Analysis

Will G-EVAL preferLLM-basedoutputs? One concern about using LLM as an evaluator is that it mayprefer the outputs generated by the LLM itself, rather than the high-quality human-written texts.

Results on Hallucinations

Advanced NLG models often produce text that does not match the context input (Cao et al., 2018), and recent studies find even powerful LLMs also suffer from the problem of hallucination.