007

008

000

015

017

039

041

053

Extended Abstract - Experimenting with Legal AI Solutions: The Case of Question-Answering for Access to Justice

Anonymous Authors¹

The integration of Artificial Intelligence (AI) into legal contexts has demonstrated significant potential in enhancing access to justice and streamlining legal processes. However, the application of general-purpose AI models such as GPT-3.5 and GPT-4 in high-stakes legal tasks presents substantial challenges, especially when it comes to question-answering tasks. Little prior work focuses on the data sourcing, inference, and evaluation of these models in the context of access to justice (Hagan 2023). This project has two complementary objectives: First, inspired by the high-quality data that powers general domains, we identify a gap in high-quality structured legal data (e.g., question-answer pairs) approved by legal experts. In this work, we hope to build humancentric legal AI systems by improving the data source. Second, based on performance findings, we advocate creating a domain-specific, crowdsourced language model interface that allows for diverse feedback collection.

1. Human-Centric Language Models for Legal OA

We propose an end-to-end human-centric legal AI framework, which includes data sourcing, training/inference, and evaluation; importantly, each step of the process is backed by high-quality data from legal experts (see Figure 1). To our knowledge, this type of human-centric legal framework is the first of its kind.

First, we construct a high-quality evaluation dataset of 300 questions asked by laypeople on real legal questions and answers vetted by legal experts. We ask law students to write expert answers to these questions and release this dataset to the public. Then, we develop an automatic evaluation protocol based on the factuality of the generated answer, as a legal expert would. Inspired by massive improvements to model quality through higher quality data at training time (e.g., Phi-3; Abdin et al., 2024), we improve the data sourcing process at retrieval time ??. Specifically, we propose

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

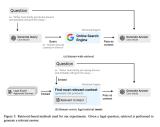


Figure 1. Retrieval-based methods used for our experiments. Given a legal question, retrieval is performed to generate a relevant answer

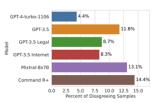


Figure 2. Factual disagreement for each model. "GPT3.5 Legal" is retrieval using only legal documents, and "GPT-3.5 Internet" is retrieval from the entire internet.

domain-specific retrieval, bolstering the performance of existing LLMs on legal question-answering by retrieving from sources trusted by legal experts. Here are the contributions:

- We construct a dataset containing real legal questions and high-quality answers labelled by legal experts. We release the evaluation 075 dataset publically.
- Evaluations: We create an evaluation protocol vetted by legal experts and find that existing models have room for improvement in factuality.
- Close vs Open RAG: We show that closed retrievalaugmented generations (RAG) from under a thousand legalexpert-approved articles matches or exceeds the performance of web-based RAG drawn from hundreds of millions of internet articles ??.

Early findings: retrieval-based approaches generally perform better than their non-retrieval counterparts except gpt-4-turbo-1106 which outperforms even retrieval models.

¹Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

2. Crowdsourcing Language Models – The OpenJustice Project

Future research will focus on closing the gap between opensourced and closed-sourced models. Drawing from performance insights from this project, we explore legal AI solutions that may show significant improvement over generalpurpose AI. Specifically, we highlight the potential of an open-source approach through www.OpenJustice.ai—a platform that encourages collaborative and crowdsourced efforts to design and test custom AI solutions for legal professionals and aid centers. This approach promotes a transparent and inclusive method of AI development, allowing diverse perspectives and expertise to contribute to more ethical and robust AI systems. An essential aspect of this domain-specific solution is the emphasis on data curation. The quality of data used to train these models is crucial for their reliability and effectiveness. Proper data curation entails not only gathering extensive and diverse datasets but also ensuring the data is representative, unbiased, and pertinent to the specific legal applications.

Along these lines, we also explore another promising avenue: the development and legal application of small language models, such as Phi-3. These models, despite their smaller size, have demonstrated capabilities comparable to much larger models. Future work will investigate the efficiency and effectiveness of small language models in legal contexts. Techniques like Direct Preference Optimization (DPO), world models, Flash Attention 2, rejection sampling, and reward modeling will be explored to enhance the quality and accuracy of AI-generated legal content. The goal is to integrate these models into legal AI systems, ensuring they align with the nuanced and ethically bound nature of legal reasoning.

- [1] M. Hagan, "Towards Human-Centered Standards for Legal Help AI," Margaret D. Hagan," Towards Human-Centered Standards for Legal Help AI." Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. Publication Forthcoming, 2023, Accessed: Nov. 26, 2023. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4582745
- [2] R. Bhambhoria, S. Dahan, J. Li, and X. Zhu, "Evaluating AI for Law: Bridging the Gap with Open-Source Solutions." arXiv, Apr. 18, 2024. doi: 10.48550/arXiv.2404.12349.
- [3] M. Abdin *et al.*, "Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone." arXiv, May 23, 2024. Accessed: Jun. 10, 2024. [Online]. Available: http://arxiv.org/abs/2404.14219