Some New Perspectives on Data Management and Knowledge Exploration in Legal Informatics in the LLM Era

Nguyen Ha Thanh

Researcher @ Center for Juris-Informatics

A. Introduction

Big data is revolutionizing various fields, and powerful AI language models like GPTs play a significant role. However, overcoming their limitations requires innovative data management and knowledge exploration to boost their potential in legal informatics.

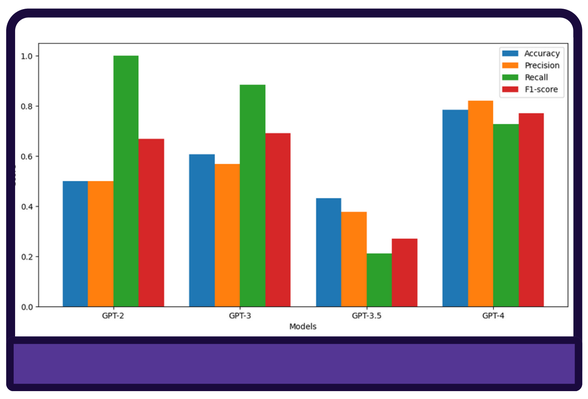
B. Objective

The study series aims to enhance data management and knowledge exploration in legal informatics to improve the reliability and usability of LLMs.

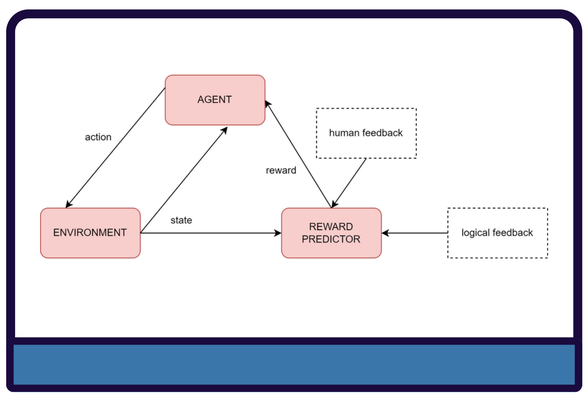
C. Methodology

Assessing strengths and weaknesses of LLMs in legal informatics tasks, identifying root causes of weaknesses, and proposing data management and knowledge exploration strategies for improvement.

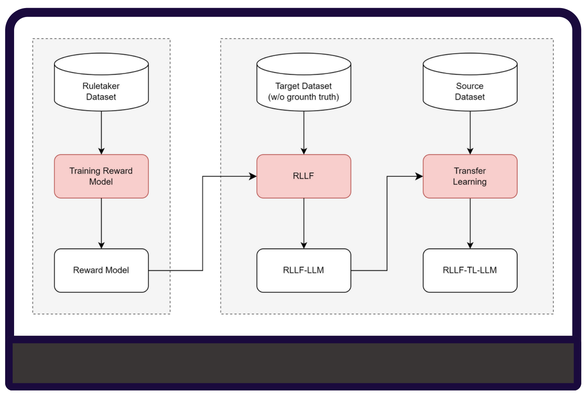
D. Current Progress



Our experiments indicate that LLM currently lacks proficiency in logical reasoning, while traditional sample-based finetuning or RLHF do not offer comprehensive solutions.



Finetuning with the traditional data signal and format is not sufficient for creating a reliable AI system for juris-informatics. It is crucial also to consider logic, fairness, justice, and legal expertise in the development process.



Our recent experiments indicate that employing reinforcement learning with logical feedback can enhance knowledge exploration of models and improve their performance.

E. Discussion

When dealing with models that have a restricted number of parameters, this method elevates the model's workload by requiring the optimization of additional loss functions. Even though there are challenges involved, in the legal field, the model's interpretability holds more significance than the accuracy of the output for LLM or any AI system to have significance.

F. Conclusion

The research highlights limitations in the traditional data and knowledge organization in building LLM. It recommends incorporating appropriate data types in the right formats to enhance their reliability and applicability in legal informatics.