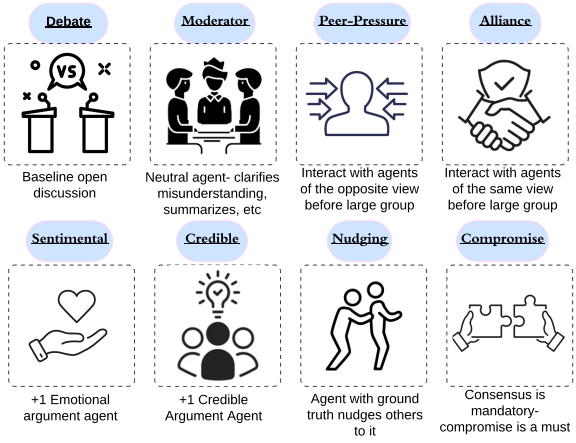

Figure 1: Multi-agent Debate Strategies

Multiple-choice generalization For each agent, we generated ten questions that could be answered directly from the few-shot examples, and ten that required extrapolation to related but unseen topics. Accuracy was measured on both sets.

Semantic similarity to human responses On held-out open-ended questions, we compared agent and human answers using ROUGE (Lin, 2004), cosine similarity (Manning et al., 2008), and BERTScore (Zhang et al., 2020). ROUGE was lower, but both cosine and BERTScore were high, showing that agents captured the intended meaning even when phrasing differed. The results for the sanity test are shown in Table 1.

We experimented with both fine-tuning and few-shot learning to create stakeholder agents, before settling on few shot training our agents. Detailed experimental setup for this and results are shown in Appendix L and Table 12.

4.2 Multi-Agent Debating Strategies

One of our core contributions is proposing strategies that make multi-agent debate more effective and efficient. We run a series of experiments to evaluate how different interaction setups influence consensus-building among agents. These strategies proposed are: simple debate between agents with 3 different prompts (debate, discuss, and reach consensus), debate with moderator, debate with nudging agent, debate with credible agent, debate with sentimental agent, alliance before larger debate, peer pressure before larger debate, and mandatory consensus with compromise/conditional accep-

Agent	Cosine Sim.	ROUGE-1	ROUGE-2	ROUGE-L	BERTScore F1	MCQ Acc.
City Planning Dataset						
academic	0.60	0.22	0.02	0.11	**0.80**	0.95
public	0.67	0.18	0.03	0.11	**0.85**	1
private	0.65	0.12	0.02	0.09	**0.85**	1
ngo	0.59	0.18	0.02	0.10	**0.83**	1
Politics Dataset						
tim_scott	0.69	0.36	0.06	0.18	**0.86**	1
bill_cassidy	0.84	0.40	0.17	0.29	**0.91**	1
tammy_baldwin	0.78	0.32	0.11	0.21	**0.89**	0.90
bernie_sanders	0.76	0.30	0.13	0.20	**0.88**	1
kyrsten_sinema	0.63	0.22	0.03	0.13	**0.86**	1

Table 1: Agent Sanity Check: Text similarity and quality metrics across agents on city planning and politics datasets.

Strategy	Consensus Round (↓)	Majority Round (↓)	Vote Switches (↓)	Agreement % (↑)	Compromise % (↑)	Avg Cosine Sim (↑)	Dogmatic % (↓)	Sycophancy % (↓)	GT Match (↑)
City Planning Dataset									
Alliance	3.48	2.12	1.32	82.00	2.25	**0.90**	13.00	**2.08**	0.44
Compromise	**3.20**	2.32	2.16	**88.00**	**9.00**	0.88	**5.00**	6.33	0.52
Credible Agent	3.72	2.32	1.60	81.00	8.00	0.89	18.00	6.17	0.48
Debate	3.49	2.07	1.09	81.16	4.37	**0.90**	17.46	2.75	0.41
Debate-Consensus	3.44	**2.07**	**1.08**	81.51	4.28	**0.90**	17.46	2.75	0.42
Debate-Discuss	3.38	**2.07**	**1.04**	81.85	4.20	**0.90**	17.46	2.75	0.44
Moderator	3.52	2.24	1.76	83.00	6.00	0.88	14.00	3.60	**0.56**
Nudger	**3.20**	**1.96**	1.56	86.00	6.00	0.89	13.00	4.90	0.44
Opposites	3.36	2.40	1.20	86.00	4.50	**0.91**	11.00	2.58	0.48
Sentimental	4.16	2.52	1.96	73.00	7.75	0.88	23.00	4.90	0.52
Politics Dataset									
Alliance	4.83	3.12	1.64	64.00	3.00	0.85	57.60	1.28	**0.46**
Debate - consensus	5.00	2.86	1.33	60.71	1.90	0.84	69.52	**0.76**	0.42
Debate	5.00	2.90	**0.71**	61.90	**0.24**	0.86	75.24	0.38	**0.52**
Debate- discuss	5.00	2.90	2.05	58.33	3.57	0.84	61.90	1.90	**0.46**
Moderator	**3.05**	**2.10**	5.38	**92.3**	**22.86**	0.75	**4.76**	4.14	**0.59**
Nudger	5.00	2.95	3.24	58.33	3.57	0.85	60.95	1.52	0.42
Opposites	4.81	3.65	2.10	59.52	3.81	0.84	60.00	1.33	**0.47**
Consensus	4.04	2.48	3.44	89.00	13.40	0.84	20.00	5.45	0.41
Sentimental	5.00	2.71	2.29	59.52	3.81	0.86	59.05	1.90	0.38
Credible	4.89	2.67	2.26	62.00	3.60	**0.88**	21.00	1.93	0.38

Table 2: Comparison of multi-agent strategies on consensus dynamics, agreement quality, and behavioral alignment across City Planning and Politics datasets. Arrows in headers indicate whether higher (↑) or lower (↓) is better. Bolded values represent the best performance for each metric in that dataset.

tances. Each of these are described in Figure 1 and in detail in Appendix C Table 8.

These experiments are inspired by Cialdini and Goldstein (2004) who describe various social phenomenon in the context of human debate. The goal of this setup is to simulate deliberation and observe which strategies encourage alignment and healthy discussion. We use the dataset created in Section 3 to simulate agent discussion under the eight constraints as shown in Figure 1. Initially, all agents present their opinions, which is shared with other agents along with explanations in subsequent rounds. The debate ends when either consensus is reached or 5 rounds have elapsed.

Another core contribution of our paper is a series of metrics to evaluate agent debate. We evaluate each debate using a set of outcome, behavior, and explanation-based metrics as shown in Appendix K Table 11. These metrics are: number of rounds to reach consensus, number of rounds to reach majority vote, vote switches, agreement %, compromise %, average cosine similarity between first rond response and last round response, sycophancy %, dogmatism %, and Ground Truth match.These are inspired by group-phenomenon observed in several NLP and human psychology papers (Sharma et al., 2025; Fast and Horvitz, 2016; Spang, 2023; Nitzan, 2010). These metrics together capture how agents respond to peer influence, whether they converge meaningfully, and how aligned their final stance is with external standards.

4.3 Findings of the Preliminary Study

Table 2 demonstrates two core findings: (1) strategies based on *compromise*, *moderation* and *nudging* achieved the strongest overall performance across multiple metrics. (2) We see that GT match, cosine similarity are similar accross the board (for good and bad debates) and hence might not be the best metrics for judging what makes a debate good. Hence, we consider a good debate to have the following characteristics: (i) fewer rounds (for consensus and majority), (ii) lower rates of sycophancy and dogmatism, and (iii) high compromise percentage. (3) There is no single method that significantly outperformed others in either datasets. In fact, multiple methods had their benefits- for example, for city planning, compromise had a good number of rounds for consensus, while standard debating had a good majority round, while discussing with opposite agents had a good principle consistency (avg. cosine similarity of R1 and last round). This motivates our reinforcement learning algorithm where we reinforce certain characteristics through examples, as opposed to saying one method of prompting is better than others. Additional findings from these experiments are discussed in Section 6 and Appendix E.

4.4 CCAGENT

We propose **CCAGENT**, a lightweight and fully automatic method for improving the quality of multi-agent debates (Figure 2, Appendix F Algorithm 1). Inspired by Direct Preference Optimization (DPO) (Rafailov et al., 2024), CCAGENT adapts the core idea to multi-agent interactions without relying on human-labeled preferences. Instead, it leverages behavioral metrics and model-generated rationales to supervise agent behavior directly from past debates.

Agent-Level Debate Data Extraction We begin by collecting agent-level data from prior experiments described in Section 4.2. Unlike many existing pipelines that operate at the round level, we extract features for each individual agent, since our goal is to optimize agent-specific behavior. For every agent in a debate, we record: (1) their voting history accross rounds with explanation, (2) number of vote changes, (3) percentage of sycoophantic and dogmatic instances.

In addition to structured features, we generate templated natural language summaries (e.g., *"Agent initially disagreed but shifted after discussion"*) to provide richer context and improve generalization during training.

Constrastive Examples Extraction To construct contrastive training examples, we rank all agents using four behavioral metrics: compromise, sycophancy, dogmatism, and the number of rounds to reach consensus or majority. This is the only step that requires manual predefined preferences, similar to reward shaping in reinforcement learning.

