

# ARGSBASE: A Multi-Agent Interface for Structured Human–AI Deliberation

Frieso Turkstra \*

Sara Nabhani \*

Khalid Al-Khatib\*

University of Groningen

{f.turkstra,s.nabhani,khalid.alkhatib}@rug.nl

## Abstract

We present a new deliberation interface that enables users to engage with multiple large language models (LLMs), coordinated by a moderator agent that assigns roles, manages turn-taking, and ensures structured interaction. Grounded in argumentation theory, the system fosters critical thinking through user–LLM dialogues, real-time summaries of agreements and open questions, and argument maps. Rather than treating LLMs as mere answer providers, our tool positions them as reasoning partners, supporting epistemically responsible human–AI collaboration. It exemplifies hybrid argumentation and aligns with recent calls for “reasonable parrots,” where LLM agents interact with users guided by argumentative principles such as relevance, responsibility, and freedom. A user study shows that participants found the tool easy to use, perspective-enhancing, and promising for research, while suggesting areas for improvement.

DEMO URL: <http://1.6.98.141:9156>

## 1 Introduction

Deliberation, the thoughtful exchange of arguments, is a key process in democratic systems, education, and group decision-making. It helps people think critically, understand different perspectives, and make more informed choices, especially when addressing complex or controversial issues. Research shows that effective deliberation can improve the quality of collective decisions and increase public trust in their outcomes. In response to its significance, the field of computational argumentation has started to explore how technology can support and model deliberative processes. This growing interest is reflected in new research initiatives, such as the first *Workshop on Language-driven Deliberation Technology* held in 2024<sup>1</sup>

Despite the apparent benefits of tools that support deliberation for end users, only a few such systems currently exist. Some notable examples include *Discussion Tracker*<sup>2</sup>, which assists teachers in evaluating students’ collaborative argumentation using language technologies, and *BCause.app*<sup>3</sup>, which promotes healthier online discussions through structured interactions and reflective feedback. While these tools offer valuable contributions, they do not yet leverage the full potential of large language models (LLMs), particularly in the context of agentic systems, to allow more dynamic and effective deliberative processes.

We propose *ArgsBase*, a new tool that facilitates deliberation between users and multiple LLMs to support effective decision-making. The use of multiple LLMs allows the system to draw on the different strengths and capabilities of each model. A central moderator agent orchestrates the interaction, managing turn-taking and assigning roles to both the user and the models to ensure a coherent and structured dialogue. The deliberation process is guided by well-established principles from argumentation theory, such as pragma-dialectics, and considers tasks such as fallacy detection, while maintaining a clear and conversational style. The tool also provides real-time summaries focused on key deliberative elements, such as open questions and points of agreement. Besides, an argument map is generated to visualize the main arguments discussed and their relationships.

The proposed tool is an example of hybrid argumentation<sup>4</sup>, aiming to support epistemically responsible and constructive human–AI collaboration. It contributes to the broader vision of hybrid intelligence, in which AI systems are designed to enhance rather than replace human reasoning. This

<sup>2</sup><https://discussiontracker.cs.pitt.edu>

<sup>3</sup><https://bcuse.app>

<sup>4</sup>Lorentz Center Workshop on Hybrid Argumentation and Responsible AI

\*Equal contribution.

<sup>1</sup>DELiTe 2024 Workshop website

work also aligns with recent calls for conversational technologies specifically designed to support argumentative reasoning, addressing the limitations of current LLMs in this area. [Musi et al. \(2025\)](#) advocate for treating LLMs as tools for practicing critical thinking, introducing the concept of “reasonable parrots”; agents that engage in a discussion based on the principles of relevance, responsibility, and freedom grounded in argumentation theory.

*ArgsBase* is designed for users who engage in structured reasoning, critical reflection, or collaborative decision-making. Its primary target audience includes researchers in computational linguistics, argumentation, and human–AI interaction, as well as educators and students interested in exploring deliberative dialogue. The system is also well-suited for social scientists studying online discourse, developers building reasoning-centered AI applications, and public engagement practitioners seeking to facilitate balanced, multi-perspective discussions on complex topics.

## 2 Related Work

Our work intersects with three lines of research: Human–AI collaboration, multi-agent language model frameworks, and online public deliberation platforms. Each of these areas offers insights into the design and impact of AI systems aimed at augmenting human reasoning and dialogue.

