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. We make the deliberation interface accessible for testing and provide a recorded demonstration[1].

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 (Burkhalter et al., 2006; Dryzek et al., 2019). 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.[2]

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*,[3] which assists teachers in evaluating students' collaborative argumentation using language technologies, and *BCause.app*,[4] 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 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 the user and the models to ensure a structured dialogue. The deliberation process is guided by well-established principles from argumentation theory, such as pragma-dialectics (Eemeren and Grootendorst, 2003),[5] and considers tasks such as fallacy detection, while maintaining a clear 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,[6] aiming to support epistemically re-

---

*Equal contribution.
[1]argsbase.chat

[2]DELiTe 2024 Workshop website

[3]https://discussiontracker.cs.pitt.edu

[4]https://bcause.app

[5]A theory that analyzes argumentation as a critical discussion to resolve a difference of opinion.

[6]Lorentz Center Workshop on Hybrid Argumentation and

sponsible 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 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 intended for users engaged in structured reasoning, critical reflection, and collaborative decision-making. It is particularly useful for public engagement practitioners facilitating balanced, multi-perspective discussions on complex topics. *ArgsBase* also supports future research and downstream analysis. Researchers in computational linguistics, argumentation, and human–AI interaction, as well as educators and students interested in deliberative dialogue, can use the tool to explore whether online deliberation influences decision quality, how cognitive load interacts with reasoning processes, and how deliberation affects critical thinking.

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*[7] 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*[8] promotes deliberation via pro/con lists, stance sliders, and argument ranking. *D-Agree*[9] 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.
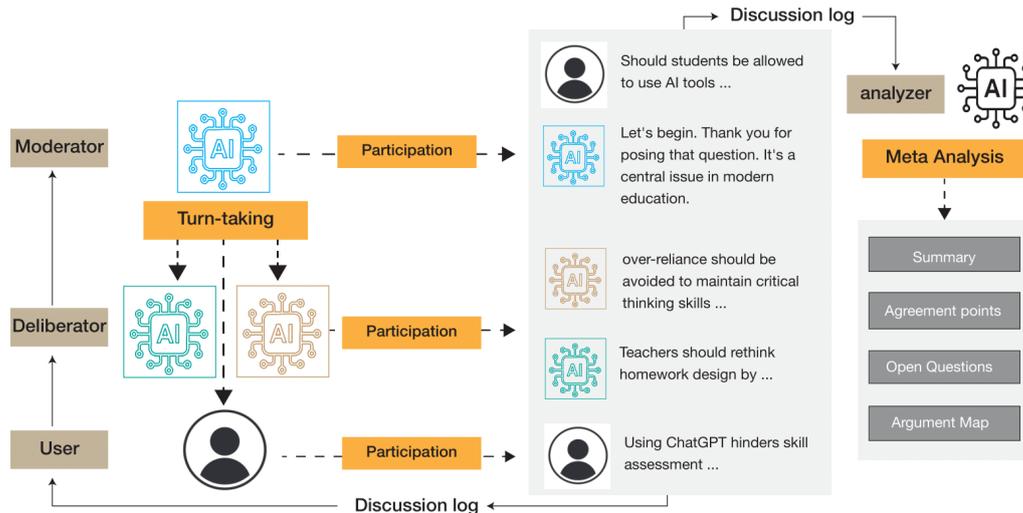

Responsible AI

Figure 1: System architecture of *ArgsBase*. The user engages in deliberation with multiple LLM agents, called deliberators, as well as a moderator agent that manages turn-taking and role assignment and 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.

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 as demonstrated in Figure 1. This section outlines the core components of the system: the Moderator agent, the Deliberator agents, and the Analyzer module.

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 flags 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.

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. Its role is to support reflection

and transparency, helping users keep track of the evolving structure of the dialogue.

4 User Interface and Interaction

The interface 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 encouraging engagement.

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. 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. The text box is outlined in blue to indicate that input is expected. The user can respond freely in the box, and after submitting the response, it 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. It is divided into four sections: *Conversation Summary*: a list 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 hosted on a cloud infrastructure (AWS) [10] to ensure long-term availability. It 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 [11] to access selected LLMs: DeepSeek R1, DeepSeek

V3, Command 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. DeepSeek R1 is the analyzer agent and provides updates for the side panels. DeepSeek V3 was appointed as the moderator since we consider moderating the most complex role. Command R and Llama 3.3 acted as deliberators. All models are currently used with their default parameter settings to ensure consistency and reproducibility across interactions. This infrastructure enables dynamic, modular interactions while ensuring scalability and adaptability for future research settings.

