Abstract

The increasing complexity of legal problems and the vast amount of legal information necessitate the development of advanced Artificial Intelligence (AI) systems that can assist in the legal domain with robustness, explainability, and truthfulness. This dissertation, titled *Towards Robust, Explainable, and Truthful Legal Systems*, addresses two pivotal tasks in the legal field using novel AI approaches: Legal Information Retrieval and Legal Textual Entailment. Addressing these tasks is instrumental for enabling advanced applications such as automated legal question answering, legal decision support systems, and predictive analytics in the legal sector.

Legal Information Retrieval (LIR) forms a fundamental component of legal AI systems by ensuring the retrieval of comprehensive and relevant legal articles, which serve as critical inputs for subsequent tasks like Legal Textual Entailment. Our approach to LIR focuses on developing a robust retrieval model that can achieve high coverage of pertinent articles while maintaining a high level of precision. Especially, our Retrieve-Revise-Refine framework achieved the macro F2 scores of 0.8517 and 0.8069 on the COLIEE 2022 and 2023 datasets, respectively, representing improvements of 3.17% and 4.24% over previous state-of-the-art methods. The experimental results affirm that our LIR system significantly outperforms existing benchmarks by retrieving broader yet more precise sets of legal documents. This high coverage is crucial as it ensures that downstream applications, such as legal textual entailment models, are grounded on a reliable and extensive corpus of legal information, ultimately enhancing the overall performance and reliability of the AI system.

In the context of Legal Textual Entailment (LTE), the challenge extends beyond robustness to incorporate the dimensions of explainability and truthfulness. Our LTE models are designed to provide robust predictions regarding the entailment relationships between legal texts. Additionally, we place a strong emphasis on building systems that can offer natural language explanations for their decisions, thereby enhancing transparency and user trust. This ability to explain decisions is vital in the legal domain, where transparency and the rationale behind decisions are of paramount importance. Our proposed method achieved an accuracy of 76.15%, representing

a significant improvement of 8.26% over the previously established state-of-the-art benchmark. Furthermore, addressing the issue of truthfulness, we propose innovative methods aimed at reducing hallucinations and ensuring that the system's outputs remain true to the input data. The truthfulness of an AI system is particularly critical in legal applications, where inaccurate or misleading information can have severe consequences. Our proposed methods demonstrate a substantial improvement in reducing untruthful outputs, thereby enhancing the reliability of the system. Our Self-itemize method exhibits a significant enhancement in accuracy, as evidenced by a 5.50% increase. Furthermore, the truthfulness of logical reasoning has substantially improved, as indicated by an 8.30% rise in the accuracy of reasoning steps.

The rigorous experimental evaluations conducted as part of this research underscore the efficacy of our proposed approaches. The findings reveal that our LIR system sets new standards in terms of coverage and precision, positioning it as a highly effective tool for legal information retrieval. Similarly, our LTE models exhibit strong performance metrics, coupled with the ability to provide clear and accurate explanations for their predictions. The improvements in the truthfulness of the system are evident, further validating the effectiveness of our methods.

In conclusion, this dissertation makes significant contributions towards the development of robust, explainable, and truthful legal AI systems. By addressing the critical tasks of legal information retrieval and legal textual entailment with novel AI approaches, we pave the way for more advanced and reliable applications in the legal field. These advancements have the potential to transform legal processes, making them more efficient, transparent, and trustworthy, thereby contributing to the broader goal of harnessing AI for societal benefit.

Keywords: Legal Information Retrieval, Legal Textual Entailment, Language Models, Explainable Legal Systems, Truthful Legal Systems