RESEARCH ARTICLE

Computational Legal Studies Comes of Age

Bao Chau* and Michael A. Livermore[†]

Abstract

Computational analysis techniques are transforming empirical legal scholarship. Two paradigms have emerged: law-as-code, which seeks to represent legal rules in a logical, executable format; and law-as-data, which leverages quantitative analysis of legal texts to reveal patterns and insights. This article surveys these approaches, emphasizing recent developments in large language models and generative artificial intelligence (AI). Law-as-code systems have enabled applications from tax preparation software to smart contracts, but realizing the vision of fully computational law has proven challenging. Law-as-data techniques like natural language processing and machine learning have charted the semantic relationship between courts and illuminated changes in judicial culture. Generative models showcase AI's explosive progress, with impressive feats like passing the U.S. bar example, but they also highlight limitations like factual inaccuracy and interpretability issues. Hybrid approaches integrating computational law, data science, and AI offer a promising research direction. As these tools spread, legal scholars can analyze more legal data than ever before, but they must remain cognizant of challenges like biased or low-quality data and linguistic/cultural limitations. Used judiciously alongside traditional methods, computational analysis has the potential to revolutionize empirical legal studies.

The integration of computational text analysis techniques into the empirical legal studies toolkit has reshaped our understanding of and interaction with the law. By using methods from the fields of network analysis, natural language processing, and artificial intelligence/machine learning (AI/ML), legal scholars are able to analyze large digital corpora of legal texts to reveal hidden patterns and shed light on the law and legal institutions. With these new tools, researchers have charted the history of legal concepts, assessed the impact of judicial decisions, and tracked the impact of environmental regulations on economic development.

As computational text analysis becomes an essential tool of analysis, two distinct but interconnected paradigms have emerged: law-as-code and law-as-data. These paradigms represent contrasting yet complementary approaches to understanding and applying the law. Law-as-code reformulates legal rules and principles as executable instructions that can be translated into computer code. One goal of this research program is to facilitate the neutral

* Cornell University - Cornell Tech NYC; Harvard Berkman Klein Center for Internet and Society, bao-kham.chau@gmail.com, https://orcid.org/0000-0002-4866-4119.

[†] University of Virginia School of Law, mlivermore@virginia.edu, https://orcid.org/0000-0002-2403-1173.

application of law by automating legal processes through the development of computerized systems to support legal decision making. On the other hand, law-as-data focuses on harnessing the power of big data analysis. This approach focuses on analyzing large datasets of digital/digitized legal texts to extract novel insights concerning the law.

In this article, we explore the use of computational text analysis techniques in the field of empirical legal studies, with an emphasis on recent developments based on generative AI technologies. We first delve into the law-as-code versus law-as-data distinction introduced in Livermore and Rockmore (2019) and provide a survey of existing literature within these two research paradigms.[1] Next, we will discuss recent developments in generative AI and how they relate to current legal practices and scholarships. We then explore potential ways that law-as-code techniques could be used to complement law-as-data research. Finally, we will conclude by briefly examining the current opportunities and challenges for posed by computational text analysis for the field of empirical legal studies.

1   Computational Analysis and Law

1.1   Law as Code

Since the 1980s, legal scholars and computer scientists have been interested in converting legal rules into executable computer code (McCarty and Sridharan 1981; Ashley 1989; Bench-Capon 1991; Gordon 1993; Governatori 2005; Wyner et al. 2011). This law-as-code approach seeks to represent the law as a set of logical rules that a computerized system could process and execute. This avenue of research has yielded many practical applications, including smart contracts that self-execute legal agreements (Zheng et al. 2020) and tax preparation software that guides users through the labyrinth of the United States tax code to file their taxes (Soled 1996; Contos et al. 2011).