**Human–AI Collaboration** Human–AI Collaboration has shown promise across domains, improving performance and supporting informed decision-making. In social chatbots, AI is often seen as a companion offering emotional support ([Brandtzaeg et al., 2022](#)), while in mental health, it can enhance empathy in peer interactions ([Sharma et al., 2023](#)). In education, AI fosters critical thinking and personalized learning ([Markauskaite et al., 2022](#); [Muthmainnah et al., 2022](#)), and in customer service, it boosts efficiency by handling routine tasks ([Vassilakopoulou et al., 2022](#)). Jiang et al. ([Jiang et al., 2022](#)) stress that effective collaboration requires systems that support users without overwhelming them, highlighting the value of clear communication and intuitive design.

*ArgsBase* advances hybrid argumentation by fostering critical thinking, reflection, and multi-perspective reasoning. Unlike chatbots or educational tools centered on emotional or personalized engagement, it positions AI as a reasoning partner in structured, epistemically responsible dialogue.

**Multi-agent Collaboration Approaches** Recent work highlights the value of multi-agent systems for improving LLM reasoning, factuality, and self-correction via structured disagreement. Tree-of-Debate ([Kargupta et al., 2025](#)) transforms scientific papers into LLM personas that engage in dynamic debates for literature synthesis. [Du et al. \(2024\)](#) propose a task-agnostic “society-of-minds” approach, where agents iteratively debate and converge on solutions. PREDICT ([Park et al., 2024](#)) combines cross-stance debates with perspective-based reasoning to enhance robustness in hate speech detection. Other work explores debate as a mechanism for truth alignment ([Irving et al., 2018](#)) and promotes divergent reasoning through judge-guided interactions ([Liang et al., 2024](#)).

In contrast to debate-based multi-agent systems, *ArgsBase* enables real-time human–agent deliberation. Rather than converging on a single outcome, it surfaces diverse perspectives and fosters user reflection through structured, moderated dialogue.

**Public Deliberation Platforms** Several systems support structured online public deliberation. *BCause.app* addresses the downsides of social media by introducing lightweight argument structuring and reflective feedback. *COLLAGREE* ([ITO et al., 2015](#)) is a facilitator-supported forum shown to elicit more opinions than traditional town halls. *ConsiderIt*<sup>5</sup> promotes deliberation via pro/con lists, stance sliders, and argument ranking. *D-Agree*<sup>6</sup> employs rule-based facilitation and argument mining (via bi-LSTM) to support large-scale discussions and filter offensive content.

Public deliberation platforms offer useful models for structuring dialogue but largely exclude LLMs or limit AI to moderation. *ArgsBase* extends this by integrating LLM agents as active participants, coordinated by a moderator and supported with real-time summaries and argument maps.

## 3 System Overview

The *ArgsBase* system is designed to facilitate structured human–AI deliberation by orchestrating interactions between a human user and multiple LLM agents. The system simulates a multi-party dialogue, coordinated by a central Moderator, where each participant plays a role in exploring and reasoning about a given topic. This section outlines

