

Online communication, especially social media are vast mines of data, derived from a multitude of diverse sources. Unfortunately, it is also an area particularly vulnerable to manipulation, fake news, extreme emotions and polarization. While most commonly used techniques for data extraction are based on statistical and network analysis, the less conventional approach of Argument Mining presents an opportunity for comprehensive qualitative examination of debates published on social platforms and portals. This approach offers an unprecedented set of tools for policy-makers and researchers. As a subfield of natural language processing (NLP) and computational linguistics, AM focuses on automatically identifying, extracting, and analyzing argumentative structures within natural language texts, which includes recognizing core components of arguments, such as claims and evidence.

But there's a problem: often these new techniques are used in a way that's too narrow. They focus mostly on technical or linguistic aspects and overlook the broader social context. They tend to only look at how persuasive a single argument is, without considering its position in the bigger debate landscape. And yet, public debates play a huge role in shaping society, engaging people in community activities, and creating social connections. So, in this project, we want to use these AI-powered techniques to analyze and understand online debates on public issues, in all their complexity, combining the methods from social sciences and computer science. The study of online civic participation in the public sphere, debates and argumentation is becoming one of the most important challenges that social sciences, particularly the study of public policy, will face in the near future. While having a vast amount of information is unquestionably of value, such resources become less useful, or even useless, if we cannot process the data efficiently and quickly enough. Reducing the information noise in contemporary social media and extracting valuable content from the online debate is a real challenge, becoming increasingly difficult due to both the explosion of content and the spread of disinformation. Argument Mining comes in handy.

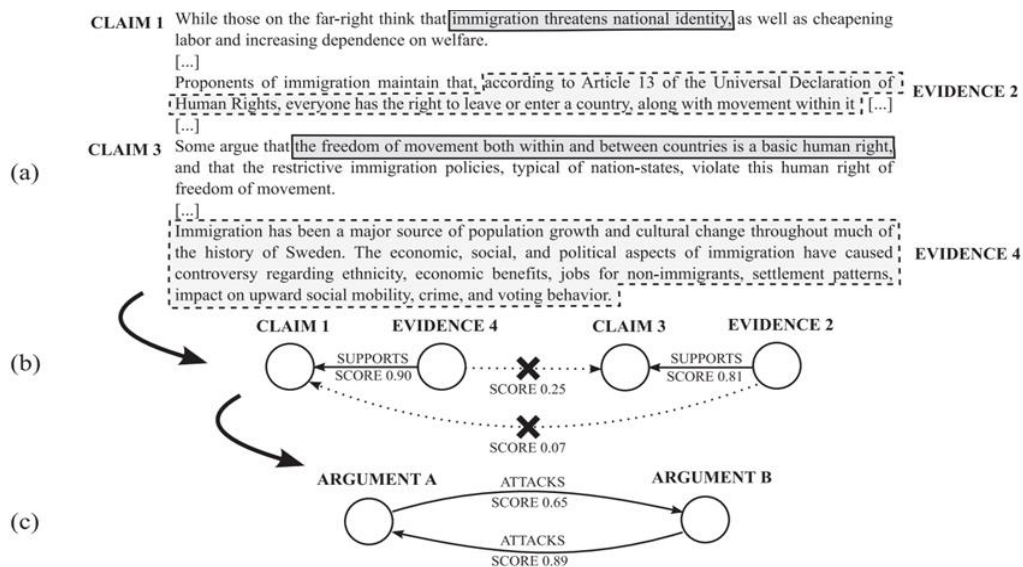


Figure 1. A process of the automatic extraction of arguments from text (source: Lippi & Torroni, 2016).

It's a very common opinion that the best way to model online public debates is to follow the deliberative democracy theories originating from J. Habermas and J. Rawls. But we believe there might be more than one way to look at this. We will also consider another model, called 'agonistic', inspired by thinkers like C. Mouffe and H. Arendt. We want to find ways to describe and measure these different styles of online debates taking into account these theoretical models and the role of arguments in them. We will also look at what happens when debates go sour - when people stop arguing sensibly and start spreading misinformation or attacking others.

From a technical point of view, we will develop a new approach to Argument Mining, which we are calling the "Hybrid AM Model". This will use a mix of advanced language models to predict arguments, alongside techniques for studying sentence meanings. The long-term goal in this project is building a new, multilingual text corpus concerning most current public debate topics, including English and Polish databases. Moreover, the corpus of texts created as a result of the project will be made available under Creative Commons license to other researchers, which will be the basis for further Argument Mining projects in Polish language.