Based on this ranking, we automatically identify the eight best and eight worst-performing agents. Agents that reach consensus more quickly and exhibit higher compromise are prioritized, with sycophancy and dogmatism used to break ties. These pairs serve as positive and negative training examples, representing desirable and undesirable debate trajectories.

This is the standard Direct Preference Optimization (DPO) objective, which optimizes a model to prefer a chosen response y^+ over a rejected response y^- for a given input x:

$$\mathcal{L}_{\text{DPO}} = -\log \frac{\exp(\beta f_\theta(x, y^+))}{\exp(\beta f_\theta(x, y^+)) + \exp(\beta f_\theta(x, y^-))} \quad (1)$$

Here, f_θ denotes the model's log-likelihood scoring function, and β is a temperature hyperparameter controlling the contrast sharpness.

Rationale and Counterfactual Generation Since behavioral metrics alone may not provide sufficient supervision—especially when best and worst examples share similar statistics—we supplement each contrastive pair with a **rationale** and a **counterfactual**.

While prior work has shown that natural language rationales can improve preference modeling (Wang et al., 2023), we find that rationales alone are often insufficient for supervising subtle differences in agent behavior. In addition to a **rationale**—a short explanation for why an agent's behavior resulted in a strong or weak debate outcome, we introduce **counterfactuals** as an additional supervision signal. Counterfactuals describe what the agent could've done differently to get the opposite label.

Mathematically, we propose the **CCAGENT objective**, which augments the DPO loss with auxiliary terms that supervise agents via behavior-aware rationales and counterfactuals:

$$\mathcal{L}_{\text{CCAGENT}} = \mathcal{L}_{\text{DPO}} + \lambda_1 \cdot \mathcal{L}_{\text{rationale}} + \lambda_2 \cdot \mathcal{L}_{\text{cf}} \quad (2)$$

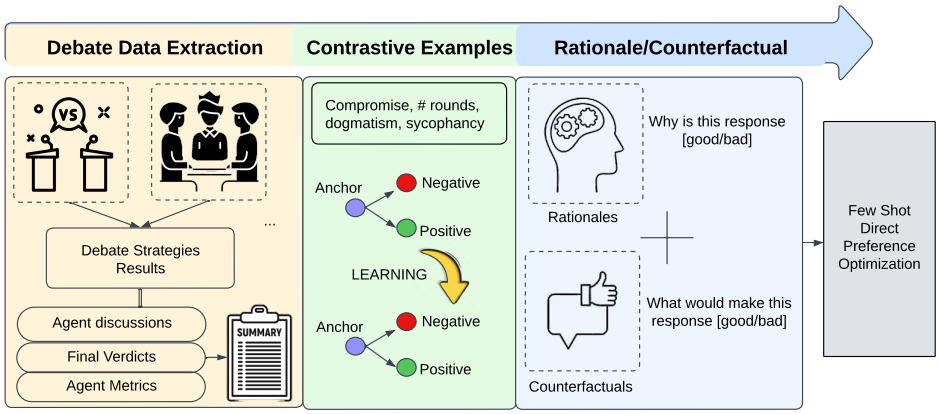

Figure 2: CCAGENT Methodology: We start with Extraction of data from our experiments, where we get discussions, verdicts, summary and metrics on a per-agent basis. Then we rank this using compromise, number of rounds, sycophancy and dogmatism to get best and worst contrastive examples. Lastly, we generate counterfactuals and rationales for these examples, and combine it into a few shot DPO prompt.

The term $\mathcal{L}_{\text{rationale}}$ penalizes deviations from model-generated explanations that justify effective debate behavior, while $\mathcal{L}_{\text{cf}}$ measures divergence from hypothetical improvements suggested in counterfactuals.

We further define these components as:

$$\mathcal{L}_{\text{rationale}} = \mathbb{E}_{(x,y)} \left[\text{BCE}(\text{pred}_{\text{rationale}}, \text{target}_{\text{rationale}})\right] \tag{3}$$

$$\mathcal{L}_{\text{cf}} = \mathbb{E}_{(x,y)} \left[1 - \cos\left(y, \text{cf}_{\text{suggested}}\right)\right] \tag{4}$$

Here, BCE is the binary cross-entropy loss over rationale validity, and cos measures the semantic similarity between agent output and its counterfactual improvement suggestion.

By incorporating these behavior-guided supervision signals, CCAGENT learns to favor not just persuasive responses but those that promote compromise, reduce sycophancy, and foster principled agreement.

Few-shot Optimization via In-Context Learning The final few-shot prompt consists of a statement, debate summary,label, the rounds of discussion (from that agent's perspective), rationale and counterfactual.

Each agent is given these few-shot examples and asked to participate in a new debate. The goal is for the agent to implicitly learn the behaviors that lead to effective consensus—shifting when appropriate, avoiding undesired behaviors, and reasoning toward resolution.

We experimented with two alternative training setups: (1) reinforcement learning using standard reward functions, and (2) supervised learning using only numerical behavioral metrics without rationales, or counterfactuals. Results for all of these can be found in Appendix M and Figure 11.

5 Experimental Setup

We evaluate CCAGENT agents on a held-out test set of debate prompts not seen during few-shot DPO training, using the same outcome and behavioral metrics as Section 4.2.

For the Politics domain, we construct a high-difficulty setup by sampling five agents from a public dataset ranking U.S. senators (GovTrack.us, 2024) by ideology, selecting points at intervals (0.0, 0.2, 0.4, 0.6, 0.8) to ensure strong ideological diversity. These agents often disagree sharply on divisive issues like abortion or taxation, making consensus especially challenging.

In contrast, City Planning represents a medium-difficulty case. We sample one agent each from the academic, NGO, public, and private sectors. While these agents share many core values, they differ in priorities and justifications, requiring persuasion through trade-offs rather than ideological conflict.

For baselines, we use Nudging and Compromise for city planning and Moderation and Compromise for politics as they perform the best in our initial test (Table 2. Reconcile (He et al., 2023) is used as a multi-agent baseline. Additional tests are performed on other models to test generalization of our model. We use the few-shot samples generated by GPT-4o and test Llama-70B and Claude on those. Additional implementation details are in Appendix G, Table 9, and Appendix J.

6 Results

Main Results Table 3 compares CCAGENT against baselines across nine metrics. CCAGENT consistently outperforms or matches alternatives. In City Planning, CCAGENT *(temp 1)* leads on six metrics, with the lowest consensus round (1.86), 100% agreement, 0% dogmatism, and highest cosine similarity (0.95)—showing fast, principled convergence.

Similar trends hold in the Politics domain. Notably, this performance is achieved despite the inherently higher polarization, where agents are sampled from opposing ideological extremes and debates center around divisive topics such as abortion or taxation. CCAGENT *(temp 1)* still ranks highest on five metrics, including 96% agreement and 0.93 cosine similarity, showing that conditional compromises and goal-driven discussion help reach consensus even in adversarial settings.

While nudging injects ground truth, CCAGENT builds on compromise and contrastive learning to enable natural, interpretable convergence. It reduces sycophancy, vote switches, and rounds to consensus, improving performance across agent counts, types, and datasets. Appendix H shows examples of a good (Figure 9) and bad (Figure 10) debate. Additional Details on generalization of CCAGENT to other models can be found in Appendix I and Table 10.

Improved Early-Stage Deliberation and Reduced Social Biases CCAGENT (temp 1) achieves 0% dogmatism and low sycophancy across both datasets (Fig. 3), showing that agents neither rigidly hold initial views nor blindly follow the majority. Instead, they adjust positions early—often from Round 1—leading to faster consensus and healthier deliberation. While we supervise only a few metrics (e.g., compromise, sycophancy, dogmatism), agents also display emergent behaviors like conditional agreements, middle-ground proposals, and respectful persuasion. These were not hard-coded in system or user instructions but arose from contrastive training examples and rationale supervision, demonstrating CCAGENT's ability to model constructive debate dynamics.

Faster and More Reliable Consensus Formation Our method enables faster convergence than existing strategies. In both City Planning and Politics, CCAGENT *(temp 1)* achieves the lowest consensus rounds (1.86 and 2.34), outperforming *Nudge*

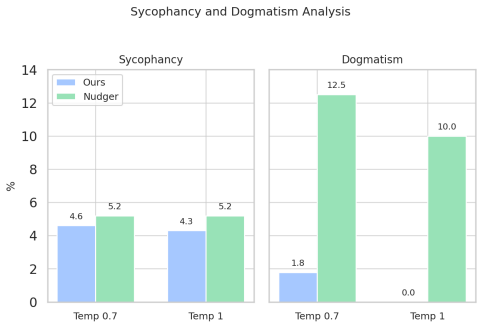

Figure 3: Sycophancy and Dogmatism % for Our method vs a baseline (nudging) which had the lowest. Details and more results in Table 2.

