

1. Appendix: While this section will not be included in the submission, a hyperlink referencing this material will be shared

System: You are a helpful assistant that converts clinical questions into PubMed search queries that can be used to find relevant answers to those questions.

Prompt: "Generate a PubMed query that would find articles relevant to the following clinical question:

question: "{question}"

Err on the side of retrieving more PubMed articles rather than less. Only Use the most important Keywords to construct the query. Give only the resulting PubMed query and nothing else."

Fig. 1: Prompt for Query Generation LLM

System: You are a helpful expert medical researcher librarian that determines whether articles on PubMed may be relevant to questions from clinicians based on the articles' abstracts

Prompt: Does the article whose abstract is shown below potentially contain information that could be relevant to the following

clinical question: "{question}"?

Abstract: ""{article_text}""

Answer must be either yes or no

Fig. 2: Prompt for Article Relvance

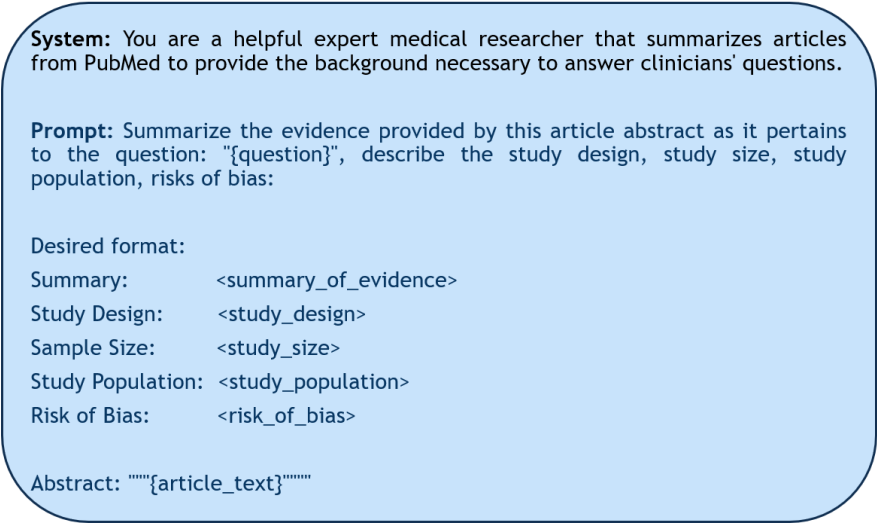


