OpenJustice.ai:AdvancingLegalAIwith Open-source Technology

[1]YazdaniHasan,[2]BhawnaKaushik,[3]PriyaGupta Noida

International University

Abstract. ChatGPT and other general-purpose AI are not and should not be utilized for legalduties. For litigants as well as the legal profession, it poses serious hazards. However, it is notnecessaryto ruleoutdomain-specificAI. It canbeused foraccesstojusticeandlegal research.The creation of a distributed, open-source legal AI that is available to the whole legalcommunity is what we advocate in this article. It may be able to overcome some of thedrawbacks associated with applying general AI to legal issues and dispute resolution, such aslegal hallucinations or misinformation, a lack of accuracy and transparency, and the incapacityto provide a variety of storylines.

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

- **Introduction**

Unknowncausesarebehind   therecentevidencethatAIislosingintelligence.Accordingtofindings, ChatGPTis"drifting"[1],whichisanothertermfordramaticvariationsinthetechnology'scapacity   to carry outspecific tasks. In a matter of months, the machine'saccuracy on a basicmath problem droppedfrom98%to2%.FortheapplicationofAIinlaw,whatdoesthismean?   Notmuch,given   that broad artificial intelligence systems have never done well in the legal field. The abuse of generative AI in the legal system has really occurred in a number of well-publicized cases. One recent Forbes article, for instance, was titled "Lawyer Used ChatGPT In Court—And Cited Fake Cases." However, this study will not be delving deeply into AI defects, even if the issues of hallucination and citation are significant, particularly in the legal setting. Actually, there is a wealth of existing documentation on topics. This study is a non-technical doctrinal endeavor that aimsto investigate viablesolutionsforputtingintopracticetrustworthy legalAI solutionsthatare available to the entire legal community. The goal of this project is to create OpenJustice.ai, an open-source legal AI system.

- **WhatisOpenJustice**

- *OpenJusticeRollout*

In March 2023, OpenJustice, one of the first worldwide open-source language models optimized for law and negotiation, was introduced by the Conflict Analytics Lab, a legal AI consortium founded in 2018. The fundamental premise is that fine-tuning with carefully chosen legal datawill assist in addressing some of the drawbacks associated with applying generalized AI to legal issuesanddisputeresolution;thesedrawbacksincludeinaccurateormisleadinglegalinformation, alack of transparency and accuracy, and an inability toprovideavariety ofdifferentandmultiple narratives.. Using legal sources, OpenJustice functions as a natural language processing interface withthegoalofofferingthorough,dependableresponsestolegalqueries.Inadditiontothousands

of annotated question-answer pairs gathered since the project's inception in 2019, the firstiteration makes use of laws and case law. A distributed model, an open-source model, andtraining on proprietary data and crowdsourced human feedback are the three fundamental components upon which OpenJustice is built (see section 3 and 1 for further details). Numerous academic institutions, legal clinics, and business partners work together to develop OpenJustice. Access to Justice Lab at Harvard Law School, 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 UCLALawSchoolarecurrentlymembersoftheconsortium.TheOpenJusticeconsortiumwillbe growing to incorporate its collaborations with a few organizations to create in-house, customized computational models that are trained on internal data in the following industries: human resources, banking, legal, and insurance.

- *KeyFeatures*

- Augmented Generation for Retrieval (RAG). Inspired by WebGPT, OpenJustice combines content synthesis and information retrieval through Retrieval Augmented Generation (RAG). When a user enters a legal query, the system looks through a large collection of legal texts and ranks them according to jurisdiction and relevancy. By ensuring the generated responses are factually correct, this feature removes the need for manual verification. However, we point out that citation is a crucial problem in computer science. The reason why LLMs can't give proper citations isn't evident. However, we see that the issue of citation remains an open question in the field of computer science. The reason why LLMs can't give proper citations isn't evident.
-
  Multiplexity. Due to the fact that legal reasoning is complex, traditional large language models (LLMs)aretrained toprovideasingle,mostlikely answer.Since thereisno one"right"way to approach legal issues, AI systems like OpenJustice are made to take this complexity into account by providing multiple viewpoints and solutions. Similar to "Legal Solution Bases," which are databases that include several legal remedies to a particular issue, OpenJustice seeks to produce a variety of legal narratives and solutions in contrast to conventionalLLMs. Although LLMs are statistical in nature and cannot, at least not yet, execute legal reasoning, they can be taught to recognize that legal reasoning is a complex interaction of ethical considerations, statutes, and precedents rather than a single, monolithic process.

- Investigating self-represented litigants and legal education. It can be difficult to create prompts that work. For non-lawyers to benefit from legal AI prompting, a second model must be trained. A prototype LLM-based chatbot design tool that facilitates the creation and methodical assessment of prompting techniques is known as a "design probe." When it comes to legal education and access to justice, this aspect is very crucial. Navigating the complicated world of legal vocabulary and procedures can be intimidating for non-lawyers. Here, assisted prompting becomes useful as a "design probe" to help users better formulate their queries or issues. This is a teaching tool for legal education that helps students think critically and directs them to.