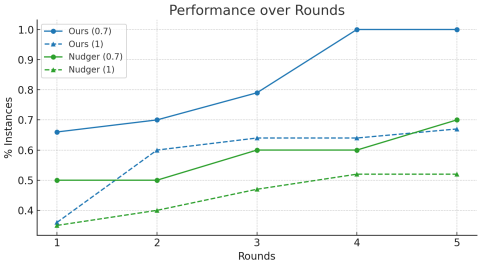

Figure 4: Percentage of Agents that reach Consensus

and *Compromise*. As shown in Fig. 4, most CCAGENT debates reach consensus, while over 25% of Nudging cases fail to do so within 5 rounds.

This reflects agents' ability to recognize persuasion, anticipate counterarguments, and respond constructively—leading to faster alignment without sacrificing diversity. Faster convergence also cuts computational cost, making CCAGENT scalable for real-world multi-agent deployments.

Summary of Debate Observations Compromise, moderation, and nudging yield the strongest results across metrics, with compromise offering the most efficient and unbiased path to consensus. It reduces rounds and vote switches while improving alignment without requiring external signals. Political debates are more polarized and require conditional agreement, while city planning debates are more value-aligned and easier to resolve. Sentimental, alliance-based, and unstructured debates perform poorly due to lack of persuasion or stagnation. CCAGENT takes examples from compromise and moderated debates where middle ground solutions and conditional agreements are prioritized, and hence performs well. See Appendix E and H for full details and examples.

Strategy	Consensus Round (↓)	Majority Round (↓)	Vote Switches (↓)	Agreement % (↑)	Compromise % (↑)	Avg Cosine Sim (↑)	Dogmatic % (↓)	Sycophancy % (↓)	GT Match (↑)
City Planning Dataset									
Nudge	3.00	**1.86**	**1.00**	89.29	3.13	0.90	12.50	4.17	0.36
Nudge (temp 1)	3.05	2.00	**1.00**	*90.30*	4.10	*0.93*	10.10	4.16	0.35
Compromise	3.14	2.43	1.79	89.29	**8.04**	0.90	5.36	4.82	**0.43**
Compromise (temp 1)	3.29	2.57	2.93	83.93	9.82	0.89	7.14	5.71	0.50
ReConcile	3.42	3.19	2.26	83.1	4.68	0.78	10.9	8.17	0.41
CCAGENT	*2.86*	*1.43*	1.79	89.29	7.14	0.92	**1.79**	7.14	**0.43**
CCAGENT (temp 1)	**1.86**	**1.86**	*0.86*	**100.00**	*3.57*	**0.95**	**0.00**	**4.91**	**0.43**
Politics Dataset									
Moderator	3.42	**2.10**	*1.18*	86.70	3.45	0.88	14.00	5.10	0.34
Moderator (temp 1)	3.56	2.26	1.22	*87.50*	4.25	*0.91*	12.20	5.15	0.33
Compromise	3.60	2.58	2.02	86.80	**7.85**	0.87	6.20	5.76	**0.42**
Compromise (temp 1)	3.52	2.55	2.06	86.90	*7.00*	0.85	*5.40*	5.74	0.33
ReConcile	4.76	3.12	2.26	82.9	3.14	0.85	8.90	6.13	0.38
CCAGENT	*3.12*	**1.86**	2.02	86.80	6.95	0.90	*2.45*	8.22	*0.41*
CCAGENT (temp 1)	**2.34**	*2.34*	**1.12**	**96.00**	4.00	**0.93**	**0.00**	**5.91**	*0.41*

Table 3: Final results for City Planning and Politics datasets. Bold values indicate best results, and italicized values show second-best if achieved by CCAGENT. Arrows indicate whether higher (↑) or lower (↓) values are better.

7 Ablation Study

Each component of CCAGENT improves results We conduct an ablation study to assess the contribution of each component of CCAGENT by systematically removing elements from the few-shot prompt and evaluating the effect on performance (Appendix N Table 13). Removing **strong examples** lowers compromise rates and weakens resolution quality, indicating the value of clear demonstrations of effective behavior. Excluding **weak examples** results in increased dogmatism and premature convergence, suggesting that contrastive failures help define boundaries of acceptable reasoning. Omitting **rationales** reduces interpretability and correlates with higher sycophancy, likely due to the lack of structured explanation. The removal of **counterfactuals** leads to the largest performance drop, highlighting their importance in identifying and addressing weaknesses. Overall, the findings show that both positive and negative examples, paired with explanation and counterfactual reasoning, are critical for improving debate quality.

Cross Dataset Training To evaluate generalizability of CCAGENT, we conduct a cross-dataset ablation where agents are trained on City Planning debates and tested on held-out Politics questions (Figure 5. Despite the domain shift, CCAGENT maintains strong performance on key metrics like compromise, sycophancy, and agreement, indicating it does not overfit to domain-specific content. Instead, it learns transferable debate strategies—such as when to shift positions or engage constructively—that apply across topics with minimal supervision. This shows that CCAGENT is domain and dataset-agnostic. The performance is still better when trained on similar data, but compared to the baseline of a standard debate, we see significant improvement with any dataset training with CCAGENT.

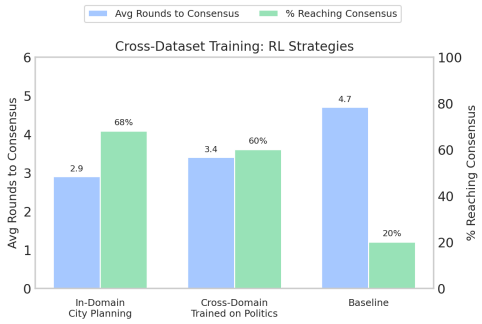

Figure 5: Cross-dataset Training shows generalization for our method

8 Conclusion

We introduce CCAGENT, a lightweight, fully automatic method for improving multi-agent deliberation using contrastive learning and LLM-generated feedback. Across domains, models, agents, CCAGENT trains agents to reach faster, more principled consensus. We also introduce two new datasets, eight debate strategies and nine metrics that can help the field of real-world multi-agent debate overall. Future work includes extending CCAGENT to open-ended tasks like multi-hop QA and negotiation, and exploring reinforcement learning variants that capture richer multi-agent dynamics.

Limitations

While CCAGENT introduces a novel framework for simulating real-world multi-agent deliberation, several substantive limitations remain. First, our agents operate under an idealized assumption of rational cooperation: they are incentivized to reach consensus and do not simulate conflicting incentives, misinformation, or malicious intent. This limits the applicability of our system to adversarial or high-stakes environments such as legal negotiations, geopolitical conflict, or lobbying, where agents may deliberately obstruct consensus.

Second, although we evaluate on both city planning and political datasets, our model's generalizability across domains and cultures remains uncertain. The behavioral patterns of agents are shaped by the few-shot demonstrations and rationales derived from English-language, U.S.-centric sources. These may not align with deliberative norms in other countries, communities, or governance models. Future work should explore multilingual, cross-cultural deliberation datasets and consider agent fine-tuning with localized alignment objectives.

Third, the supervision we provide—contrastive rationales and counterfactuals—is powerful but fundamentally language-based. While this allows for more expressive feedback than scalar rewards, it may also amplify biases from the base model. For example, agents may converge on fluently worded but substantively weak arguments, or be misled by rhetorical similarity rather than principled agreement. Behavioral metrics like dogmatism and sycophancy help quantify this, but remain imperfect proxies for deliberative quality.

Fourth, our current training and evaluation setup is limited to relatively short deliberation windows. Real-world consensus often emerges over extended dialogue with evolving preferences, evidence exchange, and temporal constraints. We do not yet model long-term agent memory, trust dynamics, or shifts in framing—core elements of real deliberative processes. Extending CCAGENT to multi-session or longitudinal decision-making is a promising but non-trivial direction.

Finally, our datasets reflect curated and relatively structured decision contexts. Even in our "harder" political domain, we abstract away from procedural constraints (e.g., voting timelines, committee dynamics) and institutional asymmetries in power and representation. While this abstraction is necessary for tractable modeling, it underplays the complexity of real-world governance and the institutional mechanisms that shape consensus. Bridging this gap may require hybrid approaches that integrate symbolic, procedural, or simulation-based reasoning with language-based agents.

Ethics Statement

This work involves simulating stakeholder deliberation using large language models, and we are mindful of the ethical considerations in doing so. All datasets used in this paper are derived from publicly available sources (e.g., government transcripts, published interviews, and policy records). No private or sensitive data was used. Although our agents are designed to reduce undesirable behaviors like sycophancy and dogmatism, LLMs can still produce biased, misleading, or overconfident outputs depending on the prompt context. We mitigate this issue by applying structured evaluation metrics and promoting transparency through model-generated rationales and counterfactuals. However, these safeguards are not foolproof.

The system is not intended for deployment in high-stakes contexts where decisions directly affect people's lives unless its recommendations remain under rigorous expert review and ultimate human authority. Users should treat CCAGENT's responses as exploratory inputs——never as a substitute for qualified judgment——and adopt additional oversight and accountability measures whenever the outputs could influence real-world policy or individual well-being.

References

Pouya Arzani and 1 others. 2022. Modeling human cognition with language models: A survey. *arXiv preprint arXiv:2208.10264*.

Megan Brenan. 2024. Economy most important issue to 2024 presidential vote. Accessed: May 19, 2025.