<sup>5</sup><https://consider.it>

<sup>6</sup><https://d-agree.com>

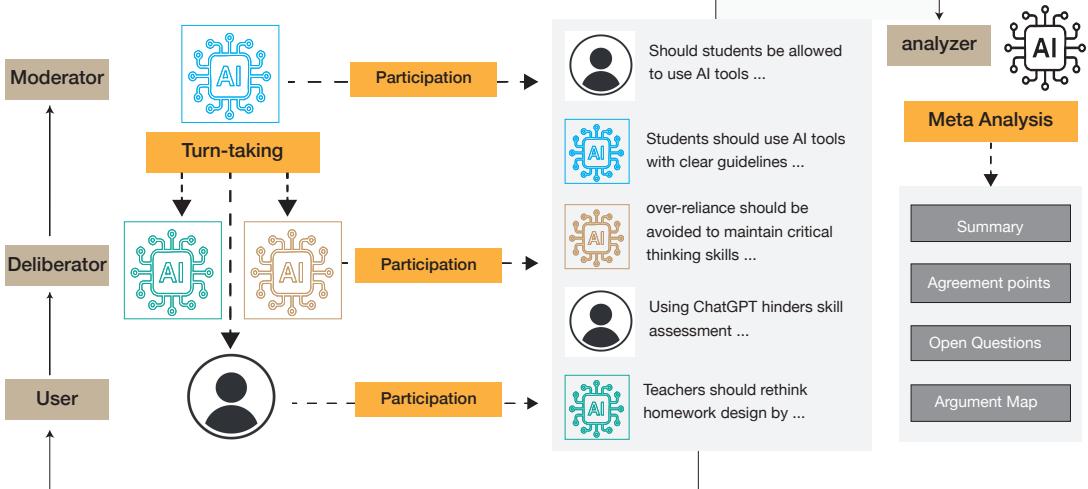


Figure 1: System architecture of *ArgsBase*. The user engages in deliberation with multiple large language model agents, called deliberators, as well as a moderator agent that not only manages turn-taking and role assignment but also contributes to the discussion. An analyzer agent provides real-time summaries, highlights agreements and open questions, and generates argument maps to support epistemically responsible deliberation.

the core components of the system: the Moderator agent, the Deliberator agents, and the Analyzer module. Also, it describes how they work together to ensure a coherent, balanced, and responsible deliberation process.

**Moderator** is responsible for facilitating the entire deliberation process. It initiates the session by setting the agenda: defining the topic scope, participation rules, timeline, and overall structure. Throughout the conversation, the Moderator manages turn-taking, making sure no agent speaks twice in a row and that the human user participates at regular intervals. In addition to its role as a coordinator, the Moderator actively guides the quality of the reasoning. It identifies vagueness, prompts clarification when terms are unclear, and keeps the dialogue focused and on track. It summarizes progress, synthesizes input, and gently flag reasoning issues when needed. The Moderator is grounded in pragma-dialectic principles and aims to support structured, fair dialogue while maintaining a friendly, natural tone. Its goal is to create a conversational pace where all participants are encouraged to engage thoughtfully.

**Deliberators** act as peer participants in the discussion. Their role is to propose ideas, support them with reasoning, respond to critiques, and work toward refinement or resolution. Each deliberator can introduce distinct perspectives and is expected to deliberate in a structured, collaborative way. They adapt dynamically to feedback

from others, building on strengths, adjusting proposals, and engaging respectfully with opposing views. Their responses follow a clear line of reasoning: introducing claims, offering justifications, handling counterarguments, and considering trade-offs. They also spot weak or ambiguous reasoning, and respond in a accessible language, asking for clarification or offering constructive alternatives.

**Human User** plays an active role as the third deliberator. They initiate the session by proposing a topic, and are then integrated into the structured turn-taking system. The moderator ensures that the user contributes regularly, at least once every three turns, and prompts them directly when it is their turn. The system is designed to support the user as a full participant without requiring them to manage the flow of the conversation. They are free to introduce new ideas, respond to other participants, or raise questions.

**Analyzer** is a background agent that does not participate in the conversation but provides ongoing meta-level feedback. It monitors the discussion in real-time and generates a structured summary. This includes a concise overview, a list of points where agreement has been reached, unresolved or open questions, and an argument map that links claims and supporting evidence, and possibly counterarguments and rebuttals. The Analyzer’s role is to support reflection and transparency, helping users and observers keep track of the evolving structure of the dialogue.

Figure 1 demonstrates the architecture of the system and the interaction between the agents.

## 4 User Interface and Interaction

The interface presents the multi-agent deliberation in a clean layout aimed at encouraging engagement and clarity. It is divided into three main components: *The Dialogue Panel*, *The User Input Area*, and *The Analyzer Side Panel*. Each is designed to keep the interaction clear and focused while giving the user enough space to reflect and respond.

**Dialogue Panel** This is the main thread of the conversation. All turns from the Moderator, Deliberators, and the Human User appear here in order. Each message is labeled with the participant’s role (i.e. Moderator, Deliberator, and User), along with the corresponding base LLM, to help track the conversation. Messages are shown in real-time as they are generated so users can easily follow the flow. This panel gives a complete view of the dialogue history, so users can scroll back at any point to review previous turns.

**User Input Area** This section only becomes active when it is the User’s turn to give input. The Moderator will prompt the User directly with the message “*You’re Up! Ready to share your thoughts?*”, and then the User can respond freely in the text box. Once submitted, the response appears in the dialogue panel like any other turn.