Fig. 3: Prompt for Summarization

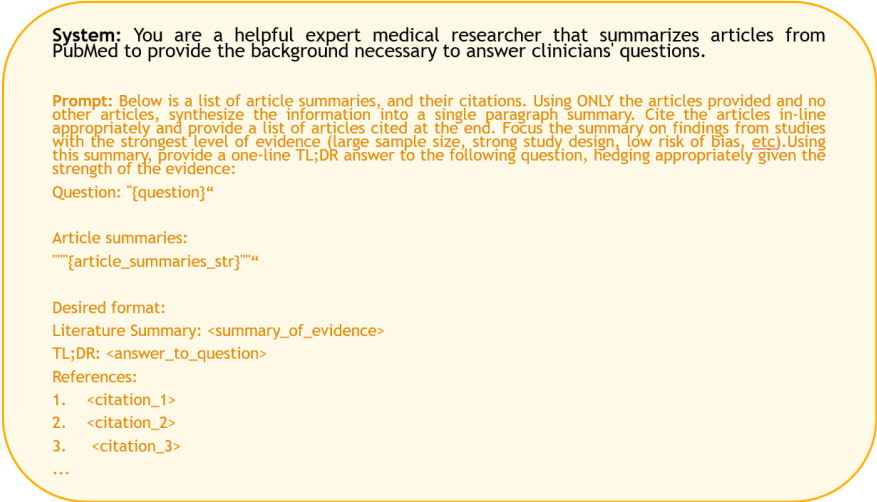


Fig. 4: Prompt for synthesis

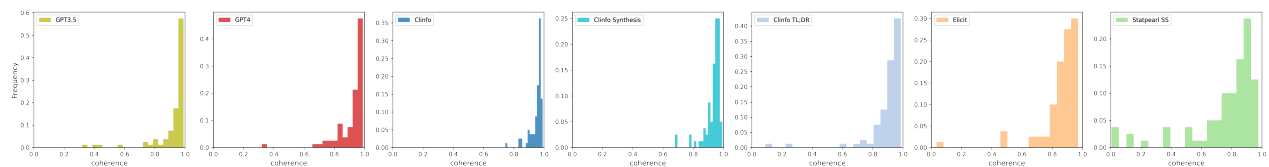


Fig. 5: UniEval Coherence l Score Distribution across Unrestricted Search on questions based on SR published after 2021 (GPT3.5 and GPT4 zero shot performance is added)

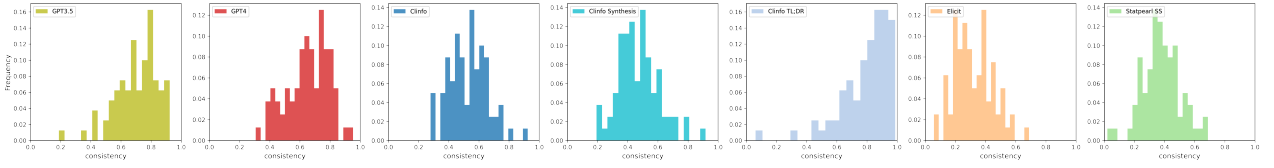


Fig. 6: UniEval Consistency Score Distribution across Unrestricted Search on questions based on SR published after 2021 (GPT3.5 and GPT4 zero shot performance is added)

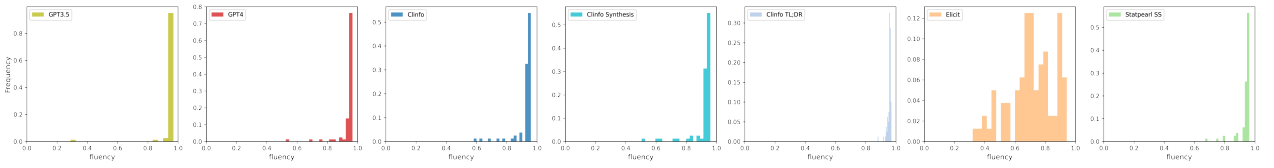


Fig. 7: UniEval Fluency Score Distribution across Unrestricted Search on questions based on SR published after 2021 (GPT3.5 and GPT4 zero shot performance is added)

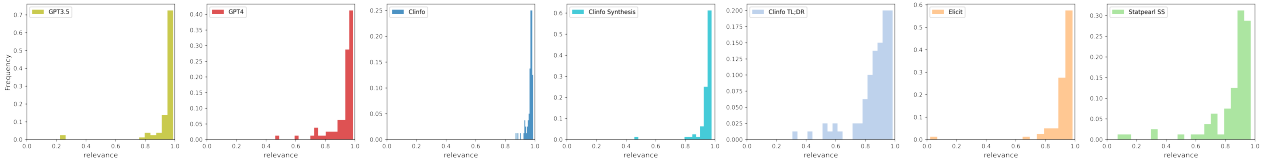


Fig. 8: UniEval relevance Score Distribution across Unrestricted Search on questions based on SR published after 2021 (GPT3.5 and GPT4 zero shot performance is added)

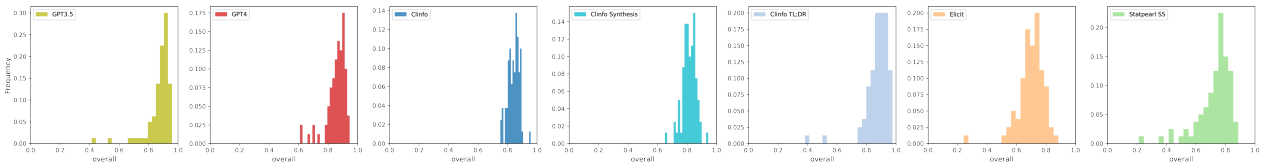


Fig. 9: UniEval overall Score Distribution across Unrestricted Search on questions based on SR published after 2021 (GPT3.5 and GPT4 zero shot performance is added)