Postdoctoral Project for the PEPR Research Program eNSEMBLE

PC4 CONGRATS - Large-scale collaboration

Caren Al Anaissy

1 Acronym and title:

IFDOD Intelligent Facilitation of Deliberation in Online Debates

2 Host institution and proposed advisor

Host institution: LIP6, Sorbonne Université et CNRS (UMR 7606). **Proposed Advisor:** Nicolas Maudet, Professor in Computer Science at Sorbonne Université, campus Pierre et Marie Curie, head of the Multi-Agent Systems SMA team.

3 Context

Deliberative processes are crucial in forming opinions, decisions, and policies within society. Unlike persuasive debates, deliberation seeks to enhance the understanding of conflicts perspectives among stakeholders. During these discussions, the exchange of arguments helps clarify viewpoints, highlight divergent interests, and ultimately guide participants toward a consensus. Online debate plateforms are plateforms where web users can create debates, argue for or againt a specific topic by posting pro/con arguments, and vote on other users' arguments. There are numerous online debate plateforms, like Idebate's debatebase[1], Kialo[2], Debategraph[3] and CreateDebate[4]. Online collective discussions play an important role in enhancing the participants critical thinking, understanding the public opinion, they can help participants rationalise their thoughts in controversial topics and reduce the spread of fake news. Many arguments are raised in discussions on these platforms, and contain valuable information. Understanding the outcome of a debate would be beneficial for different stakeholders. For instance, it allows platform moderators to improve online discourse, by over sighting debates and identifying the ones that might spread misinformation and fake news, or foster a toxic environment. It also allows to better recommend content for users, and facilitates strategic planning and decision-making in e-democracy governments and diverse communities. However, reading all the arguments requires a considerable amount of time, some discussions might shift from a topic to another and some arguments might even be maliciously used to spread fake news. AI might help to moderate and curate such debates, but it has to be done in a fully intelligible and accoutable way. This project seeks to facilitate the deliberation process between participants in an online debate by developing an AI-based system that directly interacts with the participants and helps them navigate through an online debate, by pointing out to the key arguments that are affecting the final outcome of the debate, representing these arguments with their different interactions, evaluating their acceptability and predicting the final outcome. The AI-based system will also provide explanations for such predictions based on the user's move in the debate. This will not only help the participant to well understand the debate, but will also allow them to participate in an efficient way, by showing them the possible debate's outcomes based on their choices and interactions.

4 Scientific and technical objectives

The first objective is to develop an AI-based system that supports and manages effective deliberation in online debates, focusing on enhancing decision-making processes. The second objective is to ensure that the deliberation processes are transparent and explainable for different stakeholders, helping them understand how conclusions are reached. This increases the trust of the users, which also increases user engagement. We aim to evaluate the impact of AI-mediated deliberation on the decision quality, user satisfaction and trust in the process, by incorporating feedback loops that allow the system to adapt to cognitive and emotional states of participants based on their interactions, and by exploring dialogical communication and action adaptation strategies to improve cooperation between human participants and AI entities.

5 Approach / Challenges / Originality

Approach: We plan to implement an AI-based system that can adapt its role dynamically based on the needs of the debate. e.g. , it can serve as a mediator when conflicts arise, a fact-checker when inaccuracies are presented, and an encourager when participation is low. This allows for a collaboration-centric AI-based system that focuses on the mechanisms of dialogue and cooperation with the participants. In addition to the technical development, this project will explore the dynamics of team collaboration involving intelligent agents. This includes understanding and designing the interactions and communication protocols between human team members and the AI system. By considering the AI not only as a tool but also as a participant in the deliberative processes, the project aims to enhance cooperative decision-making and problem-solving by developing and testing interaction protocols and communication strategies that facilitate effective human-AI collaboration. We illustrate below our approach:

- **Argument Strength Assessment:** We plan to define and use computational argumentation techniques to evaluate the quality and strengths of arguments presented by participants including claims, evidence, and reasoning processes. We also plan to identify the key arguments that are affecting the final outcome of the debate, representing these arguments with their different interactions, evaluating their acceptability and inferring the debate outcome.
- **Feedback Provision:** Assesing the quality of arguments presented in the debate helps identify logical fallacies, unsupported claims, and biases in arguments without directly conversing. The system should provide real-time feedback to participants

[1] https://idebate.net/resources/debatabase [2] https://www.kialo.com/ [3] http://debategraph.org [4] https://www.createdebate.com

on how to strengthen their arguments. This might include suggesting additional evidence, pointing out logical fallacies, or recommending ways to clarify their points. This ensures that discussions are based on sound reasoning and relevant evidence. We plan to incorporate natural language processing (NLP) so that the system can interpret the intent behind participants' statements. Via a chatbot, the system communicates with participants. It can ask questions, make clarifications, and offer suggestions in a way that feels intuitive and engaging.