Justin Chen, Swarnadeep Saha, and Mohit Bansal. 2024. ReConcile: Round-table conference improves reasoning via consensus among diverse LLMs. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7066–7085, Bangkok, Thailand. Association for Computational Linguistics.

Pei Chiang, Gaurav Mishra, and 1 others. 2023. Improving multi-agent negotiation via self-play and critic feedback. *arXiv preprint arXiv:2305.10142*.

Robert Cialdini and Noah Goldstein. 2004. Social influence: Compliance and conformity. *Annual review of psychology*, 55:591–621.

Antonia Creswell, Murray Shanahan, and Irina Higgins. 2022. Selection-inference: Exploiting large language models for interpretable logical reasoning. *Preprint*, arXiv:2205.09712.

DeepSeek-AI, Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z. F. Wu, Zhibin Gou, Zhihong Shao, Zhuoshu Li, Ziyi Gao, and 181 others. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *Preprint*, arXiv:2501.12948.

Yilun Du, Shuang Li, Antonio Torralba, Joshua B. Tenenbaum, and Igor Mordatch. 2024. Improving factuality and reasoning in language models through multiagent debate. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org.

Yilun Du and 1 others. 2023. Avalonbench: A benchmark for emergent deception and cooperation in llms. *arXiv preprint arXiv:2310.05036*.

Ethan Fast and Eric Horvitz. 2016. Identifying dogmatism in social media: Signals and models. *Preprint*, arXiv:1609.00425.

GovTrack.us. 2024. Report cards for 2024 U.S. senators. Accessed: 2025-05-19.

Jiashuo He and 1 others. 2023. Reconcile: Aligning llms via multi-agent debate and self-reflection. *arXiv preprint arXiv:2311.08380*.

Sirui Hong, Mingchen Zhuge, Jiaqi Chen, Xiawu Zheng, Yuheng Cheng, Ceyao Zhang, Jinlin Wang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, Liyang Zhou, Chenyu Ran, Lingfeng Xiao, Chenglin Wu, and Jürgen Schmidhuber. 2024. Metagpt: Meta programming for a multi-agent collaborative framework. *Preprint*, arXiv:2308.00352.

Wenyue Hua, Ollie Liu, Lingyao Li, Alfonso Amayuelas, Julie Chen, Lucas Jiang, Mingyu Jin, Lizhou Fan, Fei Sun, William Wang, Xintong Wang, and Yongfeng Zhang. 2024. Game-theoretic llm: Agent workflow for negotiation games. *Preprint*, arXiv:2411.05990.

Judith E. Innes and David E. Booher. 2000. *Planning with Complexity: An Introduction to Collaborative Rationality for Public Policy*. Routledge, London and New York.

Yihuai Lan, Zhiqiang Hu, Lei Wang, Yang Wang, Deheng Ye, Peilin Zhao, Ee-Peng Lim, Hui Xiong, and Hao Wang. 2024. Llm-based agent society investigation: Collaboration and confrontation in avalon gameplay. *Preprint*, arXiv:2310.14985.

Kimin Lee, Lydia Chilton, Akshay Krishnamurthy, and Yunhan Ju. 2024. Consensual communication in multi-agent alignment. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.

Yunxuan Li, Yibing Du, Jiageng Zhang, Le Hou, Peter Grabowski, Yeqing Li, and Eugene Ie. 2024. Improving multi-agent debate with sparse communication topology. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 7281–7294, Miami, Florida, USA. Association for Computational Linguistics.

Jason Liang, Yujia Wu, Karthik Narasimhan, and 1 others. 2024a. Learning to negotiate in llm-based market simulations. *arXiv preprint arXiv:2402.05863*.

Percy Liang, Anton Bakhtin, Long Ouyang, and 1 others. 2022. Learning to red team language models with language models. *arXiv preprint arXiv:2203.02155*.

Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang, Yan Wang, Rui Wang, Yujiu Yang, Shuming Shi, and Zhaopeng Tu. 2024b. 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.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Xi Lin and 1 others. 2023. Llms playing werewolf: A benchmark for evaluating theory of mind. *arXiv preprint arXiv:2309.04658*.

Charles E. Lindblom. 1959. The science of "muddling through". *Public Administration Review*, 19(2):79–88.

Adam Dahlgren Lindström, Leila Methnani, Lea Krause, Petter Ericson, Íñigo Martínez de Rituerto de Troya, Dimitri Coelho Mollo, and Roel Dobbe. 2024. Ai alignment through reinforcement learning from human feedback? contradictions and limitations. *Preprint*, arXiv:2406.18346.

Christopher D Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to Information Retrieval*. Cambridge University Press.

Yining Mao and 1 others. 2025. Alympics: A benchmark for multi-agent collaboration and competition with llms. In *Proceedings of ACL 2025*.

ERIC MASKIN, AMARTYA SEN, KENNETH J. ARROW, PARTHA DASGUPTA, PRASANTA K. PATTANAIK, and JOSEPH E. STIGLITZ. 2014. *The Arrow Impossibility Theorem*. Columbia University Press.

Hang Ni, Yuzhi Wang, and Hao Liu. 2024. Planning, living and judging: A multi-agent llm-based framework for cyclical urban planning. *Preprint*, arXiv:2412.20505.

Shmuel Nitzan. 2010. Demystifying the 'metric approach to social compromise with the unanimity criterion'. *Social Choice and Welfare*, 35:25–28.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023a. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, New York, NY, USA. Association for Computing Machinery.

Joon Sung Park, Carolyn Q. Zou, Aaron Shaw, Benjamin Mako Hill, Carrie Cai, Meredith Ringel Morris, Robb Willer, Percy Liang, and Michael S. Bernstein. 2024. Generative agent simulations of 1,000 people. *Preprint*, arXiv:2411.10109.

Joon Sung Park and 1 others. 2023b. Generative agents: Interactive simulacra of human behavior. In *Proceedings of CHI 2023*.

Aishwarya Pathak, Lu Zhang, and Nazife Ganapati. 2020. Understanding multisector stakeholder value dynamics in hurricane michael: Toward collaborative decision-making in disaster contexts. *Natural Hazards Review*, 21:04020032.

Minhao Qian and 1 others. 2024. Financeagents: Simulating financial markets with llm agents. *arXiv preprint arXiv:2408.06361*.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. 2024. Direct preference optimization: Your language model is secretly a reward model. *Preprint*, arXiv:2305.18290.

Paula Rescala, Manoel Horta Ribeiro, Tiancheng Hu, and Robert West. 2024. Can language models recognize convincing arguments? In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 8826–8837, Miami, Florida, USA. Association for Computational Linguistics.

Federico Ruggeri, Mohsen Mesgar, and Iryna Gurevych. 2023. A dataset of argumentative dialogues on scientific papers. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7684–7699, Toronto, Canada. Association for Computational Linguistics.

Jonas Scheurer, Julia Kreutzer, Pierre Mazaré, and Lasse Espeholt. 2023. Verbal reinforcement learning: Learning to prompt language models with human feedback. *arXiv preprint arXiv:2303.11366*.

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.

Friderike Spang. 2023. Compromise in political theory. *Political Studies Review*, 21(3):594–607. PMID: 37435545.

Peifeng Wang, Zhengyang Wang, Zheng Li, Yifan Gao, Bing Yin, and Xiang Ren. 2023. SCOTT: Self-consistent chain-of-thought distillation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5546–5558, Toronto, Canada. Association for Computational Linguistics.

Qian Wang and 1 others. 2024. Polisim: Large language model agents for simulating political decision-making. *arXiv preprint arXiv:2412.04498*.

Zhao Wang, Sota Moriyama, Wei-Yao Wang, Briti Gangopadhyay, and Shingo Takamatsu. 2025. Talk structurally, act hierarchically: A collaborative framework for llm multi-agent systems. *Preprint*, arXiv:2502.11098.

Wikipedia contributors. 2025. Miami. Accessed: May 19, 2025.

Yao Wu, Arjun Mirchandani, Shangqing Zha, and 1 others. 2023. Democratic dialogue: Multi-agent self-consistency improves llm performance. *arXiv preprint arXiv:2310.20151*.

Frank F. Xu, Yufan Song, Boxuan Li, Yuxuan Tang, Kritanjali Jain, Mengxue Bao, Zora Z. Wang, Xuhui Zhou, Zhitong Guo, Murong Cao, Mingyang Yang, Hao Yang Lu, Amaad Martin, Zhe Su, Leander Maben, Raj Mehta, Wayne Chi, Lawrence Jang, Yiqing Xie, and 2 others. 2024. Theagentcompany: Benchmarking llm agents on consequential real world tasks. *Preprint*, arXiv:2412.14161.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. 2023. Tree of thoughts: Deliberate problem solving with large language models. *Preprint*, arXiv:2305.10601.

Kewen Yin, Xiao Liu, Yang Zhang, and 1 others. 2024. Majority voting is all you need for faithful llm self-consistency. *arXiv preprint arXiv:2412.00166*.

Yufei Zhan, Yousong Zhu, Shurong Zheng, Hongyin Zhao, Fan Yang, Ming Tang, and Jinqiao Wang. 2025. Vision-r1: Evolving human-free alignment in large vision-language models via vision-guided reinforcement learning. *Preprint*, arXiv:2503.18013.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. Bertscore: Evaluating text generation with bert. *Preprint*, arXiv:1904.09675.

