

Prompts

ChatGPT was used as the tool for the classification (ranking) of the YouTube comments in classes of 1-5 (with 1 being not toxic and 5 being toxic). The below prompts were given:

"I want you to rank certain comments I will provide you each time according to their toxicity 1-5, (1: not toxic, 5: toxic). provide a csv format reply with the class number for each comment."

Then, I provided a list of comments (from 20-50 according to their size), in the below format:

38,"sounak is an Aho! kai o mitsotakis epetis ean eipe oti den tha ta syzytouse ta marmara mas e! pou na..... sounak"

39,"πακιστανος δεν ειναι, αλλα φιλοτουρκος και αρα Τουρκοσπορος"

40,Great report thank you very much for the update

41,Thank you very much for the update

...

Usually, the answer repeated the comments together with the annotated class, so I replied (to make copying and pasting easier):

"Return only the comment index and its class number."

After that the result was good and I repeated the input of comments until the tool forgot the task (usually when that happened ChatGPT was trying to explain the comments rather than classifying them or it gave out of range classes), where I provided again the first prompt.

When I got the classes for all the comments, I stopped using the tool and moved to the use of the dataset.