Legal Knowledge and Information Systems
G. Sileno et al. (Eds.)

doi:10.3233/FAIA230995

OpenJustice.ai: A Global Open-Source Legal Language Model

Samuel DAHAN [a,1], Rohan BHAMBHORIA [b], David LIANG [c] and Xiaodan ZHU [b]
[a] *Queen's Law, Cornell Law School, Conflict Analytics Lab*
[b] *Department of Electrical and Computer Engineering, Ingenuity Labs Research Institute, Conflict Analytics Lab, Queen's University*
[c] *Queen's Law, Smith School of Business, Conflict Analytics Lab, Queen's University*
ORCiD ID: Samuel Dahan https://orcid.org/0000-0002-1079-8998, Rohan Bhambhoria https://orcid.org/0000-0002-2597-670X, Xiaodan Zhu https://orcid.org/0000-0003-3856-3696

Abstract. Generalized AI like ChatGPT cannot and should not be used for legal tasks. It presents significant risks for both the legal professions as well as litigants. However, domain-specific AI should not be ruled out. It has the potential for legal research as well as access to justice. In this paper, we call for the development of an open-source and distributed legal AI accessible to the entire legal community. We believe it has the potential to address some of the limitations related to the use of general AI for legal problems and resolving disputes – shortcomings that include legal misinformation or hallucinations, lack of transparency and precision, and inability to offer diverse and multiple narratives.

Keywords. Legal AI, Open source, Decentralized and Distributed Learning, Feedback, Legal profession.

1. Introduction

Recent evidence shows that AI is becoming less intelligent, and the reasons are unknown. Findings suggest that ChatGPT is "drifting"[1] – also known as wild fluctuations in the technology's ability to perform certain tasks. Over just a couple of months, the machine went from answering a simple math question 98% of the time to just 2%. What does this mean for the use of AI in law? Not that much, considering that general AI systems have never performed well in law. In fact, there have been several high-profile instances of the misuse of generative AI in courts. For example, a recent Forbes headlined "Lawyer Used ChatGPT In Court—And Cited Fake Cases." However, while the issues of hallucination and citation are important, especially in the legal context, this paper will not be looking at AI flaws in depth. In fact, these have already been well documented. Instead, this research is a non-technical doctrinal effort aimed to explore potential solutions for implementing dependable legal AI solutions that are accessible to the legal community as a whole. This project is part of a greater endeavour to develop an open-source legal AI system, OpenJustice.ai.

[1] Corresponding Author: Samuel Dahan, samuel.dahan@queensu.ca

2. What is OpenJustice

2.1. OpenJustice Rollout

In March 2023, the Conflict Analytics Lab, a legal AI consortium established in 2018, launched OpenJustice, one of the first global open-source language models fine-tuned for law and negotiation. The underlying hypothesis is that fine-tuning with curated legal data will help to address some of the limitations related to the use of generalized AI for legal problems and resolving disputes – shortcomings that include legal misinformation or hallucinations, lack of transparency and precision, and inability to offer diverse and multiple narratives. OpenJustice operates as a natural-language processing interface aiming to provide reliable, in-depth answers to legal questions with legal sources. The first iteration leverages legislation and case law, as well as thousands of annotated question-answer pairs compiled since 2019 when the project originated. OpenJustice relies on three foundational features: it is (1) a distributed (2) open-source model, (3) trained on proprietary data and crowdsourced human feedback (see more detail section 3 and 1). OpenJustice is driven by a large consortium of universities, legal clinics, and industry partners. To this date, consortium members include: Harvard Law School Access to Justice Lab; McGill Law; Queen's Law; Harvard Negotiation Task Force; Pro Bono Students Canada (PBSC); Ottawa Pro Bono Employment Law Clinic; Leiden International Administrative Law Clinic; Paris Dauphine University Legal Clinic; and UCLA Law School. The OpenJustice consortium will be expanding to include its partnerships with a select number of organisations to develop in-house customized computational models trained on internal data in the following sectors: banking, law, insurance, and human resources.