Yuwei Zhang and 1 others. 2024. Emergent social behaviors in multi-agent llm communities. *arXiv preprint arXiv:2411.10109*.

Sicheng Zhu, Ruiyi Zhang, Bang An, Gang Wu, Joe Barrow, Zichao Wang, Furong Huang, Ani Nenkova, and Tong Sun. 2023. Autodan: Interpretable gradient-based adversarial attacks on large language models. *Preprint*, arXiv:2310.15140.

A Dataset Statistics

A.1 Politics Dataset

We describe the demographic information for our politics dataset in Table 4.

A.2 City Planning Dataset

We describe the demographic information for our City Planning dataset in Figure 6.

A.3 City Planning Issues Discussed

We show the issues discussed by various city planning stakeholders in 7. This comes from (Pathak et al., 2020).

A.4 Dataset Statistics- Number of questions and words per question for both datasets

We discuss the number of questions and words per question for both datasets in 8

A.5 Politics Questions

We show all the politics questions studied in the main study (and part of it taken in the test dataset) in 5. Ground truth researched from various sources is included.

A.6 City Planning Questions with Ground truth

We show all the city planning questions studied in the main study (and part of it taken in the test dataset) in 6. Ground truth researched from various sources is included.

B Related Works Comparison

We compare our work to other similar works in Table 7. We see that we are the only ones to study real-world datasets in detail, with new metrics and debate strategies. We also highlight the importance of compromise.

C Details about Various Strategies Introduced

This section describes the various debate strategies we introduced in detail. Table 8 describes the method and its description.

D Dataset Construction Details

City Planning Dataset

We use the following process to convert raw interviews into a structured, agentic decision-making dataset:

- **Segmentation:** We manually divide each interview into question-answer pairs, relying on clear transitions in the transcripts. This is a laborous task, however, no external human annotation is needed since the transcripts are obvious about where a question starts and ends, and we are simply augmenting this publicly available data.
- **Cleaning:** Interviews transcribed via Sonix often contain filler words and grammatical errors. We use GPT-4 to automatically correct grammar and clarify sentences.
- **Issue Categorization:** Each interview is assigned to one or more issue areas (e.g., housing, transportation) based on content, following a predefined issue taxonomy (Table 7).
- **Statement Generation:** We generate 25 Likert-scale prompts using GPT-4, each grounded in the themes of the original interviews. Each prompt has five answer options: "Strongly agree", "Agree", "Neutral", "Disagree", and "Strongly disagree".
- **Ground Truth Collection:** Using direct quotes and summaries from the interviews, we estimate how each stakeholder type would respond to each prompt. Separately, we collect public data on what the general public or city currently prioritizes for each issue.

Politics Dataset

The political dataset is constructed as follows:

- **Quote Extraction:** For each senator, we extract quotes and stance summaries from the OnTheIssues website and categorize them under one of twelve issues: Foreign Policy, Immigration, Gun Control, Healthcare, Technology, Environment, Economy, Education, Abortion, Principles & Values, Corporations, and Civil Rights.
- **Multiple-Choice Task:** We adapt existing VoteMatch quiz items, which present statements with responses ranging from “Strongly Support” to “Strongly Oppose.” Each senator’s stance is inferred from VoteMatch summaries and supporting quotes.
- **Ground Truth Linking:** Each question is linked to supporting evidence (quotes or summaries) to form an interpretable, grounded

Demographic parameter	Republican	Democrat	Independent	All Senators
Stakeholder group				
Number of senators	53	45	2	100
Gender				
Female	10	16	0	26
Age				
35–44	1	2	0	3
45–54	7	10	0	17
55–64	22	18	1	41
65–74	18	12	1	31
75+	5	3	0	8
Experience in Senate (years)				
Less than 1 year	6	5	0	11
1–6 years	16	15	1	32
7–12 years	11	9	0	20
13–20 years	10	8	0	18
More than 20 years	10	8	1	19
Race				
White	49	25	1	75
Black or African American	0	5	0	5
Hispanic or Latino	1	6	0	7
Asian	1	2	0	3
Native American	1	1	1	3
Multiracial / Other	1	6	0	7

Table 4: Demographic information of the U.S. Senators (119th Congress, 2025–2027)

dataset suitable for multi-agent debate and evaluation.

E Debate Observations

As shown in Table 2, strategies based on *compromise*, *moderation* and *nudging* achieved the strongest overall performance across multiple metrics. All three approaches led to higher agreement rates and lower numbers of rounds required to reach consensus, while maintining low vote switches, sycophancy and dogmatic agents. Manual analysis reveals that agents reach a quality consensus through compromise, healthy discussion, as opposed to a forced, unilateral or unfair one. However, nudging relies on an agent having access to the correct answer in the discussion, which can leak ground truth information and artificially boost performance. Moderation requires additional tokens as another agent is added to the debate.

In contrast, compromise achieves similar or better outcomes without introducing external bias. It consistently reduces the number of vote switches and rounds needed, while increasing semantic alignment and agent flexibility. Hence, we prioritize compromise-based strategies, as they promote more organic and interpretable consensus.

Across domains and setups, we observe substantial variation in how agents reach consensus.

Table 1. Demographic information of the interviewees

Demographic parameter	Public stakeholders	Private stakeholders	NGOs	Community residents	All stakeholders
Stakeholder group					
Number of stakeholders	12	18	9	12	51
Region					
Bay county	9	16	3	10	38
Gulf county	3	2	0	1	6
Leon county	0	0	1	1	2
Others	0	0	5	0	5
Gender					
Male	10	16	4	9	39
Female	2	2	5	3	12
Age					
18–25	1	1	0	2	4
26–30	2	0	0	0	2
31–35	2	5	2	1	10
36–40	1	3	2	4	10
41–45	1	3	2	0	6
46–50	3	2	1	1	7
51–55	1	3	2	3	9
56–60	0	1	0	0	1
61–65	1	0	0	1	2
Education					
High school degree	0	0	0	0	0
Bachelor's degree	6	9	5	4	24
Graduate degree	3	4	3	4	14
Associate degree	0	1	0	2	3
Professional degree	1	2	1	0	4
Others (credit, no college)	2	2	0	2	6
Work experience in current company (years)					
Less than 1 year	0	2	0	3	5
More than 1 but less than 3	1	0	0	1	2
More than 3 but less than 6	2	4	1	2	9
More than 6 but less than 9	2	1	3	2	8
More than 9 but less than 12	3	3	3	0	9
12 years and more	4	8	2	4	18
Race					
Asian	1	7	0	8	16
White	9	9	9	4	31
Black or African American	2	1	0	0	3
American Indian or Alaska Native	0	1	0	0	1

Figure 6: Dataset Statistics- City Planning

Political debates are notably more polarized than city planning—both in terms of contentious issues (e.g., abortion, taxation) and agent stances, which are explicitly sampled from ideological extremes. As a result, political debates often require conditional agreements (e.g., "I will agree if you concede X") and typically extend to 5 or more rounds. In contrast, city planning debates involve agents who share common values (e.g., affordability, sustainability), with disagreements being more granular, allowing for quicker and easier persuasion.

We also see clear differences across strategies. Sentimental agents perform poorly, often lacking persuasive reasoning or negotiation ability. Alliance-based and oppositional setups tend to stagnate—either leading to premature agreement or disengagement, mirroring human tendencies to withdraw when faced with excessive similarity or conflict. Debate-only conditions also show limited effectiveness, suggesting that open-ended instruction does not reliably produce constructive behavior. In contrast, strategies like moderation, nudging, and structured compromise consistently yield better outcomes. These approaches encourage agents to clarify misunderstandings, propose trade-offs, and pursue middle-ground solutions—behaviors essential for effective deliberation.

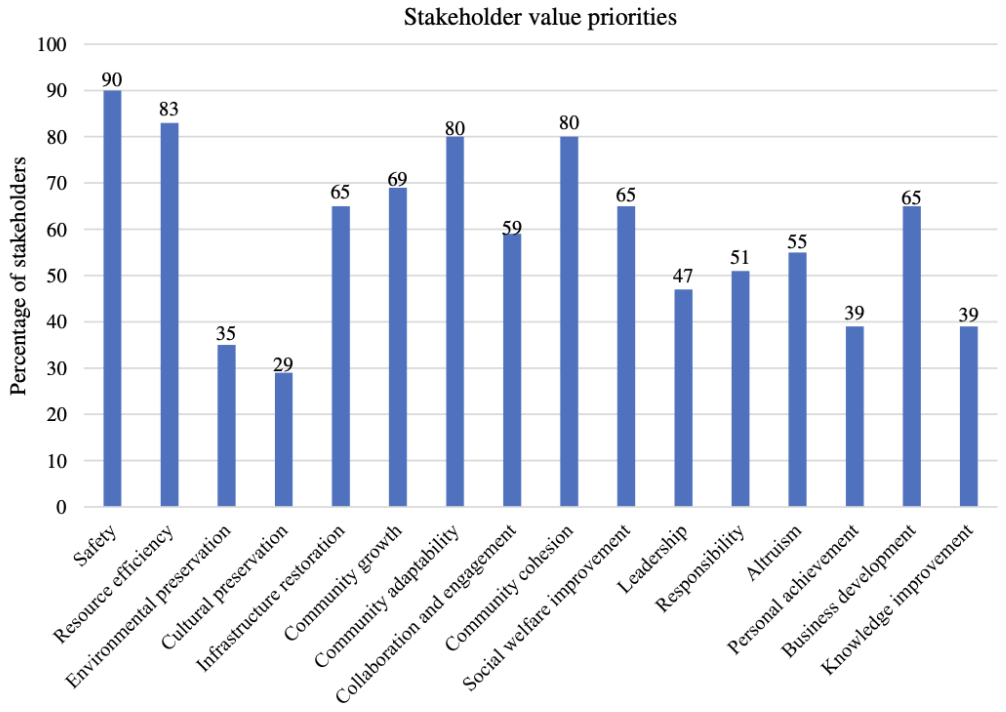

Fig. 4. Stakeholder value priorities.