Law-as-code is a version of *knowledge representation*, a branch of AI research that focuses on representing expert information in a machine computable format (Hayes-Roth et al. 1983). Within the field of medicine, a knowledge representation system might encode the diagnostic criteria for a large number of ailments that could be automatically cross-referenced against any given patient's symptoms. A legal knowledge representation system could similarly encode some of the information of human legal experts, for example through a set of ordered checklists or questions that could be used to generate a simple contract.

An interesting early example of a law-as-code system was an effort to model a portion of the British Nationality Act, the U.K. law that deals with citizenship (Sergot et al. 1986). Legal criteria for citizenship can be represented as a series of nested if-then statements, such that a computer could convert a set of inputs concerning matters such a length of stay in the country, visa status, and so on, to generate an output concerning citizenship eligibility. One of the challenges of this type of undertaking is that statutes typically contain vague terms, such as "good character" in the Nationality Act. Vague terms grant discretion to executive or judicial officials, and their interpretation can be highly nuanced and even subjective, such that different official may apply the terms differently. The definition of the term may also evolve over time, for example due to the development of caselaw. Accordingly, there may be no single authoritative interpretation of the statute, but rather a constellation of potential interpretations, based on the local jurisdiction, deciding official, or the year of decision.

---

[1] See Frankenreiter and Livermore (2020) and Livermore and Chau (2024) for detailed methodological surveys of law-as-data techniques.

To address complex and nuanced situations that do not admit of simply binary-style if-then representations, researchers turned to other sophisticated tools used within the field of knowledge representation. One such representation relies on the use of Bayesian networks, which are probabilistic models that represent dependencies (edges) between variables (nodes) using Bayesian probability theory. Another related representation uses models of defeasible reasoning, which draws conclusions based on incomplete or uncertain information. These conclusions, however, could be revised or rebutted by new evidence or considerations as they emerge. Some of these representations have been applied to law-as-code systems, and scholars in the field have paid considerable attention to the question of the kinds of "ontologies" that can represent that law and legal reasoning.[2] Keppens (2011), for example, uses a particular type of Bayesian Network to model evidential reasoning. Keppens suggests that it is possible to translate a probabilistic, Bayesian Network representation of evidential reasoning to the more formalistic, argument-based approach. This translation would, in turn, allow people with relatively little mathematical training to take advantage of the benefits of Bayesian inferences. Merigoux et al. (2021) describes Catala, a programming language based on defeasible reasoning specifically designed to model statutory text.

Law-as-code systems have found practical applications as well. The translation of the Korean Building Act into a computer-executable format and its usage in evaluating building permit requirements (Lee et al. 2016) is a form of law-as-code knowledge representation. Commentators have also argued that knowledge representation could disrupt the legal profession, by making legal information more broadly accessible and reducing the relative advantage of legal experts (Susskind 2013). Knowledge representation systems have indeed been proposed as a way to reduce the complexity of government regulation and facilitate compliance (Coglianese 2004). To date, the most ambitious visions for law-as-code systems have not come to pass, but developments in the field are ongoing, supported by institutions such as the Stanford Center for Legal Informatics (CodeX), the Leibniz Center for Law of the University of Amsterdam, and the academic journal *Artificial Intelligence and Law*.

Most relevant to empirical legal scholars is the contribution of law-as-code research to the positive description of the law. Renton (2007), for example, used a Computer Supported Argument Visualization tool to visualize two Scottish parliamentary debates and allow interested stakeholders to determine whether and how various parliamentary issues evolve throughout the debates. Researchers have also used knowledge representation approaches to reveal hidden patterns and relationships within evidentiary law (Di Bello and Verheij 2019; Hardcastle 2018; Kotsoglou 2019; Prakken 2020). Zhou et al. (2023) created a model to extract evidence in judgment documents, which could be used to assess the quality of trials. Ruggieri et al. (2010) uses a knowledge representation approach to model direct and indirect discrimination and then apply this model to find prima facie evidence of discrimination in the German credit dataset.