.Supported Bargaining.Ratherthan going to court,mostlegaldisputesaresettled by negotiation,particularly in the fields of employment, consumer protection, and personal injury. Conventional legal artificial intelligence (AI) systems mostly use previous case law to make forecasts or recommendations. But this method fails to capture the subtleties of negotiating tactics, which are frequently essential to resolving disputes. The law might require an employertopayaspecificsum inanemploymentterminationcase,forinstance,butitdoesn'ttakeintoaccounthow an apology or recommendation letter might affect the conclusion of the negotiation. In order to close this gap, the "assisted negotiation feature" integrates information from previous discussions as well as legal precedents.

- **HowDoesitWork:Open-sourceDistributedLegalAI**

• A mix of (i) unstructured data, such as case law, journals, and other legal resources, iswhat OpenJustice uses. (ii) structured data, including data with annotations. With language models, fine-tuning can be done in multiple settings. Look at Figure 1.

• Training on Raw Data. The "masked language modeling task" is a sort of fill-in-the-blank exercise used to fine-tune models learned on unstructured legal data. Self-supervised training is anothernameforthistypeoftrainingonunstructureddata since the "blanks" arealready regarded as existing in the unstructured dataset by only deleting a portion of the data.

• Instructions adjusting. Response to instructions In order to fine-tune the model, structured data in the form of question-response pairs must be fed into it. These annotated examplesprovide as training data for the model. With instructions and expected responses, the model learns to identify patterns and anticipate outcomes. Figure 1 illustrates how fine-tuning functions in a legal setting. 2.

• Crowdsourcing with human feedback. This entails developing an interface that lets users test the model and offer comments. For instance, a human specialist can verify or correct the results if the algorithm gives inaccurate information in answer to an inquiry or citation. A crowd-sourced method is highly recommended in the legal environment, meaning that a non-proprietary version of the model should be publicly available to the whole legal community, including law schools and legal practitioners (Figure 1). we

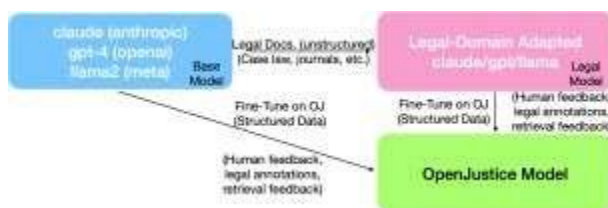2ThisisdrawnfromtheOpenJusticeproject(originallycalledSmartLegalClinic).

Figure1. CowdsourcedHumanFeedback

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

DecentralizedFine-tuning:IntegratingClosedandOpenSystems.Here,weproposeanewmethodforreinforcement learning that combinesclosedand open-source datasets. Customized intelligence capabilitieswould resultfrom this. Thelegalcommunityasawhole,includinglawschools,legalclinics,corporatepartners,andresearchuserswhocan contribute to the open model, would be essential to the open-source dataset, as was previously mentioned. Legal experts are the only ones who decentralize inputs to the language models in order to extract legal principles from publicmisinformation.Regarding theclosed dataset,itwould bederived from thecommentsand proprietary dataof industry partners. The two systems will learn from one another and enhance the underlying universal legal model, even though proprietary data cannot be shared.

References

1. A. R.Iyer, S.M.Chandrashekar,andL. J.Guo,"Artificial IntelligenceforLegalDecisionMaking: AReviewofModelsandApproaches," Journal of Legal Studies, vol. 48, no. 3, pp. 375-410, 2022.

2. V.S.Govindarajan,"AIandLaw:ThePotentialandChallengesofMachineLearningforLegalPractice,"ArtificialIntelligenceandLaw, vol. 32, pp. 233-259, 2023.

3. L.A.Maron,J.M.O'Leary,andK.W.Chan,"LeveragingAIforLegalResearch:Efficiency,Accuracy,andEthicalConsiderations,"AI & Society, vol. 37, pp. 479-497, 2021.

4. T.S.Smith,D.Y.Miller,andP.H.Green,"MachineLearningforLegalTextClassification:ACaseStudyinContractReview,"International Journal of AI & Law, vol. 31, no. 1, pp. 71-93, 2023.

5.   B.Liu,A.R.Ramirez,andR.S.Patel,"UsingAIforPredictingLegalOutcomes:ASystematicReviewofCurrentApproaches,"AIand Legal Technology Journal, vol. 9, pp. 58-82, 2022.

6.   E.T.Brown,"TheRoleofAIinEvolvingLegalAnalytics:CaseLawPredictionandBeyond,"JournalofLegal Analytics,vol.19,no. 2, pp. 134-150, 2023.

7.   J.L.Johnson,andM.T.Holtz,"ExplainableAIforLegalSystems:AnOverviewofInterpretableModelsforLegalDecisionSupport," AI in Law Journal, vol. 21, pp. 227-243, 2022.

8.   K.Zhang,J.P.Kim,andS.B.Lin,"EthicalAIforLegalServices:ChallengesandOpportunitiesinPredictiveModels,"AIEthicsReview, vol. 7, no. 3, pp. 113-126, 2023.

9.   C.R.Levenson,L.T.Williams,andJ.A.Phelps,"AI-BasedDecisionSupportforLegalProfessionals:Frameworks andBestPractices," Journal of Legal Technology, vol. 25, pp. 203-215, 2022.

10.   A.P.DeLuca,M.B.Harris,andR.M.Stevens,"FromPredictiveAnalyticstoCaseLawForecasting:TheEvolutionofLegalDecision Systems," Legal Informatics Review, vol. 18, no. 1, pp. 101-117, 2022.