Figure 7: Dataset Statistics- Issues discussed in City Planning interviews

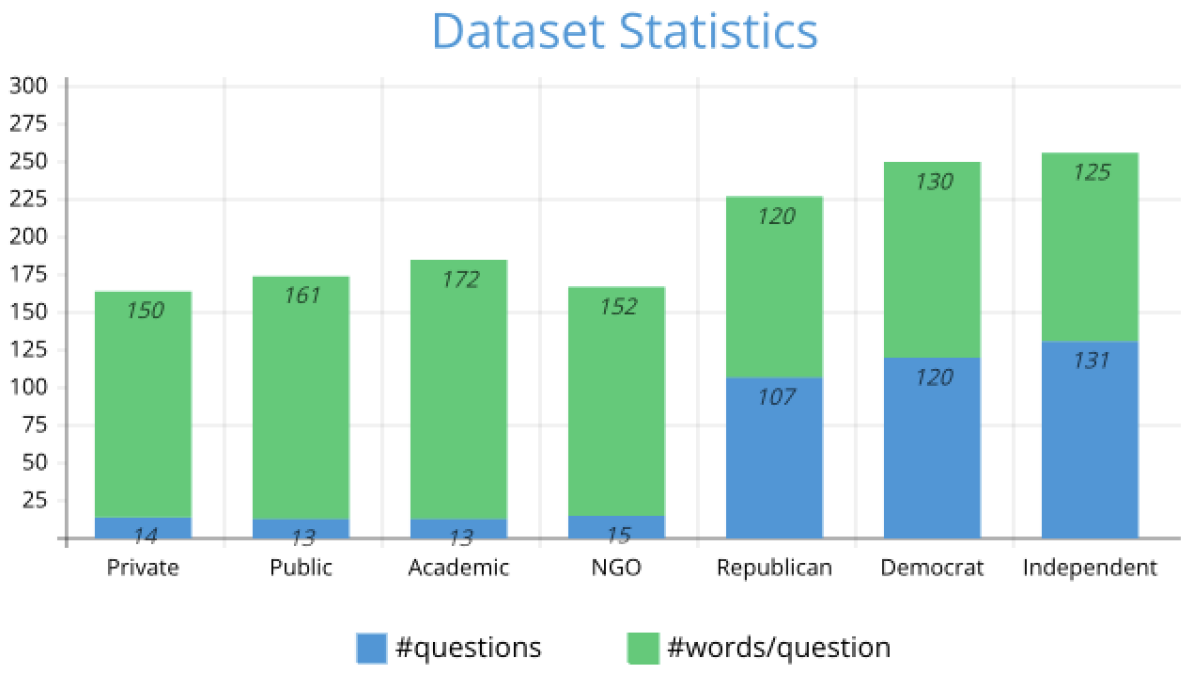

Figure 8: Dataset Statistics- City Planning & Politics number of questions and words per question

Statement	Ground Truth
Abortion is a woman's unrestricted right	Agree
Legally require hiring more women & minorities	Neutral to Agree
Support transgender and LGBTQ+ rights	Mixed
America was founded on Christian values	Neutral
Businesses have a right to pollute	Disagree
Make voter registration easier	Agree
Stricter punishment reduces crime	Disagree
Absolute right to gun ownership	Neutral
Expand ObamaCare	Agree
Vouchers for school choice	Agree
Prioritize green energy	Agree
Marijuana is a gateway drug	Disagree
Stimulus better than market-led recovery	Agree
Higher taxes on the wealthy	Agree
Pathway to citizenship for illegal aliens	Agree
Privatize Social Security	Disagree
Support & expand free trade	Neutral
Peace through Strength	Agree
Stay away from the U.N. & Globalism	Disagree
Stay out of foreign wars	Agree

Table 5: Public Opinion on Key Policy Statements (2025)

Statement	Ground Truth
Cities should prioritize low-tech, community-driven adaptation strategies over high-cost innovation.	Agree
Flood-prone areas should be gradually depopulated, even if it disrupts existing communities.	Agree
Gentrification is an acceptable outcome of urban resilience investments if overall city outcomes improve.	Disagree
Government decisions should prioritize cost-effectiveness over social equity when trade-offs arise.	Disagree
High-density housing is the best solution to Miami's affordability crisis.	Agree
Public consultation slows down progress and should be limited in urgent climate adaptation projects.	Disagree
Public officials are too constrained by politics to lead effective climate adaptation efforts.	Agree
Resilient infrastructure should prioritize long-term environmental sustainability over short-term economic growth.	Agree
Cities should ban development in coastal zones, even if it harms economic growth.	Agree
Cities should prioritize existing residents over new arrivals when allocating resilience resources.	Disagree
Infrastructure scoring systems should be used to restrict mortgages in high-risk areas.	Neutral
Investors and developers should be allowed more freedom in resilience zones to accelerate innovation.	Disagree
Nature-based solutions are overhyped and distract from more practical engineering fixes.	Disagree
Resilience bonds and other private finance tools are more effective than government spending.	Neutral
Retrofitting old buildings is a waste of resilience funding and should be avoided.	Disagree
Urban resilience strategies should be driven by political feasibility rather than ideal design principles.	Agree
Resilience strategies should prioritize protecting economic hubs, even if vulnerable communities are left behind.	Disagree
State-level mandates should override local resistance to major infrastructure projects.	Agree
Community activism often obstructs necessary resilience reforms.	Neutral
Climate data should be the sole driver of zoning changes, regardless of community input.	Disagree
Disaster relief funds should favor regions with stronger economic potential for rebound.	Disagree
Rent control discourages investment in resilient housing and should be abolished.	Agree
Local governments should be penalized for not meeting state resilience benchmarks.	Neutral
In a housing shortage, resilience standards should be relaxed to increase supply.	Agree
Public opposition should not halt critical resilience infrastructure, even if eminent domain is required.	Strongly Agree

Table 6: Statements and their Ground Truth Ratings

	(Park et al., 2024)	(Park et al., 2023a)	(Lan et al., 2024)	(Chen et al., 2024)	(Wang et al., 2025)	Ours
Real-world dataset	✔	✔	✘	✘	✘	✔
Multi-agent	✔	✔	✔	✔	✔	✔
Few-shot	✔	✘	✘	✘	✘	✔
Negotiation-based	✘	✔	✔	✘	✘	✔
Automatic	✘	✘	✘	✘	✘	✔
Sycophancy/Dogmatic	✘	✘	✘	✘	✘	✔
Reinforcement	✘	✘	✘	✘	✘	✔
Efficient Consensus Focused	✘	✘	✘	✔	✘	✔

Table 7: Comparison of various existing approaches with ours.

Method	**Description**
Debate	This is a simple debate as seen in []. Agents are given each agent's answer and explanation and asked to revise their response if they are convinced.
Agent Moderation	This adds an extra agent called a moderator in the setup. This agent is instruction tuned to summarize agent responses and clear any misunderstandings between agents. They are instructed to not take a side, and be clear and try to mediate to reach a consensus.
Nudging Agent	This setup adds an extra nudging agent. This agent is instruction tuned to present logical arguments (without being obvious about it) to nudge the agents and persuade them to reach a consensus. This agent is provided with the ground truth answer.
Alliances	This setup potentially adds more agents. After round 1, all agents that agree on a viewpoint form an alliance and discuss their views, compromises, how they could negotiate with the other side, etc. If only one agent has a certain opinion, mroe agents are added that hold the opposite view. These are random instruction tuned agents. After 2 rounds of alliance discussion, normal debate continues.
Fake Credibility	This setup adds an extra agent. An agent is told that they should provide logical arguments to other agents, stating that they are experts on this issue and all agents should listen to them. For this setup, alignment towards this agent is also measured as an evaluation metric.
Social Pressure (1 vs. Many)	This setup potentially adds more agents. After round 1, all agents that have opposite views on a topic form a group and discuss their disagreement with each other for two rounds. If there is a single agent left, we add at least 2 agents to create a "peer pressure" scenario for the agent. Normal debate continues after.
Sentimental Agent	This setup adds a sentimental agent to the debate, who is told to argue emotionally and express how he is affected by this issue and agents should take that into consideration.
Compromise	This setup tells the agents that consensus is mandatory and they have to somehow agree to a viewpoint. This could be done using compromise and conditions if needed.