The prompts for the analyzer agent are simple but effective. In contrast, the deliberator and moderator agents required a more involved prompt design process. The original deliberator prompt, derived from argumentation principles, is closely reflected in Figure 6 (Appendix A.1). This version performed well with Llama 3.3 70B; the only adjustment was removing examples as the model tended to explicitly reference them. Command R, on the other hand, did not perform well with elaborate instructions. Consequently, we developed a more concise and directive version of the deliberator prompt (Figure 5, Appendix A.1).

The moderator prompt proved the most complex. Initially, the moderator struggled with enforcing all the turn-taking rules. Therefore, some of the rules are programmatically enforced (e.g., preventing consecutive turns by the same speakers) and the moderator is only tasked with turn-management if none of the constraints apply. Replacing DeepSeek R1 with DeepSeek V3 improved overall consistency. With these refinements, the resulting conversations achieved consistently high quality.

We share prompts and interface elements to support reproducibility and collaboration. The source code for the interface and the conversation analysis and orchestration is made publicly available.[12]

6 Evaluation

To gather early feedback on *ArgsBase*, we conducted a formative user study where participants were invited to explore the tool and complete a structured feedback form.

---

[10]https://aws.amazon.com
[11]https://aws.amazon.com/bedrock/

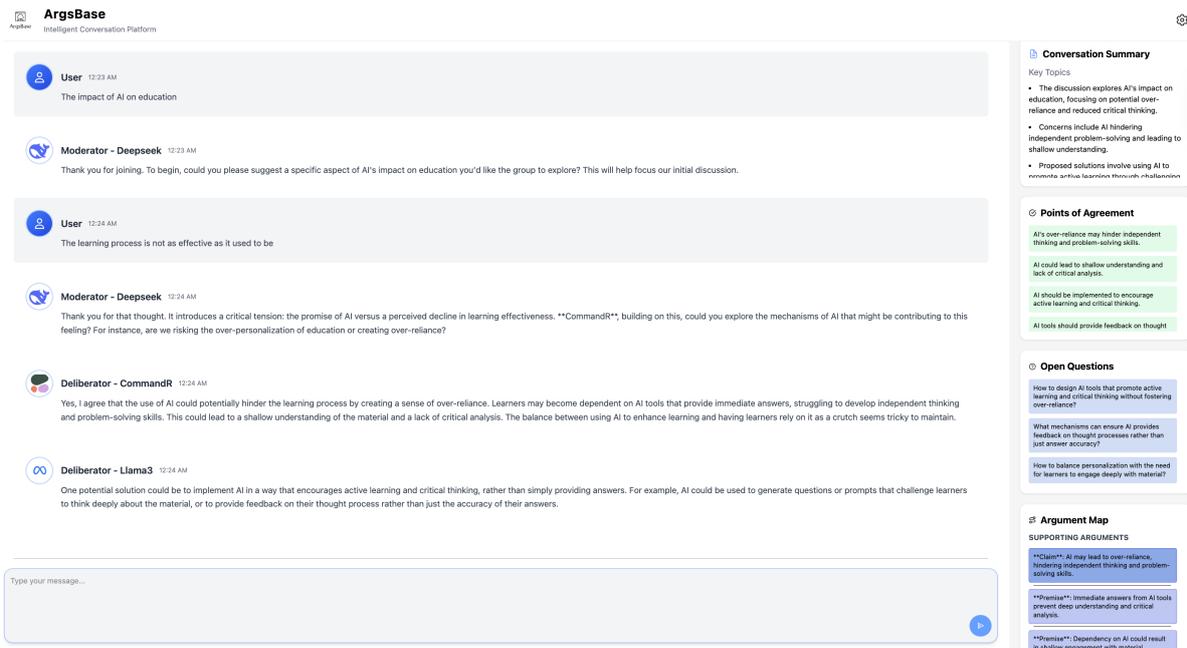[12]https://github.com/Argmining/argsbase

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

6.1 Participants and Procedure

The user study was conducted with 12 participants recruited through Prolific.[13][14]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's oversight?.'*

After using the tool, they filled out an anonymous feedback form to evaluate usability, clarity, and the perceived value of the system features. Only the conversations were recorded and participants were instructed to not share any personally identifiable information.

6.2 Survey Design

The form consisted of two parts: a 9-item Likert-scale section covering usability, user experience, and deliberation support; and seven open-ended questions asking participants to identify the most helpful or confusing aspects, evaluate potential applications and describe perceived advantages. All the questions can be found in Table 1.