Empirical legal scholars have also designed knowledge representation systems to study the judiciary (Baryse and Sarel 2023; Chen 2019; Costa et al. 2023). Taylor and Mfutso-Bengo (2023), for example, built an annotated corpus of Malawi Criminal cases and used this corpus to determine the types of criminal cases that were tried by Malawi courts from 2010 to 2019. Similarly, Adler et al. (2023) designed an AI system that analyzes a quarter-million court dockets to extract metadata concerning the U.S. courts, making this information more accessible to both technical and non-technical end-users.

---

[2] For helpful reviews of the law-as-code literature at different points in its history, see Bench-Capon et al. (2012) and Rissland et al. (2003).

As with any act of translation, there are some challenges to representing law as executable code. One such challenge relates to the multitude of different legal jurisdictions and languages. As most popular programming languages are based on the English language (Fedorenko et al. 2019), it could be potentially difficult to represent non-English legal concepts within the confine of English-based programming frameworks. In addition to technical challenges, law-as-code applications could also raise socioethical concerns regarding fairness. Liu et al. (2020), for example, notes that smart contracts could be unfair to certain participants and proposes a framework to analyze smart contract fairness. Addressing these challenges requires interdisciplinary collaboration between legal scholars, technologists, ethicists, and other stakeholders.

1.2   Law-as-data

Law-as-code researchers attempt to create a symbolic representation of the content of the law—the law-as-data approach, by contrast, focuses on the translation of legal texts into machine-readable data capable of quantitative analysis to determine the content, causes, or consequences of legal decisions. Traditionally, this law to data conversion was done manually, with human coders meticulously reading individual documents and categorizing them according to some predefined rubric (Ruger et al. 2004; Law and Zaring 2010). This method of law to data conversion, however, is both time and resource intensive. Not only do research assistants take time to be onboarded, but they could also produce unreliable and/or biased results (Hutchinson and Moran 2005; Glazier et al. 2021).

As the number of digital legal corpora grew, manual coding shifted towards automated means of transforming unstructured digital texts to structured legal data. This transformation involves a range of choices and trade-offs that differ depending on the research question (Livermore and Chau, 2024). To effectively apply computational text analysis techniques, it is thus necessary for scholars to find low-dimensional representations of legal texts that still preserve the information required to address research questions.

One of the more straightforward ways to represent legal documents is to measure their lengths. Although this particular representation strips all of the semantic context from the texts, it can still provide useful insights into said texts. Differences in lengths could indicate a difference in format, writing style, or complexity of the various corpora. Brown (2022), for example, studied American state constitutions and found that lengthier constitution tended to act as a more effective constraint on future legislatures. Osenga (2011) examined the lengths of patent documents and found this variable to be unaffected by year of filing, the technological area of the patent, or prosecution time. The author suggests that this relatively static patent length is due to an informal incentive structure that encourages patent prosecutors to draft claims that follow a highly standardized format. Finally, Bowie et al. (2023) used opinion lengths to study British lower courts' influence on the British Supreme Court. The authors found that the length of a lower court's opinion was positively correlated with influence, as measured by the extent to which the British Supreme Court borrowed language from the lower court's opinions.

To retain more context when translating raw texts into structured data, legal scholars could leverage tools using curated dictionaries of terms. A curated dictionary amounts to a list of words that are chosen to characterize some feature of a text. Sentiment analysis is an example of a tool that uses a curated dictionary. Although there are bespoke sentiment analysis computer programs, a typical sentiment analysis program usually uses a generic dictionary where words such as "bad" or "horrible" would map to a negative number while words such as "good" and "pleasant" would map to a positive number. A typical sentiment analysis

program would receive a document, break the corpus of text into individual words (or tokens), and look up the emotional valence of those words in its curated dictionary. The sentiment of a particular passage would then be expressed as a ratio of positive and negative terms.