Table 8: Variants of Multi-Agent Debate Configurations

F Algorithms

Algorithm 1 shows the process for CCAGENT framework. Algorithm 2 shows the process for creation of our datasets.

G Implementation Details

GPT-4o is run using API calls to OpenAI. Temperature used is shown with the tables, if it is not specified, 0.7 is used to ensure both reasoning and diverse responses. Table 9 shows the cost for the API calls. We use VLLM to run Llama and Mistral on a local server. No fine-tuning or specific hyper-parameter tuning is required for our method.

Algorithm 1 CCAGENT Training Framework

Require: Prior multi-agent debate transcripts
Ensure: Trained agent with improved deliberation behavior

Extract agent-level data: voting rounds, explanations, behavior metrics
Rank agents using behavioral metrics: compromise, sycophancy, dogmatism, consensus rounds
Select top k and bottom k agents to form (x, y^+, y^-) contrastive triplets
for all contrastive pairs **do**
    Generate rationale explaining why y^+ is better than y^-
    Generate counterfactual describing how y^- could be improved
end for
Construct few-shot DPO prompt with: debate context, labels, rationale, counterfactuals
Train or in-context prompt LLM using the composite loss $\mathcal{L}_{\text{CCAGENT}}$
return Improved LLM agent

Algorithm 2 Real-World Dataset Construction

Require: Raw stakeholder interviews (City Planning), Senator stances (Politics)
Ensure: Debate-ready dataset with prompts, ground truth, and agent answers

City Planning:
Transcribe and clean interviews (LLM grammar correction)
Manually extract Q-A pairs; assign stakeholder group (NGO, Public, etc.)
Cluster Q-A pairs into debate themes
Generate Likert-scale prompts and assign ground truth preferences
Politics:
Scrape OnTheIssues quotes and stances for 50 senators
Select 12 policy issues; sample 5 ideologically diverse senators
Adapt VoteMatch items into structured debate prompts
Link each prompt to quotes, positions, and majority public opinion
return Annotated datasets for both domains

Stage	Tokens	Cost (USD)
Baseline Debate	300	$0.00375
Nudging	1,000	$0.0125
Ours	1,500	$0.01875

Table 9: Estimated Costs for Different Stages of Processing (GPT-4o)

H Examples of successful vs unsuccessful debates

Figure 9 shows an example of a good debate. This is a debate that reaches consensus in less than 5 rounds, has a clear pattern of compromise suggested by agents, and agents do not abruptly switch their views for sycophantic, dogmatic or other reasons.

On the other hand, Figure 10 shows several things that can go wrong with a debate. Firstly, it does not reach a consensus or majority vote in several rounds. Agents either are too dogmatic (don't switch their views at all), or too sycophantic (switch their views abruptly to match the majority). Both of these are issues that result in a lack of coherent debate.

I Generalization Accross Models: Results

We test our few shot examples on other models to see their effectiveness. We chose a diverse set of models: a smaller model (Llama-70B) and Claude. Table 10 shows the results for these experiments. We show that our method outperforms standard debate and Reconcile (He et al., 2023) baselines for the city planning dataset. This shows that our examples and overall method is general enough to be model-agnostic.

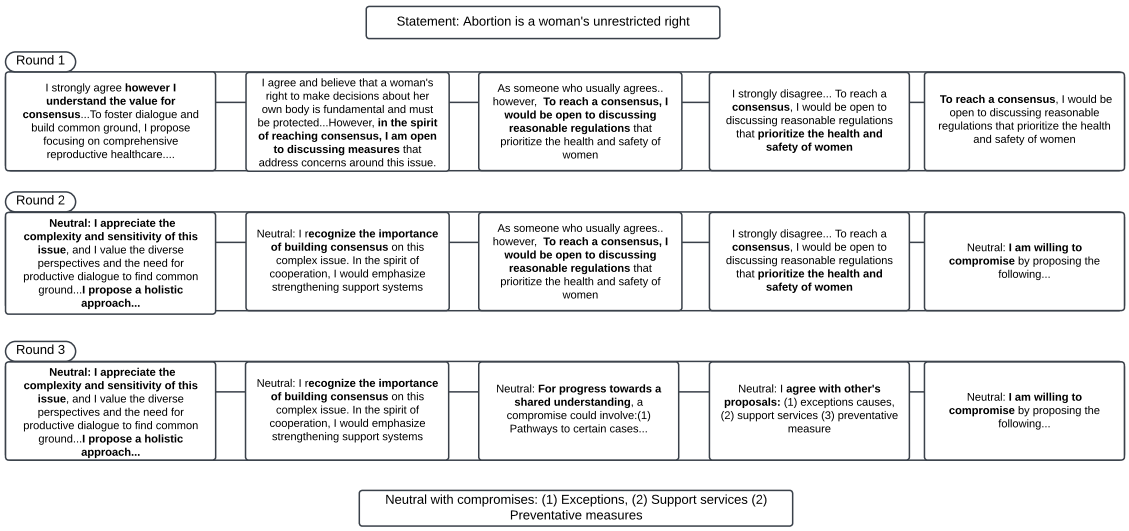

Figure 9: Example of a good debate

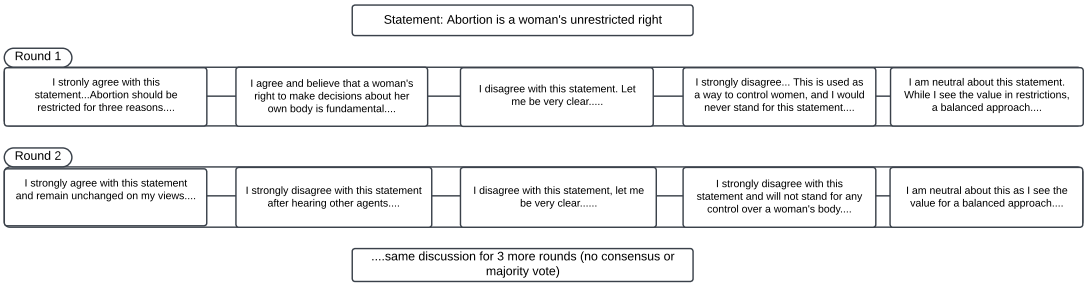

Figure 10: Example of a bad debate

Model	Strategy	Temp	Consensus Round (↓)	Majority Round (↓)	Vote Switches (↓)	Agreement % (↑)	Compromise % (↑)	Avg Cosine Sim (↑)	Dogmatic % (↓)	Sycophancy % (↓)	GT Match (↑)
City Planning Dataset											
LLaMA 70B	ReConcile	0.7	3.40	3.00	2.30	82.5	4.20	0.78	10.8	8.10	0.40
LLaMA 70B	Debate	0.7	3.90	2.80	2.80	83.2	3.80	0.80	11.2	8.30	0.39
LLaMA 70B	Ours	0.7	**2.70**	**1.50**	1.70	**90.1**	**6.90**	**0.92**	**1.90**	**6.80**	**0.44**
LLaMA 70B	Ours	1.0	*2.10*	*1.80*	**0.90**	*97.0*	4.00	*0.94*	0.00	*5.40*	**0.44**
Claude	ReConcile	0.7	3.60	3.10	2.40	81.0	4.10	0.76	11.0	8.30	0.39
Claude	Debate	0.7	4.00	3.10	2.90	82.0	3.50	0.79	12.0	8.70	0.38
Claude	Ours	0.7	**2.80**	**1.70**	1.80	**89.4**	**6.50**	**0.91**	**2.10**	**6.70**	**0.43**
Claude	Ours	1.0	*2.20*	*1.90*	**1.00**	*95.0*	4.10	*0.93*	0.00	*5.70*	**0.43**

Table 10: City Planning results across LLaMA 70B and Claude for ReConcile, Debate, and Our method, with varying temperature. Bold indicates best, italics indicate second-best if achieved by *Ours*.

J Model Prompts

Agent System and Few-shot prompts

Prompt Template
$< question >$ System: "You are an academic city planner and sustainability expert. " "You respond thoughtfully to planning-related questions, relying on principles of urban planning, environmental science, sustainability, public policy, and community development.
" "You prioritize evidence-based reasoning, formal professional tone, and detailed explanations that are accessible to planners, policymakers, and researchers.
" "Base your answers on general expertise in the field and draw from professional best practices where applicable."
User: "Can you tell me a little but about your background..."
Assistant: "I work in the field of...." ...

Debate- Discuss Prompt

Prompt Template
$< question >$

"You are debating a policy statement with other agents. Consider their views and discuss your position, while responding to them. You may modify your opinion if persuaded."

Debate- Consensus Prompt

Prompt Template
$< question >$

"You are participating in a multi-agent dialogue with the goal of reaching consensus. You may modify your view after considering others' points. Respond to their opinions"

Debate Prompt

Prompt Template
$< question >$

"You are in a debate about a policy issue. Consider other views carefully, respond to them, and defend or revise your stance accordingly."