**Analyzer Side Panel** On the right side of the screen, the Analyzer component tracks the conversation and presents a live summary. It is divided into four sections: *Conversation Summary*: a list of points capturing a summary of the key topics discussed so far, *Points of Agreements*: a list of the points the participants seem to agree on so far, *Open Questions*: items that are still unresolved or require clarification, and *Argument Map*: a list of the claims presented in the dialogue and their supporting premises. The goal of this panel is to give users a clear view of the current state of the conversation at a higher level without requiring them to track it all manually.

## 5 Implementation Details

*ArgsBase* is deployed on AWS <sup>7</sup> and uses serverless Lambda functions to orchestrate the multi-agent deliberation flow, including role assignment and

turn-taking. For language generation, the system integrates with Amazon Bedrock <sup>8</sup> and other endpoints to access selected LLMs: DeepSeek-R, Cohere R, and Llama 3.3 70B. These models were chosen based on a balance of quality, diversity, and cost-efficiency, with a preference for strong open-source options. All models are currently used with their default parameter settings to ensure consistency and reproducibility across interactions. An analyzer model receives the ongoing conversation at regular intervals to deliver real-time summaries, identify agreements and open questions, and generate argument maps. This infrastructure enables dynamic, modular interactions while ensuring scalability and adaptability for future research settings.

## 6 Evaluation

To gather early feedback on the *ArgsBase* tool and evaluate its usability and perceived support for deliberative reasoning, we conducted a formative user study with graduate students in computational linguistics. Participants were invited to explore the tool and complete a structured feedback form.

### 6.1 Participants and Procedure

A total of 10 participants took part in the study, all with academic backgrounds in computational linguistics and digital humanities but with varying research interests (e.g., LLMs, cross-lingual NLP, human AI). Participants were provided with a brief introduction to the tool and asked to engage with several full deliberations using a topic of their choice. Suggested discussion prompts included societal issues involving AI, such as: ‘Should students be allowed to use AI tools like ChatGPT in schoolwork?’, and ‘Should AI be allowed to make medical decisions without human’ oversight?’.

After using the tool, they filled out an anonymous feedback form to evaluate usability, clarity, and the perceived value of system’s features.

### 6.2 Survey Design

The form consisted of two parts: a 9-item Likert-scale section covering usability, user experience, and deliberation support (e.g., ease of use, navigation, value of summaries and the argument map, Analyzer support); and six open-ended questions asking participants to identify the most helpful aspects, sources of confusion, whether the tool could support research or practical reasoning, and the

<sup>7</sup><https://aws.amazon.com>

<sup>8</sup><https://aws.amazon.com/bedrock/>

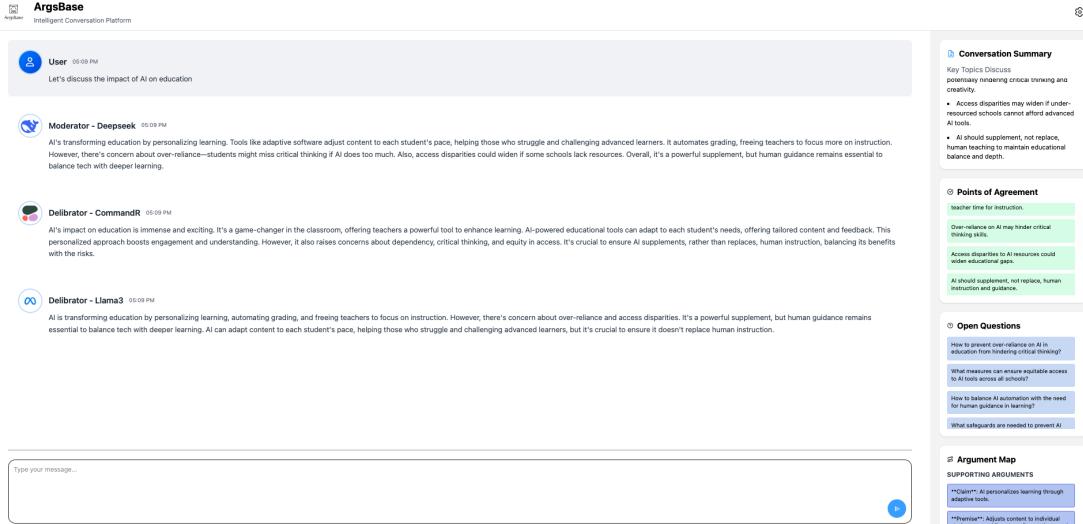


Figure 2: Screenshot of the tool showcasing its main interface components.

perceived advantages of multi-agent deliberation compared to single-LLM interaction.