Empirical legal scholars have used sentiment analysis to analyze judicial opinions (Budziak et al. 2019; Busch and Pelc 2019; Carlson et al. 2016; Rice and Zorn 2019). Bryan and Ringsmuth (2016) measure the sentiment of U.S. Supreme Court dissents and find that negative language in dissents positively correlates to media coverage. High public salience—which could lead to media coverage and may influence the language choices of dissenters—is one possible pathway that could explain this association. Similarly, Corley and Wedeking (2014) use a bespoke dictionary in the Linguistic Inquiry and Word Court (LIWC) tool to measure "certainty" in U.S. Supreme Court opinions and show that Supreme Court opinion with higher levels of certainty tend to be more positively treated by the lower courts in subsequent cases.

An alternative to the curated dictionary approach is to count the appearance of unique terms in a document. This computational text analysis approach is referred to as a "bag-of-words" or term-frequency vector representation. While this representation can provide a general impression of the subject matter, it can also yield misleading results. One obvious issue is that, because the order of words is ignored, simple bag-of-words representations do not capture the significance of negative auxiliary verbs (e.g., "not").

Notwithstanding their limitations, bad-of-words models have found many uses in law-as-data research. In stylometric analysis—which characterizes the writing style of a text rather than its contents—term frequency vectors often serve quite well to construct "stylistic fingerprints" that are associated with both individual authors and epochal periods (Carlson et al. 2016). This approach has been used to study the U.S. Supreme Court as well as the European Court of Justice (Carlson et al. 2016; Frankenreiter 2019). Different versions of bag-of-words models have been used to measure the lexical diversity of judicial opinions (Cheruvu 2019), the linguistic complexity in attorneys' opening and closing statements (Zubrod et al. 2020), and the similarity between opinions within the U.S. appellate courts (Hinkle 2016).

Recently, empirical legal scholars have also begun to integrate AI/ML techniques into their computational toolkits. One of the main distinctions within the fields of AI/ML approaches that is relevant to empirical legal scholars is the distinction between supervised and unsupervised algorithms. Supervised models are trained on "labeled" datasets, meaning that there is an underlying predictive task for which an outcome label is provided. An example of a supervised model would be an algorithm that predicts whether a photograph contains a cat or a dog, based on an annotated dataset. Unsupervised models are not provided with labeled data, and instead are designed to identifying different types of patterns within the underlying data, based on the purposes of the model. An example of an unsupervised model would be an outlier detector that identifies unusual observations within a dataset, based on the general statistical properties of the data.

One family of unsupervised text analysis model that legal scholars have put to broad use are "topic models," which are designed to automatically identifies latent subject matter categories within a corpus of documents (Blei 2012; Blei and Lafferty 2007). Law (2018), for example, uses topic models to analyze themes within the global discourse on human rights. Quinn et al. (2010) applies topic modeling on congressional speeches from 1997 to 2004 to determine which issues captured congressional attention during that period. Rice (2019) uses topic modeling to investigate how dissenting opinions at the U.S. Supreme Court influenced the majority's opinions. Livermore et al. (2017) combined a topic model with a

classifier to chart the evolution of the relationship between U.S. Supreme Court opinions and appellate court opinions over time.

Scholars have used supervised ML algorithms to categorize legal texts (Daniels and Rissland 1997; Gonçalves and Quaresma 2003; Hausladen et al. 2020) and to predict the outcomes of cases (Aletras et al. 2016; Medvedeva et al. 2020; Varga et al. 2021) or whether a case would be cited in a subsequent matter (Schepers et al. 2023). Evans et al. (2007) survey articles that use various supervised learning algorithms to categorize the ideological slant of briefs submitted in affirmative action cases. Alschner and Charlotin (2021) surveyed several supervised ML algorithms that are used to predict the outcomes of cases.