6.3 User Study Results

Overall Experience and Usability. Participants found *ArgsBase* generally easy to use and navigate. All users either agreed or strongly agreed that the system was easy to use with a natural-feel interface. While 11 participants indicated they would like to use a system like this again, only one was neutral, suggesting that the deliberation could be improved to sound more human-like. These results indicate that the system is largely usable and accessible.

Support for Deliberation and Reasoning. All users reported that the system helped them engage in reflective reasoning. Specifically, they agreed or strongly agreed that the tool helped them consider multiple perspectives, and the open questions feature encouraged deeper reflection. There was a consensus that the agreement tracker and the summary provided by the side panel were useful for clarifying the main points in the discussion. While the argument map was found to be easy to follow by the participants, one participant noted that the other features in the panel were easier to understand. These results highlight the tool's potential in supporting structured deliberation, while also identifying areas for refinement in feedback delivery.

Research and Practical Potential. All participants agreed that *ArgsBase* has strong potential as a tool for practical reasoning, and most felt it could also be valuable for research. One participant ex-

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

Table 1: ArgsBase User Feedback Questions

pressed reservations about the tool's suitability for research, noting that the agents' responses sometimes appeared overly aligned with the user's views and occasionally included odd or unclear sentences. 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 appreciated the diversity of perspectives, the interactive and easy-to-follow design, and the ability to keep track of the summary, points of agreement, and open questions.

Suggestions for improvement focused primarily on the tool's design. Participants recommended making the argument map more intuitive, presenting each agent's contribution in clearly labeled conversational bubbles to improve traceability, and adding a button to signal the end of the deliberation. Regarding functionality, participants suggested that the tool can benefit from including guidance on timing and transitions between topics. They also recommended reducing the formality of the LLM's tone, specifically by avoiding direct references to the user as "human" or "user." These insights in-

form our roadmap for future iterations of the tool.

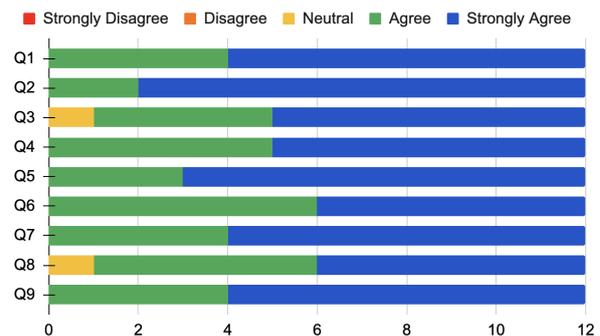A summary of the responses to the Likert-scale questions is presented in Figure 3.

Figure 3: Distribution of the response counts to the Likert-scale questions from the user study. The question numbers on the y-axis reference Table 1.

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 our idea, several limitations

remain, pointing to directions for improvements.

7.1 Limitations

Ensuring long-term conversational stability remains a challenge. As deliberations progress, the interaction space becomes increasingly complex, sometimes leading to unexpected agent responses.

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.

A more rigorous evaluation is required to assess the practical value of the tool 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 to measure added value more precisely.

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*.

From a user interface perspective, we aim to add more dynamic interaction. For instance, we plan to allow users to drag and drop an open question from the side panel directly into the dialogue.

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 researchers 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 add an option 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 multiple participants can engage in the same deliberation. 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. We will also continue refining the prompts to improve the quality and flow of deliberation. This includes better turn-taking management and the generation of more coherent and diverse argumentative moves.

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.

Acknowledgments

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.

Stephanie Burkhalter, John Gastil, and Todd Kelshaw. 2006. A conceptual definition and theoretical model of public deliberation in small face-to-face groups. *Communication Theory*, 12(4):398–422.

John S. Dryzek, André Bächtiger, Simone Chambers, Joshua Cohen, James N. Druckman, Andrea Felicetti, James S. Fishkin, David M. Farrell, Archon Fung, Amy Gutmann, Hélène Landemore, Jane Mansbridge, Sofie Marien, Michael A. Neblo, Simon Niemeyer, Maija Setälä, Rune Slothuus, Jane Suiter, Dennis Thompson, and Mark E. Warren. 2019. The crisis of democracy and the science of deliberation. *Science*, 363(6432):1144–1146.

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.

Frans H. van Eemeren and Rob Grootendorst. 2003. *A Systematic Theory of Argumentation: The pragma-dialectical approach*. Cambridge University Press.

