# Interpreting and explaining on-line debates

There are numerous platforms for online argument-based discussions, like the subreddit ChangeMyView, the idebate's debatebase (https://idebate.org/debatabase), Kialo (https://www.kialo.com/) and Debategraph (http://debategraph.org) , which was used by the White House and CNN. Some cases of e-democracy in France are the online discussion about la loi numérique (https://www.republique-numerique.fr), which was the first law in France to be preceded by an online debate, and Le grand débat national (https://granddebat.fr/). The Conference for the Future of Europe (https://futureu.europa.eu/) is an online platform that receives the contributions from the European citizens to debate about the Europe's priorities.

All those platforms have one thing in common: there is no or very little automatic reasoning used to disentangle the different viewpoints raised in the debate, i.e. the full potential of AI is not used in order to treat the data.

The work of this internship will contribute to the ANR project AGREEY. The goal of this project is to allow for automatic reasoning and to exploit the data present in those platforms by using artificial intelligence, computational argumentation theory and natural language processing. Namely, during the discussions on those platforms, many arguments are raised. Those arguments are the key part of the discussion and contain valuable information. However, due to the large number of arguments, it is time consuming to read them all. Also, arguments are either text-based or only partitioned in two groups: pros/cons but there are links between them that remain hidden.

We propose to represent the discussions in the form of a graph, which is the standard representation format in the area of computational argumentation.There are several things that can be done automatically. First, we can identify key arguments. There are several factors that can help us to do this, namely the number and the quality of attackers and supporters, whether an argument is defended (its attackers are attacked), its conclusion, how many votes did it get, the ratio of the positive and the negative votes etc. Second, we can try to predict whether the proposal is going to be accepted or rejected by looking at the strengths of arguments, their relationships, their strengths and so on. Let us emphasize that an AI system is not capable of always guessing the decision which will be taken by the humans, for several reasons, for example some of the arguments may be implicit and people can be irrational. Third, the system can help a new user joining the discussion to easier grasp its current state (e.g. key arguments, main lines of critique). It is also very useful to policy makers who want to quickly understand the public opinion, especially in the case of a high number of arguments.

## Objectives of the internship

The first goal of the internship will be to be able to first extract a data set of debates from existing platforms. The student would then focus on developing a first approach to automatically identify the most important arguments and to estimate the acceptability degrees of arguments.  Learning approaches would be investigated to build **models of the debaters** from which it would be possible to predict the outcomes of debates and decisions that will be taken, potentially using inverse RL techniques.  There are a few recent approaches which took the same perspective, see eg. [3]. One other way to approach the problem is run multiagent simulations on the basis of a simplified model to see how these strategies affect the debates, see eg. [2].

These models of the debaters could then be used to generate explications of those decision.  We would like to be able to justify why a given argument is strong, or why a given proposition is probably going to be accepted. In order to construct and return explanations to a given audience, our idea is to take into consideration the work of Miller [2019], who brings together and discusses the many existing works from the social sciences on the subject of human explicability

- Lieu : LIP6, équipe SMA
- Encadrants :  Aurélie Beynier, Nicolas Maudet
- Durée : 6 mois
- Possibilité de poursuite en thèse dans le cadre du projet

# Bibliography :


1. Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence, 77(2):321–358.
2. Louise Dupuis de Tarl., Elise Bonzon, and Nicolas Maudet. Multiagent Dynamics of Gradual Argumentation Semantics. In Piotr Faliszewski, Viviana Mascardi, Catherine Pelachaud, and Matthew E. Taylor, editors, 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), Auckland (virtual), New Zealand, May 2022. URLhttps://hal.archives-ouvertes.fr/hal-03584238. Online.
3. Winning Arguments: Interaction Dynamics and Persuasion Strategies in Good-faith Online Discussions.WWW '16: Proceedings of the 25th International Conference on World Wide WebApril 2016 Pages 613–624 https://doi.org/10.1145/2872427.2883081
4. Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence, 267:1–38, 2019.
   URL https://www.sciencedirect.com/science/article/pii/S0004370218305988?via%3Dihub.