Recently, legal scholars have also begun to use word-embedding techniques to train supervised learning models for use in computational text analysis. These techniques capture the semantic meaning of words by mapping them to vectors of real numbers. In a word embedding, each word is represented as a vector in a continuous vector space. Nyarko and Sanga (2022) used a word-embedding model to determine whether 18th- and 21st-century English speakers have the same understanding of the word "commerce" and whether judges and layperson agree on what "reasonable conduct" means. Rice et al. (2019) use word-embedding techniques to provide evidence of implicit racial bias in U.S. state and federal courts. Choi (2024) examines the viability of corpus linguistics for aiding interpretation of legal contracts by using word-embeddings models to estimate the degree of clarity of terms in legal texts. Law-as-data research that uses word-embeddings is at the cutting edge of empirical legal scholarship. In the coming years, the role of word-embeddings in generative large language models (LLMs) is likely to thrust this technique to further prominence in empirical legal scholarship.

2   Generative Models

The release of ChatGPT by OpenAI in the late 2022 catapulted generative AI to the forefront of public imagination. Built on the groundbreaking neural network architecture described in Vaswani et al. (2017), ChatGPT is able to engage in human-like conversations with its users and produce seemingly new and realistic artifacts that are responsive to individual user's queries (i.e., "prompts"). These responses are, however, not limited to everyday conversations. In an experiment carried out by Katz et al. (2023), ChatGPT has passed the American Uniform Bar Examination (UBE) by a significant margin, scoring in the 90th percentile among all UBE bar takers.

The successes of ChatGPT and other large language models (LLMs) in performing tasks for businesses (Cromwell et al. 2023), healthcare (Dave et al. 2023), and law (Perlman 2023) has led empirical legal scholars to explore some potential uses for LLMs in legal scholarship. Choi (2023), for example, discusses how LLMs could carry out at least some of the tasks currently assigned to graduate student research assistants (RAs). In one experiment, Choi (2023) compared outputs produced by law student RAs to those generated by GPT-4 and found that GPT-4 performed just as well as law students in classifying Supreme Court opinions. Because RAs are expensive, take time to onboard, and might not produce deterministic results, Choi (2023) suggests that LLMs could replace human RAs for some empirical legal tasks.

Livermore et al. (2024) similarly compared LLMs and non-LLMs outputs for three type of legal tasks – classification of legal areas, case comparison, and estimation of the "innovation" of legal language – in order to explore the potential use of LLMs within the context of empirical legal studies. In categorizing between the various case types (i.e., administrative, bankruptcy, civil, and criminal law), the GPT-4 model significantly outperformed a topic-model based classifier. However, the GPT-3.5, BARD, and GPT-4 models somewhat underperformed

the results of a combined topic model- and citation information-based algorithm at a law search task. Finally, the authors compared the results of a dynamic topic model described in Herron et al. (2024) with those generated by an LLM in Vicinanza et al. (2022:3) and found that both approaches revealed that innovative legal language tends to be developed within the lower courts, rather than the U.S. Supreme Court. From these comparisons, Livermore et al. (2024) conclude that the field of empirical legal studies would benefit when empirical legal scholars incorporate both LLMs and non-LLMs techniques in their approach.

In a large collaboration involving many empirical legal scholars, Guha et al. (2023) constructed a legal reasoning benchmark for 162 tasks covering six different types of legal reasoning—issue-spotting, rule-recall, rule-application, rule-conclusion, interpretation, and rhetorical-understanding—in order to measure the legal reasoning capabilities of LLMs. For these legal benchmarks, the authors studied three commercial API-access models – GPT-3.5, GPT-4, and Claude-1 – and seventeen open-source models. In evaluating these models' outputs, Guha et al. (2023) use the "exact-match" criteria in Liang et al. (2022) for classification tasks. Rule-application tasks, on the other hand, were manually evaluated by coders with the help of a grading rubric.