2.2. Key Features

1. Retrieval Augmented Generation (RAG). Inspired by WebGPT, OpenJustice uses Retrieval Augmented Generation (RAG) to merge information retrieval and content synthesis. When a user inputs a legal query, the system scans a vast database of legal texts, ranking them by relevance and jurisdiction. This feature ensures the factual accuracy of the generated responses, eliminating the need for manual verification. We note though that the question of citation is a key issue in computer science. It is unclear as to why LLMs are unable to provide accurate citations. We note though that the question of citation is still an unresolved question in computer science. It is unclear as to why LLMs are unable to provide accurate citations.

2. Multiplicity. Traditional Large Language Models (LLMs) are trained to offer a singular, most likely solution, which does not align well with the multifaceted nature of legal reasoning. Legal reasoning cannot be reduced to a single "correct" answer; thus, AI systems like OpenJustice are designed to reflect this complexity by offering multiple perspectives and solutions to legal problems. Unlike traditional LLMs, OpenJustice aims to generate a range of legal narratives and solutions, akin to "Legal Solution Bases"—databases that contain multiple legal solutions to a given problem. While LLMs are statistical in nature and cannot perform legal reasoning—at least not yet—they can be trained to understand that legal reasoning is not monolithic but a complex interplay of statutes, precedents, and ethical considerations.

3. Probing for legal education and self-represented litigants. Crafting effective prompts can be a challenging task. As a result, an additional model needs to be trained to assist non-lawyers with legal AI prompting. This is referred to as a design probe—a prototype LLM-based chatbot design tool that supports the development and systematic evaluation of prompting strategies. This feature is particularly important in the context of access to justice and legal education. For non-lawyers, navigating the complex world of legal jargon and procedures can be daunting. Assisted prompting comes into play here, serving as a "design probe" to guide users in framing their questions or concerns more effectively. As for legal education, this serves as a pedagogical tool that encourages students to think critically, guiding them to discover answers independently rather than being spoon-fed solutions.

4. Assisted Negotiation. In many areas of law, especially in employment, consumer protection, and personal injury, the majority of disputes are resolved through negotiation rather than litigation. Traditional Legal AI systems primarily generate predictions or recommendations based on past case law. However, this approach falls short in capturing the nuances of negotiation strategies, which often play a crucial role in dispute resolution. For example, in an employment termination case, while the law may dictate that an employer must pay a certain amount, it doesn't consider the potential impact of an apology or a recommendation letter on the negotiation outcome. The "assisted negotiation feature" aims to address this gap by incorporating data from both legal precedents and past negotiations.

3. How Does it Work: Open-source Distributed Legal AI

OpenJustice relies on a combination of (i) unstructured data – this includes case law, journals, and other legal resources. (ii) structured data, which includes annotated data. There are several layers of fine-tuning that can be performed with language models. See Figure 1.

1. Raw Data Training. Models trained on unstructured legal are tuned on a "masked language modeling task" in which the model is essentially trained on a fill-in-the-blank task. As the "blanks" are already considered as present in the unstructured dataset by simply omitting a part of the data, this form of training on unstructured data is also known as "self-supervised training".

2. Instruction Fine-tuning. Instruction response fine-tuning is a process that involves feeding the model structured data in the form of question-response pairs. The model is trained using these annotated examples. The model learns to recognize patterns and make predictions based on the given instructions and desired responses. Figure 1 shows how fine-tuning works in the legal context [2].

3. Crowd-sourcing Human-feedback. This involves creating an interface that allows the user to test the model and provide feedback. For example, if the system provides incorrect information in response or citation to a query, a human expert can correct or validate the results. In the legal context, we strongly recommend a crowd-sourced approach, that is: a non-proprietary version of the model should be openly accessible to the entire legal community; that is, both law schools and legal professionals (Figure 1). In fact, we

[2]This is drawn from the OpenJustice project (originally called Smart Legal Clinic).