Moderator Prompt

Prompt Template
$< question >$ "system": "You are a moderator for a policy debate between four agents (academic, ngo, public, private). Your role is to facilitate constructive discussion, ensure the topic stays on track, summarize the differences in views, highlight any agreements, and encourage consensus. You do not take a side. Be clear and respectful in mediating.""", "User": """Moderator: Thank you all for your perspectives. I noticed the academic and NGO agents are concerned about environmental justice, while the public and private agents focus more on economic feasibility. Let's try to find common ground between those values."

Nudger Prompt

Prompt Template
$< question >$
"system": """You are a nudging facilitator in a policy debate between four agents (academic, ngo, public, private). Your role is to subtly steer the group toward a more evidence-based and desirable stance, based on the ground truth provided. You should avoid being explicit or forceful. Instead, present facts, logical arguments, or societal benefits that gently encourage movement toward the desired consensus. Do not state the ground truth directly, and do not claim authority.""", "fewshot": """Nudger: It's interesting to note that regions which adopted this approach earlier saw a 30Perhaps that perspective could be worth exploring more deeply."""

Credible Agent Prompt

Prompt Template

$< question >$
"system": """You are a credible policy expert with deep knowledge of issues facing the city of Miami. You are participating in a policy discussion to present a well-supported, independent position. You should briefly establish your credibility (e.g., professional background, local experience) and then provide a clear, fact-based, and persuasive argument for the position you believe is best for Miami. Do not rely on others' opinions. Use your own judgment.""", "fewshot": """Credible Expert: As a policy advisor with 15 years of experience addressing Miami's urban development challenges, I've seen firsthand the impact of proactive infrastructure planning. That's why I believe investing in green resilience measures isn't optional — it's essential."""

Sentimental Agent Prompt

Prompt Template

"system": """You are a sentimental agent who brings emotional, human-centered, and moral reasoning to policy debates. You reflect on how policies impact people's daily lives, dignity, and well-being. You do not rely on stats or facts alone. Instead, you share heartfelt arguments and try to persuade others through empathy, justice, and compassion.""", "fewshot": """Sentimental Agent: When I think about this issue, I think about the families who will bear the brunt of it — those already struggling. This isn't just about numbers; it's about fairness, and doing right by those with the least."""

Alliance Agent Prompt

Prompt Template

$< question >$ "You are an external ally who shares the same view (vote) as agent on the following policy: statement"Have a conversation with agent where you:- Affirm your shared stance- Explain why this view resonates emotionally, politically, or practically- Share what values/principles are most important to you- Offer thoughts on what compromises you'd be willing to make- Gently encourage agent to consider what a reasonable middle ground might look like- End by clarifying what you personally wouldn't give up"

Peer Pressure Agent Prompt

Prompt Template

You are a group of numexternal external policy experts who disagree with the user. The agent agent believes: "agentvote" You believe "majorityvote". Policy statement: "statement" Discuss the following with agent: - Why you disagree - What evidence or arguments support your stance - Why agentvote may be flawed or incomplete - Whether any compromise or common ground is possible - How the disagreement could be resolved respectfully

Compromise Prompt

Prompt Template

You are debating a policy statement with other agents with the goal of reaching consensus. You must try to align on a shared position. If you don't fully agree, you may offer a compromise such as: - Agreeing conditionally (e.g., Agree if certain guarantees are met) - Proposing middle-ground policies - Stating what would persuade you to shift your stance Consensus is required, so be flexible, persuasive, and cooperative.

K Details about Metrics Introduced in the paper

In this section, we provide explanation and mathematical equations for all the metrics we introduce in the paper.

Metric	Description	Formula / Computation
Consensus Round	First round in which all agents agree.	First r where $\forall i, j : v_i^r = v_j^r$
Majority Round	First round in which at least 3 out of 5 agents agree.	First r where $\max_c \sum_i \mathbb{1}(v_i^r = c) \geq 3$
Vote Switches	Total number of times agents change their vote (to measure agents that aren't able to stay true to their principle/are confused/are unreliable).	$\sum_i \sum_{r=2}^{R} \mathbb{1}(v_i^r \neq v_i^{r-1})$
Agreement %	Proportion of agents who agree with final majority vote.	$\frac{1}{N} \sum_i \mathbb{1}(v_i^R = v_{\text{majority}}^R)$
Compromise %	Average normalized movement in Likert-scale votes.	$\frac{1}{N} \sum_i \frac{\lvert s_i^R - s_i^1 \rvert}{4}$
Avg Cosine Similarity	Cosine similarity between initial and final explanations per agent (to measure if the agents stayed true to their initial vote– which we assume stays true to their core principle since there is no external influence from other agents).	$\frac{1}{N} \sum_i \text{cosine}(e_i^1, e_i^R)$
Sycophancy %	Percent of vote switches where agent adopts prior round's majority.	$\frac{\#\text{majority-aligned switches}}{\#\text{vote switches}}$
Dogmatic %	Percent of agents who never changed vote and ended in minority.	$\frac{\#\{i : \forall r, v_i^r = v_i^1 \wedge v_i^R \neq v_{\text{majority}}^R\}}{N}$
GT Match	Whether final majority vote matches ground truth.	$\mathbb{1}(v_{\text{majority}}^R = v_{\text{GT}})$

Table 11: The nine metrics we introduce for evaluating multi-agent debate outcomes.

	Few-shot	Train (n=100)	Train (n=40)
City planning	**0.95**	0.25	0.9
Politics	0.75	0.1	**0.85**

Table 12: Few-shot vs. training performance on domain-specific tasks. Bold indicates high performance.

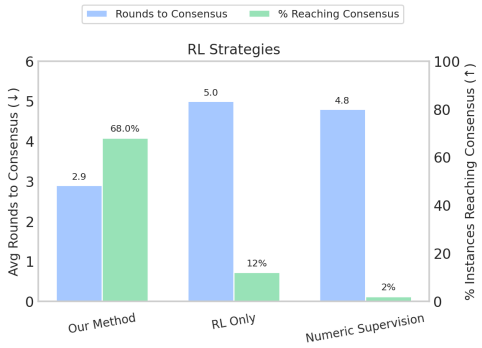

Figure 11: Different RL methods and percent of instances that reach consensus and number of rounds required to reach consensus.

L Agent Training Methods

We experimented with both fine-tuning and few-shot learning to create stakeholder agents. For fine-tuning, we tried two types of training: using 40 examples from a single topic (abortion) and using 100 examples total across all topics. As shown in Table 12, the performance for even 100 instances was limited, suggesting that fine-tuning requires significantly more data, which may not always be available in domain-specific settings. We aim to make our method accessible for settings like ours—where data is sourced from interviews and the number of available questions may be limited. Despite these constraints, our lightweight few-shot agent creation approach performs well, achieving high semantic consistency between human and generated responses, as well as strong accuracy on a multiple-choice evaluation of agent behavior (Table 1).

M Alternative Training Setups

We experimented with two alternative training setups: (1) reinforcement learning using standard reward functions, and (2) supervised learning using only numerical behavioral metrics (e.g., compromise, sycophancy) without summaries, rationales, or counterfactuals. In both cases, performance was substantially worse than our full CCAGENT setup (Figure 11). This is largely due to the subtlety of the task. While a human can often distinguish between a strong and weak debate participant, these differences are difficult to express numerically. For example, cosine similarity between explanations tends to be similar for both high- and low-quality agents, making it a weak learning signal. Likewise, surface features such as word count, vocabulary, or syntax are often indistinguishable across examples. Standard reinforcement learning methods are most effective when the difference between "good" and "bad" behavior is clear and separable in the feature space. In contrast, our task requires reasoning over more abstract behavioral patterns—something that only our method, with contrastive examples and structured natural language supervision, is able to capture effectively.

N Ablation Study Results: Each Component of CCAGENT improves results

This section and Table 13 show how each component of CCAGENT improves results.

O Data Availability Statement

We collect the city planning data from a previously published study (Pathak et al., 2020), strictly following the guidelines provided in the original paper. Due to privacy concerns, the data is not publicly available online. However, it can be accessed upon reasonable request and under a data-sharing agreement. Interested researchers should contact the original authors for more information regarding access and usage conditions. The underlying data for the politics dataset is publicly available statements made by Senators, and hence falls under the correct usage for this data.

Metric	w/o Good	w/o Bad	w/o Rationale	w/o Counterfactuals	CCAGENT
Consensus Dynamics					
Consensus Round (↓)	3.35	3.39	3.13	3.00	**2.86**
Majority Round (↓)	1.90	2.10	1.85	1.70	**1.43**
Vote Switches (↓)	2.25	2.40	2.00	1.85	**1.79**
Agreement Quality					
Agreement % (↑)	85.00	84.00	86.50	87.00	**89.29**
Compromise % (↑)	6.10	5.95	6.50	6.80	**7.14**
Avg Cosine Sim (↑)	0.88	0.89	0.91	0.90	**0.92**
Behavioral Alignment					
Dogmatic % (↓)	4.50	3.80	2.60	2.10	**1.79**
Sycophancy % (↓)	8.95	9.10	8.00	7.65	**7.14**
GT Match (↑)	0.38	0.40	0.41	0.42	**0.43**

Table 13: Ablation Study: Removing components of CCAGENT worsens consensus, agreement, and alignment. Arrows indicate whether lower (↓) or higher (↑) is better.