List of prompts and their output results.

Prompts for “Categorize the different texts in one of the following category”

1. *Using only the following categories: "Question´", "Complaint", "Praise", "Suggestion" sort the following text snippets. If no category fits use "None"*
    - For about 48% of the total data, ChatGPT used the given categories and stayed on track for the task. After that, it started creating its own categories, which are mostly good options, but not part of the initial prompt, so I decided to reduce the accuracy score for it, as it does not correspond with the initial task.
    - It is also good to notice, that even if the answer could be considered good, some were completely wrong and ChatGPT also completely ignored some data and just did not give any answer for them, thus reducing the final accuracy score.
    - I had to separate the task into a number of smaller tasks, but also start new sessions and give the prompt again, in order for ChatGPT to stay in line with what was asked for, but it has a decent accuracy rate, for the Post + comments table.
        i. I think that for the Post table, the accuracy value should lay higher, as the part mostly wrong, where concerning comments, where the context was barely given to the model
        ii. This turned out to be a true observation as ChatGPT definitely has a higher success rate in all three categorisation activities when dealing only with posts, thus making the context an important part of the categorisation capacity of the model.
    - When going over the missed/wrong data again, and starting a new session with the same prompt for the part of texts that were wrong/not done, the accuracy rate can go up to 0.8, as when not doing it, it can go down, in this case to a rate rounding 0.61
2. As expected, using only main posts and removing comments from the selection of text improved the accuracy to a high 0.95 rating.
    - However, some texts had to be redone, as ChatGPT did not follow the prompt on the first tries, even if there was a particular attention to regularly updating the room in order to remind the model of the prompt and for it not to ignore it completely.
    - Also interesting to note, that it at first included suggestions when submitting a *none* answer, but stopped after a few times. I cannot explain why.
    - It also changed the way it displayed the answer in its answer, even tho the prompt and text were given in the exact same way each time.
    - Less than 5% of the data had to be repassed into ChatGPT, the method did improve the optimisation time of work required.

Prompts for “Categorize the different texts in one of the following thema”

2. *Using only the following categories: "Builds", "PvP related discussions", "PvE related discussions ", "Hardware-related" categorize the following text snippets. If no category fits use "None". You can give more than one category to each snippet*

- For about 78% of the cases, ChatGPT stayed on track with the task asked. After that, it either gave categories not proposed, or ignored in mass certain parts of text. Without going back on those, it drastically reduces the total accuracy of the exercise.

- The accuracy rate lies about 0.51 as 22% of the given classification were not an option, thus marked as incorrect.
    - This rate can get higher if we take the time to redo the texts that were not correctly categorized by starting a new session, re-giving the same prompt and re-giving the part of texts that were given a not-allowed classification or were first ignored.
- The process can be quite daunting and might sometimes require many tries as ChatGPT will redo the same thing on the smaller data-set.
- ChatGPT often gave 2 options, a right one and would use the “none” option to add a second category that was not on the options given (it would then specify what in parentheses behind the “none” classification) which also played to reduce the accuracy, as it was not asked of it to do so.

“Remember, these categorizations are subjective and may vary based on context.”
“Please note that the categorizations are based on the content and may involve subjective interpretation.”
“Please note that some snippets could fit into multiple categories based on their content.”

Prompts for “Categorize the different texts according to the general feeling”

3. *Can you say if the following text snippets have a "positive", "negative", or "neutral" position towards the game?*

- For about the first 85% of the data set, ChatGPT managed the task according to the original prompt in a fairly accurate manner, before going completely off-prompt and tanking the accuracy rate to a 0.6. When going over the last 15% in a new chat and re-writing the prompt, the accuracy of the categorizing goes up to 0.8. ChatGPT also sometimes gave two feelings, and justified its choice behind it. It was sometimes a good explanation, sometimes the missing context of certain comments (as discussed previously) is probably responsible for the mistakes made.