Overall, Guha et al. (2023) found that the performances of open-source models are comparable to their commercial counterparts. The authors also discovered that some models are more suited for certain tasks than others. For example, Guha et al. (2023) noted that the WizardLM-13B open-source model performed worse than other open-source models on issue-spotting tasks. The same model, however, yielded the best average score for rule-recall tasks and the second-best score for rule-conclusion tasks. This suggests that the choice of pretraining data, instruction-tuning, and architecture play an important role in a model's performance for specific tasks.

One limitation of existing LLMs for empirical legal research is their tendency to hallucinate. LLM hallucination is a phenomenon where a model generates text that is factually incorrect or nonsensical. In the latter half of 2023, for example, news outlets have reported several instances of ChatGPT generating fictitious caselaw (Geoghegan 2023; Merken 2023). Similar to other types of LLM hallucinations, this type of legal hallucination could be caused by a combination of the model having incomplete or noisy training data, receiving vague prompts or questions, becoming too tailored to the training dataset (i.e., "overfitting"), or lacking real-world context. It is therefore essential for legal scholars to familiarize themselves with known methods for minimizing hallucinations (e.g., training LLMs with adversarial examples designed to cause hallucination, applying post-processing filter to remove inconsistent outputs, and incorporating human feedback into the training process).

In addition to hallucinations, it might be difficult to trust LLM outputs due to interpretability challenges. While the fundamental mechanics of how information is processed within LLMs are well-understood, the complexity of LLM architectures make it highly challenging to determine how a specific LLM arrived at a particular conclusion. Indeed, even the simplest LLM architecture contains multiple layers of neural networks, making it extremely difficult to understand how these models process inputs to generate outputs. The large number of parameters in LLMs further compounds this interpretability problem. GPT-3, for example, has a staggering 175 billion parameters (Brown et al. 2020) while the newest GPT-4 has an estimate of around 1.8 trillion parameters (Schreiner 2023).

Limited insights into the datasets used to train LLMs could also undermine the level of trust in the LLMs' outputs. While OpenAI researchers divulged that ChatGPT's training dataset consisted of many datasets publicly accessible on the Internet, they did not disclose all of their sources (Radford et al. 2018). Understanding the provenance of the dataset is important, as biased inputs would invariably produce biased outputs (Chau 2022).

3   Hybrid Approaches

At a theoretical level, the law-as-code and law-as-data research programs reflect different concepts of the nature of law itself. The law-as-code approach fits more comfortably within a formalist jurisprudential paradigm that understands legal rules as abstract principles that can be applied in a relatively neutral and deterministic fashion. The law-as-data approach, by contrast, is agnostic regarding the formal status of the law and whether, for example, there are demonstrably correct answers to legal question. Instead, law-as-data scholars treat the law as a social phenomenon that has causes and consequences that can be studied through social scientific tools of modeling and empirical analysis.

Notwithstanding these different theoretical foundations, there is potential value to be had in hybridized research that combine insights from the law-as-code and law-as-data approaches. Even diehard legal realists typically accept that the positive law affects how legal questions are constructed, argued, and even decided, and that within some zone of understanding, the content of the law (as filtered through widely shared values and interpretive practices) is relatively clear. And although some legal formalists might believe that the law should aspire toward the kind of neutrality, clarity, and completeness that would lead to deterministic outcomes, most recognize that no actual legal system lives up to these aspirations. Accordingly, research that draws from both law-as-data and law-as-code has the potential to offer a more nuanced characterization of the law and legal institutions.

Zufall et al. (2022) provides an example of how to integrate law-as-code and law-as-data approaches. The first step is to model the EU Framework Decision on Hate Speech as a series of binary decisions. This is a standard law-as-code representation of a statutory text. The authors then use this representation to assist a group of human coders in annotating a dataset. With a contemporary LLM, it is possible that this annotation step could be undertaken by an LLM, perhaps with some level of supervision by human coders. In Zufall et al. (2022), the annotated dataset was then used to train a ML classifier, which could then be used to analyze whether new hypothetical scenarios violate the Framework Decision.