- **Conflict Resolution and Participation Encouragement:** The AI-based system should detect when the debate might be shifting into an unproductive conflict or when participants are talking past each other. It should help the participants stay focused on the debate's main subject, and help them in observing and understanding the path towards which the debate is heading. The system should suggest reframing of issues and introduce questions that guide participants back to constructive paths. The system monitors also the participation levels of all debate members and can encourage quieter participants to voice their opinions, ensuring a more balanced and inclusive dialogue.
- **Decision Support:** The system should help to structure decision-making processes, by creating e.g. polls or decision trees that participants can interact with to visualize the impact of their choices. The system should also be able to simulate various decision paths (mostly by using computational argumentation). This helps the participants to understand the potential consequences of the arguments they might present before actually presenting them.

Challenges: The main challenge would be to develop an AI-based system that allows for comprehensive and successful human-AI interaction and collaboration. It is essential to develop adaptive explanations and interactions that customize the level and type of explanation presented to the user based on their profile, background, expertise and feedback.

Positioning regarding the literature and Originality: The originality of this project lies in the use of computational argumentation, natural language processing and multi-agent systems to facilitate the deliberation process between the different parties in an online debate. Our project's originality also lies in its approach to advance the understanding of AI as an interactive and adaptive collaborator. **Argumentation theory:** There is an increasing interest in applying argument mining to online debates in order to detect the premises, claims and relations between the arguments (attacks, supports) [3]. IFDOD seeks to develop adequate semantics for online debates and to bridge between the argumentation semantics and the argumentation graphs that result either from annotation or from argument mining. [5, 2] will be employed to model the positive and negative interactions between arguments as well as the votes. This data will then be used to determine the acceptability of each argument and identify sets of arguments that can be collectively accepted, representing a coherent viewpoint. **NLP** [4] will be used to offer assistance in detecting manipulation by identifying duplicate arguments with argument similarity calculation [6], checking the relationships between arguments provided by users etc. NLP techniques will be used to detect logical fallacies in real-time. This helps in providing instant feedback to participants on how to strengthen their arguments. Also, utilizing NLP in chatbots allows for dynamic interaction with participants, offering suggestions and feedback based on the analysis of the ongoing discussion. A recent study [1] shows that using an AI chat assistant that provides suggestions to individuals in politically divisive conversations, and recommends real-time, context-aware, and evidence-based ways to rephrase messages, improves reported conversation quality and enhances commitment to democratic reciprocity in conversation. NLP can also detect shifts towards conflict or detect disengagement. This helps in moderating the discussion effectively. **Multi-agent systems** [7] theory will be used to design and analyze interaction protocols, as well as to detect possible manipulations (e.g. an organised group of users massively voting for the exact same arguments in a short time period).

6 Positioning regarding the literature as well as the priority themes of the PC3

Our project, IFDOD, directly aligns with the PC3's themes, more precisely **Theme 2: Modeling & understanding of collaborative or competitive interactions between humans and AI-driven entities.** We are specifically focusing on the critical aspect of AI-based systems functioning as active team members. This involves not only adapting AI explanations to diverse user needs for enhanced control and understanding but also developing and testing AI capabilities in dialogic interactions and cooperative decision-making within team settings. The project IFDOD embodies the priorities outlined in the theme. It allows for establishing more effective and understandable interactions between humans and AI-based models.

7 Project organization, duration, milestones

Milestones: Year 1: We will develop an advanced argumentation framework tailored for online debates and start the initial deployment of our AI-based system. This includes setting up the basic functionalities for argument strength assessment, feedback provision, and conflict resolution. We will initiate the integration of dialogue and exchange mechanisms with the system, ensuring it can function as a proactive member of the team rather than just a facilitator. **Year 2:** We will focus on refining the system's ability to provide adaptive explanations and interactive feedback. This phase will include enhancing the system's capability to understand and respond to the cognitive and emotional states of human participants. We will evaluate the impact of these interactions on decision quality, user satisfaction, and trust, adjusting our approaches based on user feedback and system performance.

References

[1] Lisa P Argyle, Christopher A Bail, Ethan C Busby, Joshua R Gubler, Thomas Howe, Christopher Rytting, Taylor Sorensen, and David Wingate. Leveraging ai for democratic discourse: Chat interventions can improve online political conversations at scale. *Proceedings of the National Academy of Sciences*, 120(41):e2311627120, 2023.

[2] Philippe Besnard and Anthony Hunter. *Elements of Argumentation*. MIT Press, 2008.

[3] Elena Cabrio and Serena Villata. Five years of argument mining: A data-driven analysis. In *IJCAI*, volume 18, pages 5427–5433, 2018.

[4] K. R. Chowdhary. *Natural Language Processing*, pages 603–649. Springer India, New Delhi, 2020.

[5] Phan Minh Dung. On the Acceptability of Arguments and its Fundamental Role in Non-Monotonic Reasoning, Logic Programming and n-Person Games. *Artificial Intelligence*, 77:321–357, 1995.

[6] Nils Reimers, Benjamin Schiller, Tilman Beck, Johannes Daxenberger, Christian Stab, and Iryna Gurevych. Classification and clustering of arguments with contextualized word embeddings. In *ACL*, pages 567–578, 2019.

[7] Michael J. Wooldridge. *An Introduction to MultiAgent Systems, Second Edition*. Wiley, 2009.