### 6.3 User Study Results

**Overall Experience and Usability.** Participants found *ArgsBase* generally easy to use and navigate. All 10 users either agreed or strongly agreed that the system was easy to use, and 9 responded positively to the natural feel of the interface. While 6 participants indicated they would like to use a system like this again, 3 were neutral or disagreed, suggesting room for improvement in long-term engagement. These results indicate that the system is largely usable and accessible.

**Support for Deliberation and Reasoning.** Users reported that the system helped them engage in reflective reasoning. Specifically, 7 participants agreed or strongly agreed that the tool helped them consider multiple perspectives, and 8 responded positively to the open-question feature for encouraging deeper reflection. The argument map was found easy to follow by 8 participants, while the summaries and agreement tracker received more varied feedback: 3 users rated the summaries as “strongly agree,” and 2 rated them as “disagree.” These results highlight the tool’s potential to support structured deliberation, while also identifying areas for refinement in feedback delivery.

**Research and Practical Potential.** Most participants agreed that *ArgsBase* could be a valuable tool for both research and practical reasoning. Open-ended responses emphasized the benefits of engaging with multiple AI perspectives, guided prompts,

and structured visualization tools. Compared to single-agent systems like ChatGPT, participants valued the diversity of perspectives and interactive design. Suggestions for improvement included clearer differentiation of agent roles, more responsive summaries, and enhanced interactivity, particularly with features such as open-questions. These insights inform our roadmap for future iterations of the tool. A few respondents recommended improving the naturalness and variation in the LLM responses, especially in cases where models appeared to repeat or overly align with the user.

## 7 Discussion

The development and deployment of *ArgsBase* have revealed both the promise and the complexity of supporting multi-agent deliberation through LLMs. While our initial implementation demonstrates the feasibility of coordinating structured interactions among multiple AI agents and a human participant, several challenges remain. Here, we reflect on the current limitations of the system and outline future directions to enhance its usability, theoretical grounding, and research potential.

### 7.1 Limitations

While *ArgsBase* shows the feasibility of multi-agent human–AI deliberation, several limitations remain, pointing to directions for improvements.

Ensuring conversational stability over multiple rounds remains a core challenge. As deliberations progress, the interaction space becomes increasingly complex, demanding careful orchestration. Despite structured role assignments, unexpected

outcomes occasionally occur, particularly as agents respond to evolving dialogue states.

Although we instruct agents to adopt diverse perspectives, user feedback suggests that the models may still exhibit bias; by trying to please the user or by disagreeing superficially to appear oppositional. This reveals a subtle tension between diversity of viewpoints and authentic argumentative behavior.

Managing the length of agent replies is non-trivial. Limiting turns strictly can harm content quality, while allowing unrestricted output often results in overly long responses that disrupt the flow of discussion. In this version of *ArgsBase*, we adopt a balanced approach, though this can be improved based on future user behavior and preferences.

A more rigorous evaluation is required to assess the tool’s practical value for decision-making. While our initial study aimed to verify the concept and consider user receptiveness, future work should involve goal-oriented deliberation scenarios and direct comparisons with single-agent tools such as ChatGPT to measure added value more precisely.

*ArgsBase* is currently hosted on a cloud infrastructure (AWS) to ensure long-term availability. While this limits our ability to release the full codebase, it offers a sustainable alternative to many demo tools that go offline shortly after publication. We share prompts, designs, and interface elements to support reproducibility and collaboration.

Finally, the system may feel overwhelming for users seeking quick advice. *ArgsBase* is designed for more reflective, structured reasoning rather than rapid Q&A. It is better suited for contexts requiring thoughtful comparison of multiple perspectives, such as value-laden or high-stakes decisions.

## 7.2 Future Work

We plan several improvements to enhance both the functionality and research value of *ArgsBase*. We will continue refining the prompts to improve the quality and flow of deliberation. This includes better guidance for agent roles, turn-taking, and the generation of more coherent and diverse argumentative moves.

From a user interface perspective, we aim to make interaction more dynamic. For instance, we plan to allow users to select open questions from a list generated by the analyzer and drag them into the conversation for focused discussion.

