The Restatement (Artificial) of Torts

Colin Doyle [1]

Abstract

This article experiments with processes for using large language models to construct a Restatements of Torts based upon the model's understanding of leading torts cases. The model's performance is evaluated by comparing its results with existing provisions from the human-written Restatements of Torts. By tracking the discrepancies between the restatements, we can gain insights into both the machine and human processes for understanding and synthesizing the law. Where the machine and human restatements converge would tend to indicate the reliability of both sources on a particular subject. Where the restatements diverge may reflect the limitations of a language model or human author, meaningful differences in how humans and machines process information, or different underlying values. This analysis has implications for the potential of large language models as tools for legal research and writing, accountability mechanisms for computer and human authorship, the function and authority of Restatements of the Law, the future of human-machine collaboration in legal practice, and the potential for machine learning to reshape the law itself.

Extended Abstract

Large language models seem poised to take over many legal research and writing tasks. (Agarwal et al., 2022). But there are serious concerns about the reliability of legal software that relies upon generative A.I. Large language models are prone to hallucinate information, cannot accurately gauge the uncertainty of their own answers, have short context windows that inhibit their ability to work with larger documents, and can struggle to perform well in niche domains — among other issues. (Yang et al., 2023b). Evaluating the performance of large language models is also difficult. (Li et al., 2023). Because large language models are terrific few-shot and zero shot-learners, they can accomplish tasks without a large dataset of examples. (Wei et al., 2022). This allows language large models to function in a variety of contexts, but it presents a challenge for evaluating their performance. With traditional machine learning methods, some data is withheld during training and used as test data to evaluate the model's performance. But when a large language model is deployed in a new context, there's often no test data available to evaluate how well the model has performed.

Restatements of the Law can be a fertile resource for evaluating LLM performance at legal tasks. Published by the American Law Institute, Restatements of the Law are influential treatises that attempt to clarify the law by articulating black letter rules, providing commentary on those rules, and demonstrating the application of those rules with illustrative examples. (Revesz, 2019). Although large language models seem up to the task of summarizing analysis of caselaw that humans have already performed, language models' competency at independently analyzing and synthesizing caselaw is uncertain. (Shukla et al., 2022; Yang et al., 2023a). To create restatement provisions, a large language model must do more than simply summarize cases. Writing restatement provisions requires resolving conflicts across jurisdictions, addressing discrepancies in how legal rules are framed, handling ambiguity and gaps in common law, reflecting trends across case law over time, and choosing the best available legal rule among multiple options. The tension at the heart of the restatement process is between, "the impulse to recapitulate the law as it presently exists and the impulse to reformulate it, thereby rendering it clearer and more coherent while subtly transforming it in the process." (201, 2015).

This article experiments with processes for using large language models to construct provisions of a Restatement of Torts based on the model's understanding of leading torts cases. We create an application that chains together separate calls to ChatGPT-4 to process caselaw and write restatement provisions. (Mialon et al., 2023; Wei et al., 2023; Mamooler et al., 2022; Wu et al., 2022; Dong et al., 2023; Lou et al., 2023). The application is initially given the raw text of a set of a cases on a particular legal issue. Following a prompt on how to take notes on legal cases, the model distills each case

[1]Loyola Law School, Los Angeles, CA. Correspondence to: Colin Doyle <colin.doyle@lls.edu>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

into a shorter casebrief that captures each case's relevant facts, legal issues, holding, and reasoning. These casebriefs are then stored in a vector database. When asked to write a particular Restatement provision, the application retrieves the relevant casebriefs and relies exclusively on the casebriefs for substantive information. A prompt based on the American Law Institute's style guide instructs the model on how to write a restatement provision. For more complicated provisions or longer texts, some steps are broken up into sequential calls to the language model. Following this multi-step process, the application can produce restatement provisions worth analyzing and evaluating.

Large language models' performance at crafting restatements can be evaluated by comparing the machine-written restatements with existing, human-written restatements. For this article, the Second and Third Restatements of Torts serve as a kind of test data for evaluating the Artificial Restatement of Torts. Inversely, machine-written restatements can function as test data for evaluating the human-written restatements. Lawyers and judges often rely on Restatements of the Law, but these treatises have their own reliability concerns. (Balganesh, 2022). Critics have argued that certain restatement provisions do not accurately reflect the caselaw or trends in the caselaw. (Levitin et al., 2019). This criticism is often coupled with an additional criticism that a restatement provision has captured the authors' preferences for what the law should be rather than what the law actually is.[1] (Merrill & Smith, 2014).

