Not ready for the bench: LLM legal interpretation is unstable and uncalibrated to human judgments

Abhishek Purushothama
Georgetown University

Junghyun Min
Georgetown University

Brandon Waldon
University of South Carolina

Nathan Schneider
Georgetown University

Abstract

Legal interpretation frequently involves assessing how a legal text, as understood by an 'ordinary' speaker of the language, applies to the set of facts characterizing a legal dispute. Recent scholarship has proposed that legal practitioners add large language models (LLMs) to their interpretive toolkit. This work offers an empirical argument against LLM-assisted interpretation as recently practiced by legal scholars and federal judges. Our investigation in English shows that models do not provide stable interpretive judgments and are susceptible to subtle variations in the prompt. While instruction tuning slightly improves model calibration to human judgments, even the best-calibrated LLMs remain weak predictors of human native speakers' judgments.

Proceedings of the Natural Legal Language Processing Workshop 2025, pages 317–317
November 8, 2025