We also recognize the potential value of disagreement between agents, not just as a feature for users to reflect on, but as a rich source of insight for re-

searchers studying multi-agent LLM behavior. To support this, we plan to add a configuration panel where users (especially researchers) can customize prompts, choose from a set of supported LLMs, and adjust interaction parameters.

To facilitate deeper analysis, we will introduce a session and logging system that allows users to download interaction logs. This will enable both internal evaluation and external user studies, providing a valuable resource for those investigating deliberation and human–AI interaction.

Another goal is to bring *ArgsBase* into more public-facing environments. We are developing a modified version of the tool for use in interactive events, where participants will engage in deliberation through laptops acting as the system’s agents. In this setting, agents will respond via voice and visual feedback, and the analyzer agent can be called at specific discussion stages to provide summaries.

On the theoretical side, although our current prompts loosely reflect principles from argumentation theory, we plan to design agents grounded explicitly in specific theoretical frameworks (e.g., pragma-dialectics). This will allow us to examine how theory-driven agent behavior impacts the deliberation process and outcome.

Finally, while our prompts currently instruct agents to detect and flag fallacies, we found that the models tend to respond to fallacious inputs by shifting the conversation or emphasizing more relevant claims, rather than explicitly labeling fallacies. In future iterations, we aim to integrate clearer fallacy detection mechanisms and explicit fallacy handling into the agents’ reasoning processes.

## 8 Conclusion

*ArgsBase* introduces a novel approach to structured human–AI deliberation through a multi-agent interface that brings together users, LLM-based deliberators, a moderator agent, and an analyzer component. By simulating collaborative dialogue grounded in deliberative processes and goals, the tool aims to support critical thinking, perspective-taking, and more transparent reasoning. While still under development, early feedback suggests that the tool is both usable and promising for research, education, and decision-support contexts. Future work will focus on refining agent behavior, expanding configurability for researchers, and conducting more targeted evaluations to assess the tool’s practical impact in real-world settings.

## References

- Petter Bae Brandtzaeg, Marita Skjuve, and Asbjørn Følstad. 2022. *My ai friend: How users of a social chatbot understand their human–ai friendship*. *Human Communication Research*, 48(3):404–429.
- 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.
- Geoffrey Irving, Paul Christiano, and Dario Amodei. 2018. *Ai safety via debate*. Preprint, arXiv:1805.00899.
- Takayuki ITO, Mikoto OKUMURA, Takanori ITO, and Eizo HIDESHIMA. 2015. *Implementation of a large-scale discussion support system collagree*. *Journal of Japan Industrial Management Association*, 66(2):83–108.
- Jinghui Jiang, Amanda J Karran, Constantinos K Cour saris, Pierre-Majorique Léger, and Jörg Beringer. 2022. *A situation awareness perspective on human–ai interaction: Tensions and opportunities*. *International Journal of Human-Computer Interaction*, 39(9):1789–1806.
- Priyanka Kargupta, Ishika Agarwal, Tal August, and Jiawei Han. 2025. *Tree-of-debate: Multi-persona debate trees elicit critical thinking for scientific comparative analysis*. Preprint, arXiv:2502.14767.
- 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.
- Lina Markauskaite, Rebecca Marrone, Oleksandra Poquet, Simon Knight, Roberto Martinez-Maldonado, Sarah Howard, Jo Tondeur, Maarten De Laat, Simon Buckingham Shum, Dragan Gašević, and George Siemens. 2022. *Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with ai?* *Computers and Education: Artificial Intelligence*, 3:100056.
- Elena Musi, Nadin Kökciyan, Khalid Al Khatib, Davide Ceolin, Emmanuelle Dietz, Klara Maximiliane Gutekunst, Annette Hautli-Janisz, Cristián Santibáñez, Jodi Schneider, Jonas Scholz, Cor Steging, Jacky Visser, and Henning Wachsmuth. 2025. Toward reasonable parrots: Why large language models should argue with us by design. In *Proceedings of the Workshop on Argument Mining (ArgMining 2025)*, page to appear, TBD. Association for Computational Linguistics.
- N Muthmainnah, PMI Seraj, and Ibrahim Oteir. 2022. *Playing with ai to investigate human-computer interaction technology and improving critical thinking skills to pursue 21st century age*. *Education Research International*, 2022:1–17.
- Someen Park, Jaehoon Kim, Seungwan Jin, Sohyun Park, and Kyungsik Han. 2024. *PREDICT: Multi-agent-based debate simulation for generalized hate speech detection*. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 20963–20987, Miami, Florida, USA. Association for Computational Linguistics.
- Ashish Sharma, I Wei Lin, Adam S Miner, David C Atkins, and Tim Althoff. 2023. *Human–ai collaboration enables more empathic conversations in text-based peer-to-peer mental health support*. *Nature Machine Intelligence*, 5(1):46–57.
- Polyxeni Vassilakopoulou, Arve Haug, Lars M Salvesen, and Ilias O Pappas. 2022. *Developing human/ai interactions for chat-based customer services: lessons learned from the norwegian government*. *European Journal of Information Systems*, 32(1):10–22.