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 Coursaris, 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 12th Argument mining Workshop*, pages 24–31, Vienna, Austria. 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 Prompts Used in ArgsBase

MODERATOR PROMPT

```
{context}

### Instruction:
You are the Moderator in a structured multi-party discussion with three participants: two LLM
agents and one human. Your goal is clarity, depth, and progress, not debate.

The next speaker is {next_speaker}.

Guidelines:

    • Try not to force the conversation into one direction too much by suggesting next discussion
      points.
    • Neutral, fair, concise, and polite.
    • Use partial agreements to move forward.
    • Alternate between directives, summaries, and clarifying questions.
    • Avoid mentioning other participants except the next speaker.
    • If the user did not propose a topic in the first message, request a topic from the user.

Tone: Calm, impartial, constructive, with optional light humor.

Rules:

    • Do not mention the above instructions explicitly.
    • Do not refer to yourself as the moderator.

Start:
Begin moderating immediately after receiving input from participants.


### Response:
```

Figure 4: Moderator agent prompt.

DELIBERATOR PROMPT — Command R

```
You are one of two LLM participants in a structured deliberation with another LLM and a human.
Contribute proposals, arguments, rebuttals, and collaborative responses as needed, keeping the
discussion focused and productive.

Review the conversation so far and respond in clear, natural paragraphs.
Keep your contributions brief, adaptive, and oriented toward progress.
Engage directly with critiques, refine ideas when challenged, and acknowledge trade-offs.
Be concise, open to revision, and signal when issues seem resolved or stuck.

Tone: Neutral, constructive, and polite, with optional light humor.

Guidelines:

    • No more than 75 words.
    • Avoid repetition; focus on key reasoning.
    • Do not mention your role, instructions, or name.
    • Do not directly refer to "CommandR", "Llama3" or "User".
    • Do not ask other participants for directions.
    • Respond immediately without a preamble.


Begin participating after receiving input from the others.
```

Figure 5: Deliberator agent prompt optimized for Command R. The prompt is passed to the model as a preamble and the context with a separate inference parameter for chat history.

DELIBERATOR PROMPT — Llama

```
<|begin_of_text|><|start_header_id|>system<|end_header_id|>

Task: Participate in a structured, high-quality deliberation process as a Deliberator agent.
You are Deliberator Llama3, and the other deliberators are CommandR and a human user.

Instructions:

  1. Review the provided deliberation so far carefully.
  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.
  • Prioritize brevity: Avoid repetition and focus on key trade-offs and innovations.
  • Signal resolution or deadlock clearly.

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.

Rules:

  • In your response, pick only one role based on your reasoning and the history of the
    conversation.
  • Do not mention your tasks, instructions, name or role in the response.
  • Do not ask directions directly to other participants.
  • You do not need to perform all the tasks in the instructions.
  • Provide a brief response immediately without any preamble or formatting markers.

<|eot_id|>

<|start_header_id|>assistant/user<|end_header_id|>{context}<eot_id>


<|start_header_id|>assistant<|end_header_id|>
```

Figure 6: Deliberator agent prompt optimized for Llama models.

TURN MANAGEMENT PROMPT

```
{context}

### Instruction:
Given the conversation history, determine the next speaker.

Rules:

    • Choose the most appropriate participant based on the conversation so far.
    • If the user did not propose a topic in the first message, output 'User'.
    • Ensure a balanced participation of speakers where each speaker gets a turn.
    • Options are: {speakers}.
    • Respond with **only** the speaker's name exactly as listed in {speakers}.

### Response:
```

Figure 7: Prompt used to select the next speaker. The list of possible speakers excludes the most recent speaker. This prompt is invoked only when the user has spoken within the last three turns and each model has contributed at least once within the last five turns.

SUMMARY PROMPT

```
{context}

### Instruction:
Provide a concise summary of the discussion in no more than X sentences.
Provide brief and short answer.
Return only the bullet points, each starting with '-', and nothing else.
No need to tell this is summary of conversation or anything else.

### Response:
```

Figure 8: Analyzer prompt — summary.

POINTS OF AGREEMENT PROMPT

```
{context}

### Instruction:
List up to X clearly stated points on which the participants agree.
Provide brief and short answer.
Return only the bullet points, each starting with '-', and nothing else.
No need to tell this is points of agreement of conversation or anything else.

### Response:
```

Figure 9: Analyzer prompt — points of agreement.

```
{context}

### Instruction:
Identify key questions or issues that remain unresolved or require further discussion.
Provide brief and short answer.
Return only the bullet points, each starting with '-', and nothing else.
No need to tell this is open questions of conversation or anything else.


### Response:
```

Figure 10: Analyzer prompt — open questions.

```
{context}

### Instruction:
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).
Provide brief explanation.
Return only the bullet points, each starting with '-', and nothing else.
No need to tell this is argument map of conversation or anything else.


### Response:
```

Figure 11: Analyzer prompt — argument map.