Knowledge Discovery in the Age of LLMs

Jakub Zavrel[1]

[1]*Zeta Alpha, Amsterdam, The Netherlands*

Abstract

Large Language Models (LLMs) like GPT-3 are quickly changing the way NLP is done, and hence also how NLP is done for the purpose of knowledge discovery in the academic literature. Tasks that have traditionally been done by specialized NLP models, like entity extraction, summarization, and question answering, can now all be prototyped, usually with high accuracy, using zero-shot or few-shot prompting of LLMs. For example, this allows us to extract highly accurate meta-data, such as up-to-date author and affiliation, for scientific impact analysis from recent scientific papers[1]. The combination of LLMs and Information Retrieval systems has recently evolved into the paradigm of Retrieval Augmented Generation, which is one of the most promising approaches to use LLMs for Question Answering and to reduce hallucination, especially for data sets that were not accessible to LLMs during training, such as private data sources or very recent documents. Retrieval Augmented Generation can even be expanded to large document collections, like full conference proceedings, to quickly create conference summaries, blog-posts or tabular digests. At Zeta Alpha we are building a modern platform for scientific and enterprise knowledge discovery with neural search and generative language models at the core. In this talk, we show how we integrate LLMs with neural search to answer questions, explain documents while reading, summarize large document collections, and generate meta-data on the fly.

---

[1]See: https://www.zeta-alpha.com/post/must-read-the-100-most-cited-ai-papers-in-2022