By tracking the discrepancies between machine-written restatements and human-written restatements, we can gain insights into both the machine and human processes for understanding the law. Where the two restatements converge would tend to indicate the reliability of both sources on a particular subject. Where the restatements diverge may reflect the limitations of a language model or human author, meaningful differences in how humans and machines process information, or different underlying values that would lead an author to choose one legal rule over another.

Preliminary results indicate that, when there is a general consensus in the caselaw over a legal issue, LLMs can produce restatement provisions that are virtually identical to human-written provisions on the same issue. With longer cases and more complicated legal rules, LLMs are more prone to error. Some of these errors can be overcome by decomposing the task of writing a restatement provision into separate tasks. When the caselaw is inconsistent or unclear, large language models require specific prompting for why they should choose one legal rule over another. Prioritizing different considerations — such as the current trend in caselaw, majority over minority rules, or the influence of important cases or courts — can sometimes produce quite different restatement provisions.

This analysis has implications for the potential of large language models as tools for legal research and writing, accountability mechanisms for computer and human authorship, the function and authority of Restatements of Law, the future of human-machine collaboration in legal practice, and the potential for machine learning to reshape the law itself.

[1] Kansas v. Nebraska, 574 U.S. 445, 475 (2015) (Scalia, J., concurring in part and dissenting in part) ("I write separately to note that modern Restatements . . . are of questionable value, and must be used with caution.").

References

Capturing the Voice of The American Law Institute: A Handbook for ALI Reporters and Those Who Review Their Work. Technical report, The American Law Institute, 2015.

Agarwal, A., Xu, S., and Grabmair, M. Extractive Summarization of Legal Decisions using Multi-task Learning and Maximal Marginal Relevance, October 2022.

Balganesh, S. Relying on Restatements. *Columbia Law Review*, 122(8):2119, 2022. ISSN 1556-5068.

Dong, Q., Li, L., Dai, D., Zheng, C., Wu, Z., Chang, B., Sun, X., Xu, J., Li, L., and Sui, Z. A Survey on In-context Learning, February 2023.

Levitin, A. J., Kim, N. S., Kunz, C. L., Linzer, P., McCoy, P. A., Moringiello, J. M., Renuart, E. A., and Willis, L. E. The Faulty Foundation of the Draft Restatement of Consumer Contracts. *Yale Journal on Regulation*, 36, 2019.

Li, M., Song, F., Yu, B., Yu, H., Li, Z., Huang, F., and Li, Y. API-Bank: A Benchmark for Tool-Augmented LLMs, April 2023.

Lou, R., Zhang, K., and Yin, W. Is Prompt All You Need? No. A Comprehensive and Broader View of Instruction Learning, May 2023.

Mamooler, S., Lebret, R., Massonnet, S., and Aberer, K. An Efficient Active Learning Pipeline for Legal Text Classification, November 2022.

Merrill, T. W. and Smith, H. E. Why Restate the Bundle? The Disintegration of the Restatement of Property. *Brooklyn Law Review*, 79:681, 2014.

Mialon, G., Dessì, R., Lomeli, M., Nalmpantis, C., Pasunuru, R., Raileanu, R., Rozière, B., Schick, T., Dwivedi-Yu, J., Celikyilmaz, A., Grave, E., LeCun, Y., and Scialom, T. Augmented Language Models: A Survey, February 2023.

Revesz, R. L. The Debate Over the Role of Restatements. https://www.ali.org/news/articles/debate-over-role-restatements/, August 2019.

Shukla, A., Bhattacharya, P., Poddar, S., Mukherjee, R., Ghosh, K., Goyal, P., and Ghosh, S. Legal Case Document Summarization: Extractive and Abstractive Methods and their Evaluation, October 2022.

Wei, J., Bosma, M., Zhao, V. Y., Guu, K., Yu, A. W., Lester, B., Du, N., Dai, A. M., and Le, Q. V. Finetuned Language Models Are Zero-Shot Learners, February 2022.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., and Zhou, D. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models, January 2023.

Wu, T., Terry, M., and Cai, C. J. AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. In *CHI Conference on Human Factors in Computing Systems*, pp. 1–22, New Orleans LA USA, April 2022. ACM. ISBN 978-1-4503-9157-3.

Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., Yin, B., and Hu, X. Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond, April 2023a.

Yang, S., Nachum, O., Du, Y., Wei, J., Abbeel, P., and Schuurmans, D. Foundation Models for Decision Making: Problems, Methods, and Opportunities, March 2023b.