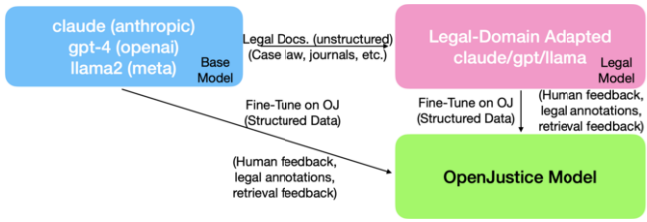

Figure 1. Cowdsourced Human Feedback

think it is key to invest in truly open LLMs for law as one of the most immediate issues for the research and legal community is the lack of transparency in these systems.

4. Decentralized Fine-tuning: Combining Open and Closed Systems. We suggest here a novel approach to reinforcement learning with a combination of open-source and closed datasets. This would create customized intelligence capabilities. As discussed earlier, the open-source dataset would rely on the legal community at large including law schools, legal clinics, industry partners, and research users who can contribute to the open model. Inputs are decentralized only by legal professionals to distill legal principles into the language models rather than misinformation from the general public. As for the closed dataset, it would be drawn from industry partners' proprietary data and feedback. While proprietary data cannot be disclosed, the two systems will learn from each other and improve the underlying generalized legal model.

References

[1] L. Chen, M. Zaharia, and J. Zou, "How is ChatGPT's behavior changing over time?," arXiv preprint arXiv:2307.09009, 2023.

[2] R. Bhambhoria, S. Dahan, and X. Zhu, "Investigating the State-of-the-Art Performance and Explainability of Legal Judgment Prediction.," in Canadian Conference on AI, 2021.

[3] C. F. Luo, R. Bhambhoria, S. Dahan, and X. Zhu, "Evaluating Explanation Correctness in Legal Decision Making," in Proceedings of the Canadian Conference on Artificial Intelligence (5 2022). https://doi.org/10.21428/594757db. 8718dc8b, 2022.

[4] R. Bhambhoria, H. Liu, S. Dahan, and X. Zhu, "Interpretable low-resource legal decision making," in Proceedings of the AAAI Conference on Artificial Intelligence, 2022, pp. 11819–11827.

[5] C. F. Luo, R. Bhambhoria, S. Dahan, and X. Zhu, "Prototype-Based Interpretability for Legal Citation Prediction." arXiv, May 25, 2023. doi: 10.48550/arXiv.2305.16490.

[6] Samuel Dahan, Maxime Cohen, and Colin Rule, "Conflict Analytics: When Data Science Meets Dispute Resolution," Management Business Review, forthcoming 2020.

[7] M. C. Cohen, S. Dahan, W. Khern-Am-Nuai, H. Shimao, and J. Touboul, "The use of AI in legal systems: determining independent contractor vs. employee status," Artificial intelligence and law, pp. 1–30, 2023.

[8] Samuel Dahan, Jonathan Touboul, Jason Lam, and Dan Sfedj, "Predicting Employment Notice Period with Machine Learning: Promises and Limitations," McGill Law Journal, 2020.

[9] J. Lam, Y. Chen, F. Zulkernine, and S. Dahan, "Detection of Similar Legal Cases on Personal Injury," in 2021 International Conference on Data Mining Workshops (ICDMW), IEEE, 2021, pp. 639–646.

[10] Samuel Dahan, Yifei Yin, and Farhana Zulkernine, "Determining Worker Type from Legal Text Data using Machine Learning," Pervasive Intelligence and Computing (IEEE PICom), 2020.

[11] J. T. Lam, D. Liang, S. Dahan, and F. H. Zulkernine, "The Gap between Deep Learning and Law: Predicting Employment Notice.," in NLLP@ KDD, 2020, pp. 52–56.