

Improving the quality of public debate with AI

Oana Balalau

AIvolution*, Brussels, 16 November 2023

My name is Oana Balalau and I am a researcher working on several topics in AI, in particular, related to natural language processing. I am here to discuss how AI can have a strong positive impact on our society, in particular in improving the quality of public debate. Full disclosure: my talk, which is part of a public debate, is better because of AI, as I could use a search engine to find interesting examples and a spell checker to improve my writing. And given that I am sure I was not the only one that did this, I could rest my case. But I am a researcher, we like detailed and complex arguments, and possibly slightly boring ones such that we know they are correct.

Today, we are exposed to an abundance of data, especially on the Internet, which we need to filter, interpret, and assimilate to make informed decisions in our lives. In my team, we work on several tools designed to assist journalists or citizens via AI and data management. Each time I mention AI in the following presentation, I am referring to open source algorithms that use deep neural networks, which are today the most powerful tools in AI.

Suppose you are a journalist. You are about to interview a politician or a scientist about the potential health risks of pollution, in particular their link with severe illnesses. You might first wonder if the person might have a biased perspective as they are receiving funding from heavy polluting companies. How could you discover this if the information is not already present in their CV or webpage? If the person is a scientist they might have declared this funding in one of their papers, in particular in the acknowledgement section. It would be great to already have a database where all of these conflicts have been extracted from publications. This is what we do in our work, which is a collaboration with Le Monde. We extract conflict-of-interests from textual data using natural language processing tools [UBM23]. More precisely, we work on extracting triples of the form (person, type of conflict of interest, company). Given the sentence: "Michel Aubier is working as an advising doctor for Total and he is also a member of the administration council", we would like to extract (Michel Aubier, advising doctor, Total) and (Michel Aubier, administration council, Total). Part of this work is ongoing, as we aim to have a high quality database and extractions are sometimes incomplete.

Let's get back to our journalist. This time, she wants to write an article about inequality and how it is perceived in public debates. For this, she takes a look at statements that are made by public figures on this topic and tries to prove or disprove them. This is called fact-checking. Fact checking is very important for journalists, and most well-established media will have a dedicated fact-checking team. For example, in France, we have collaborated with Le vrai du faux, from Radio France. These journalists monitor all current events, and in particular the statements of politicians and well-known personalities. For example, the journalists might need to verify the claim that "The total wealth of the French billionaires (43 people) sums up to 21% of GDP of France in 2022, that is 590 billion dollars"¹. To assist fact-checkers, we have created a web application where journalists can find 100 thousands tweets of politicians, which have been labeled as being checkworthy claims or not [BEG⁺22]. In addition to detecting claims, we have a search interface for INSEE and Eurostat data, two reliable data sources for statistical data, which hold today 120 thousands tables of statistical data, and that can be used to verify claims. To note we will not decide automatically if a statement is true or false, this is the job of the journalist. To sum up, we find claims and if possible we match them with evidence.

Finally, our journalist has noticed something interesting when looking at claims related to inequality: some of the claims are backed up by incorrect arguments, also called fallacious arguments. As I mentioned, a statement for which we need to determine its truth value is known as a claim. To validate a claim we will look for evidence, that is other statements released by reliable sources, and a

*<https://eurotech-universities.eu/news-and-events/events/aivolution/>

¹https://en.wikipedia.org/wiki/List_of_French_billionaires_by_net_worth, accessed in November 2023

reasoning method that allows us to verify if the evidence supports or not the claim. A claim, together with its evidence and reasoning is called an argument. Fallacious arguments are arguments in which the reasoning is incorrect. So what about them? For example, when talking about immigrants, there is often a hasty generalization fallacy: we talk about a few examples of immigrants committing crimes, and at the end of the story there is an explicit or implicit claim that all or a majority of immigrants commit crimes. This is not how statistics work. When we talk about women we often use the appeal to tradition fallacy: women traditionally had less rights, traditionally had to take care of children, traditionally had less revenue, so then they should still be in this situation. There are many other types of fallacies: ad hominem, appeal to nature, slippery slope, and so on. Our journalist asks herself the following question: How often do these kinds of arguments are used in political news and what is their impact? We have worked on automatically detecting fallacious arguments [SBH21], in addition to studying how they are used in political discussion [BH21], where they fall under the umbrella term of propaganda. Automatically detecting fallacious arguments is part of the field called argumentation mining and it is not a solved task, but we have tools that can give us good results. We have performed a long-term analysis of propaganda on online forums discussing political news. For almost 400K political articles and their comments, we investigate via the use of an automatic tool for classifying arguments in fallacies or not: i) Who is posting propaganda? We find that extreme right or extreme left journals will more often use propaganda. ii) How is propaganda received on political forums? We find that an article that has more propaganda content, it might receive more user engagement, measured either as the number of comments or as upvotes and downvotes. Such large-scale analysis would have not been possible without the use of AI algorithms.

I have described these applications from the perspective of a journalist, but we as citizens often ask ourselves the same questions. I hope you found these applications interesting and I note that we still have ongoing work and there is still much to do. These applications come in addition to all the discussions you might have had with ChatGPT over a glass of wine and because we are in Belgium, over a waffle. AI can facilitate our work and also make it more interesting and challenging. Because of these I see two important directions that our society should take: investing more in quality education such that no citizen is left behind, and secondly, given that we are becoming much more efficient and productive as a society, perhaps we have finally deserved more free time, more rest, more time to care about our dear ones, and more time to care about the planet we live on.

References

- [BEG⁺22] Oana Balalau, Simon Ebel, Théo Galizzi, Antoine Deiana, Emilie Gautreau, and Antoine Krempf. Fact-checking multidimensional statistic claims in french. In *TTO 2022-Truth and Trust Online*, 2022.
- [BH21] Oana Balalau and Roxana Horincar. From the stage to the audience: Propaganda on reddit. In *EACL 2021-16th Conference of the European Chapter of the Association for Computational Linguistics*, 2021.
- [SBH21] Saumya Sahai, Oana Balalau, and Roxana Horincar. Breaking down the invisible wall of informal fallacies in online discussions. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 644–657, Online, August 2021. Association for Computational Linguistics.
- [UBM23] Prajna Upadhyay, Oana Balalau, and Ioana Manolescu. Open information extraction with entity focused constraints. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 1255–1266, 2023.