Data Challenge Plan

Critical Text Analysis Tool

10.11.2023

Amira Chaib

Iris van den Boomen

Overview

Think Big

What is your innovative idea?

A tool to help with critical text analysis that highlights, for example: logical fallacies, or journalistic standards not applied in the text. The tool could also make use of the Socratic method to generate questions to help the reader maintain a critical perspective.

To have a proof of concept we can pick one kind of logical fallacy or journalistic standard to get started and then expand on it.

Amira discussed the idea with Olaf, and he mentioned that language models should already be able to detect logical fallacies so could start with highlighting those and then add journalistic standards. Also, it could be an extension for the true/fake news classifier so that it gives you a classification and then expands on it by highlighting the parts that can be considered misleading.

Another idea that came up is making it more like a learning tool for how to critically view and analyse texts, to show what to pay attention to.

What problem, need, or goal does your idea address?

When reading through e.g. articles, especially when the text hails from a supposedly trustworthy source, maintaining a critical view can be difficult. Instead, we accept what we read without questioning which is especially problematic when it aids the spread of misinformation or propaganda.

The tool should help with critical analysis of texts for the average uninformed/liable audience, to build a broader awareness and limit the impact of propaganda and misinformation.

Which (hypothetical or real) data will you need?

Mainly, we would need texts and the extracted patterns that we want to identify, which we would need positive and negative examples of. Lots of data can be used to generate a dataset based on, e.g.: newspaper articles. However, it would be a huge effort to generate that by ourselves and not achievable in the time we have, so we are going to look for what we can find or which already trained models we could make use of.

Why is this interesting and innovative?

Momentarily, I am aware of tools that will give a score on the trustworthiness of an article and that will categorise it as true or fake. Often these categorisations are based on other sources, however, how is their trustworthiness guaranteed?

A reader needs to be able to identify the reasons for which a text is problematic. By highlighting and explaining the sections and displaying the reasoning behind a text being biased or untrue, that “classification” is corroborated. Moreover, the reader learns the patterns and rhetorics applied and through learning can recognise them in similar cases.

Are there ethical aspects involved?

Yes, there are a few ethical aspects involved.

Transparency: the users need to understand how the tool works, so it should be explained clearly.

Privacy: since everyone will be able to use the tool and input their text themselves, privacy concerns should be addressed.

The first step

What is the realistic first step?

The first step would be to do research and domain understanding. We need to come up with some research questions and answer those. We need to learn about logical fallacies, journalistic standards, and critical analysis ourselves. Then, we should also find data that is useful for the project and start analyzing it.

Which real data will you use?

To make sure that we know what we are doing, we will start with simple data. That could be a dataset found on Kaggle, where we perform EDA (data cleaning/understanding/preparation) and modelling. When we have a clear idea of what we want to do with that data, we can move on to e.g. articles.

Which techniques and tools will you most probably use?

For this project, deep learning will be used. We will use Explainable AI to visualize what is happening and why. In the end, we hope to have a full-stack application.

When is your first step successful?

Our first step is successful if we have gathered enough knowledge to continue implementing it in the next steps.