Avoiding Perjury of Large Language Models using Retrieval Augmentation

Tomás Vergara Browne
Pontificia Universidad Católica de Chile
tomvergara@uc.cl

Abstract

Retrieval Augmentation provides the abilities to Large Language Models to access external information for them to generate more factual answers. Although the use of Retrieval Augmentation techniques has been intensely studied in the last year, a particular domain of legal domains has not been significantly studied, in comparison to other domains. In this paper, we explore the use of Retrieval Augmentation for the Mistral 7B model, which has surprised the open source community for its ability to supersede in performance models which are considerably larger. Through the use of the open source framework of RETA-LLM, we are able to add retrieval capabilities to the model. With this model, we are able to provide a new state of the art for the benchmark of LexGLUE.