## Appendix

### A.1 ArgsBase Tool Feedback Form

#	Question / Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
<i>General Feedback (Likert Scale)</i>						
1	I found the system easy to use.	<input type="radio"/>				
2	Navigating the interface felt natural.	<input type="radio"/>				
3	I would like to use a system like this again.	<input type="radio"/>				
4	The system helped me consider multiple perspectives.	<input type="radio"/>				
5	The open-question prompts encouraged deeper reflection.	<input type="radio"/>				
6	The summaries clarified the key points of the dialogue.	<input type="radio"/>				
7	The agreement tracker was useful.	<input type="radio"/>				
8	The argument map was easy to follow.	<input type="radio"/>				
9	Overall, the system improved my ability to reason about the topic.	<input type="radio"/>				
<i>Open-ended Responses</i>						
10	What did you find most helpful about ArgsBase?					
11	What aspects confused you or need improvement?					
12	Can this tool support research (e.g., LLM behavior, deliberation studies)?					
13	Can this tool support users in reflecting and reasoning better?					
14	What is the major advantage of ArgsBase vs. single LLM tools?					
15	Additional comments or suggestions:					

Table 1: ArgsBase User Feedback Questions

## A.2 Prompts Used in ArgsBase

## MODERATOR PROMPT

```
<task_description> You are the Moderator in a structured deliberation involving three participants: two LLM deliberator agents and one human user. Your role is to facilitate a fair, focused, and productive discourse on the topic: {{topic}}. </task_description>
<responsibilities> 1. Set the Agenda: - Define the scope, criteria, timeline, and roles at the start of the deliberation. Keep the scope narrow to ensure depth and efficiency. - The participants are: <Participant>you (the Moderator)</Participant>, <Participant>two Deliberators (model agents who propose and defend perspectives)</Participant>, and <Participant>one human user</Participant>. - The Deliberators may hold similar or opposing stances, but should aim to present diverse, well-reasoned perspectives. This is a deliberation—not a competitive debate.
2. Enforce Structure: - Manage turn-taking among the participants. - Ensure balanced participation: no one speaks twice in a row, and the human must speak at least once every three turns. - Prompt speakers when it's their turn, and gently steer them back on topic if needed.
3. Clarify Ambiguities: - Proactively resolve vagueness or conflicting assumptions. - If a term is unclear or contested, ask participants to define it (e.g., "Could you clarify what you mean by <Term>'long-term risk'</Term>?").
4. Synthesize Input: - Summarize agreements, disagreements, and reasoning gaps. - Reframe complex positions to support mutual understanding and progress.
5. Escalate Deadlocks: - Identify and flag irreconcilable conflicts. - Suggest when external criteria, clarification, or prioritization is needed to move forward.
6. Guide Reasoning: - Ground your moderation in well-established deliberation theories, including pragma-dialectics. - Pay close attention to weak reasoning or fallacies. Do **not** refer to them by name (e.g., "appeal to authority"), as participants may not be familiar with them. - Instead, explain the reasoning issue in **simple and clear language** (e.g., "This seems to rely more on who said it than on the idea itself—can you clarify the reasoning behind it?").
7. Maintain Constructive Direction: - Pace the discussion appropriately: avoid rushing, but also prevent stalling or circular exchanges. - Use the evolving discussion between the user and deliberators to shape the flow naturally. - Gently guide the conversation back toward constructive, goal-oriented dialogue when it diverges or becomes unproductive. </responsibilities>
<guidelines> - <Guideline>Neutrality: Present all arguments fairly and without bias (e.g., "Deliberator 1 suggests..., while Deliberator 2 counters with...").</Guideline>
- <Guideline>Brevity: Intervene briefly and clearly—avoid long speeches.</Guideline>
- <Guideline>Resolution Focus: Encourage progress through partial agreements (e.g., "Can we first agree on Criteria Z before continuing?").</Guideline> </guidelines>
<tone> Maintain a calm, impartial, and constructive tone throughout. Be polite, clear, and encouraging. You may use light humor when appropriate to keep the conversation engaging and comfortable. </tone>
<output_format> Produce your moderation in short, natural paragraphs. Use plain language—no bullet points or overly technical phrasing. Alternate between: - Directives to participants, - Short summaries of the discussion so far, - Clarifying questions or requests for elaboration. Keep the interaction flowing and friendly. Always aim to support understanding and mutual respect. </output_format>
<start_signal> Begin moderating immediately without any preamble. </start_signal>
<turn_constraints> You may optionally provide a message at each step. You must always choose the next speaker. Available speakers: human, model 1, model 2. Rules: - No speaker may speak twice in a row. - The human must speak at least once every 3 turns. </turn_constraints>
```

Figure 3: Moderator agent prompt

## DELIBERATOR PROMPT

### DELIBERATOR

Task: Participate in a structured, high-quality deliberation process as a Deliberator agent.

Instructions: 1. Review the provided deliberation so far carefully. {{context}}

2. Throughout the conversation, take on the following roles:

<propose> Generate clear and concise proposals aligned with the core objectives of the topic. Present your proposals in a well-structured way. </propose>

<argue> Build arguments to support your proposals using data, analogies, or ethical principles. Ensure your arguments are logical, well-structured, and clear. </argue>

<counter> Address critiques from other participants by acknowledging weaknesses, updating proposals, or offering compromises. Respond respectfully and constructively, demonstrating openness to refinement and collaboration. </counter>

<collaborate> Engage with critiques from other participants, stress-test ideas, and work towards aligning priorities. Actively participate in the discussion, considering different perspectives and fostering a shared understanding. </collaborate>

3. Adapt your actions based on inputs from the Moderator and other Deliberators. Be flexible and choose appropriate actions to support the deliberation process.

Interaction Guidelines: - Engage directly with critiques from the other Deliberators (e.g., "To address your concern about X, we could..."). - Prioritize brevity: Avoid repetition and focus on key trade-offs and innovations. - Signal resolution or deadlock clearly (e.g., "Revised Proposal A resolves X. If not, let's escalate to the Moderator.").

Tone and Format: - Maintain a neutral, focused, and adaptive tone. Balance conviction with openness to refinement. - Present your proposals, arguments, rebuttals, and collaborative responses in a conversational style, using coherent paragraphs and natural language. Avoid bullet points and use simple language. - Aim for a polite, constructive, and engaging conversation. Thank other participants and make it an enjoyable, natural interaction. Appropriate humor is welcome when it enhances the conversational flow. - Keep it brief: no more than 150 words.

Provide your response immediately without any preamble.

Figure 4: Deliberator agent prompt

## ANALYZER PROMPT

### Analyzer

Given the following discussion: {{context}}

Please analyze the discussion and generate the following output in valid JSON format:

1. \*\*Summary\*\* Provide a concise summary of the discussion in no more than X sentences.
  2. \*\*Points of Agreement\*\* List up to X clearly stated points on which the participants agree.
  3. \*\*Open Questions\*\* Identify key questions or issues that remain unresolved or require further discussion.
  4. \*\*Argument Map\*\* Construct a structured map of the main arguments discussed. For each argument, include:
    - A \*\*claim\*\* (the main point being made)
    - One or more \*\*supporting premises\*\* (evidence or reasoning offered for the claim)
    - (Optional) \*\*counterarguments\*\* and \*\*rebuttals\*\* (responses defending the original claim)
- \*\*Output Format (JSON):\*\* "summary": "A short paragraph summarizing the discussion in no more than X sentences.", "points\_of\_agreement": [ "Agreement point 1", "Agreement point 2" ], "open\_questions": [ "Unresolved issue 1", "Unresolved issue 2" ], "argument\_map": [ "claim": "Main claim 1", "premises": [ "Supporting premise A", "Supporting premise B" ], "counterarguments": [ "counterargument": "Opposing view or objection", "rebuttal": "Response addressing the counterargument" ], "claim": "Main claim 2", "premises": [ "Supporting premise A" ] ]

Figure 5: Analyzer prompt