Another example of a potential hybrid approach is found in Dadgostari et al. (2021). That project (referred to as *LexQuery* in Livermore et al. 2021) drew from the law-as-code tradition to describe a formal mathematical model of law search, and then, drawing from the law-as-data tradition, analyzed a corpus of U.S. Supreme Court opinions to parametrize that model. The mathematical model was based on a multinetwork in which judicial opinions were linked to each other based on citation information as well as semantic similarity, as estimated via a topic model. This base model was then combined with a reinforcement learner that was trained on the task of predicting the citations in a text from only the topic model representation of that text. One goal of that combined model was to estimate the relative weight that law searchers place on broad coverage of legal issues, versus in depth exploration. The algorithm also effectively asks as a search recommendation system, one that performs reasonably well compared to a baseline of human searchers.

In addition, LLMs can be understood as a law-as-data approach, inasmuch as, within the large datasets used to the train the model, legal information is included, and the text generated from the models are the results of patterns in legal texts that are implicitly detected by the models. The opacity of LLMs and the problem of interpretation that they pose may limit their use for legal tasks and empirical legal scholarship. However, it may be possible to integrate LLMs with law-as-code and more traditional law-as-data approaches in ways that draw from the relative strengths of these different paradigms.

One potential hybrid approach engendered by LLMS would be to integrate the formalized representation of a legal text with an LLM-based classifier capable of addressing vague

terms. Versions of this approach were proposed in Wolfram (2016) and Livermore (2020). For example, as discussed above, a classic law-as-code project was to create a formal representation of the British Nationality Act, but that effort faced the difficulty of vague terms such as "good character." An LLM-based classifier could, in principle, take as inputs natural language descriptions of individual cases and then draw conclusions concerning whether the descriptions represented (or not) "good character" according to legal standard. The LLM would, in essence, make a prediction concerning how a human decision maker would approach a particular case. For cases in which human decision makers disagree, the LLM could be trained to deliver the most likely human response (in essence the majority answer). For these systems to be deployed in practice, it would be necessary for their outputs to be extensively tested. It may also be necessary to incorporate additional layers of human feedback or review in order to fulfill procedural justice requirements (Livermore 2020).

The *LexQuery* search system is also an example of a legal model that could potentially be improved through integration with an LLM. One possibility would be to use the outputs of *LexQuery* as a training data for the LLM as part of the fine-tuning process. Another would be to instruct the LLM to make queries to the law-as-data model and then incorporate the resulting returns within its replies. More ambitiously, an effort could be made to draw lessons from the construction of *LexQuery* to create a generative model that is partially based on a semi-structured representation of documents, rather than the raw-text approach used in the current generation of LLMs. Research in other areas show that LLM hallucinations can be mitigated by such techniques. For example, Agrawal et al. show that knowledge representation techniques such as knowledge-graph-based augmentation techniques do help in reducing the frequency of LLM hallucinations (Agrawal et al. 2023).

4 Opportunities and Challenges

Computational text analysis techniques stand poised to revolutionize empirical legal research. Used properly, scholars could now delegate rote tasks to computerized assistants, giving them more time to carefully analyze relevant texts and answer questions previously intractable. Researchers, for example, could quickly "read" through large volumes of Supreme Court oral argument transcripts to determine if an attorney's gender plays a role in the number of times that the attorney was interrupted (Paxton and Smith 2017). Similarly, text analysis techniques could ingest copious amount of data and provide a bird eye's view, yielding novel patterns and insights (Livermore and Rockmore 2019). Olsen and Küçüksu (2017) analyzed a large number of decisions rendered by the European Court of Human Rights to reveal new insights into the prohibition of discrimination under relevant laws.

As with any powerful tools, it is necessary to understand their limitations. Sophisticated research tools will not compensate for poorly structured research questions or faulty data. Because these tools are primarily designed by English speakers to analyze English texts, they are not well equipped to handle non-English documents (Dombrowski 2020). Most tools, for example, rely on whitespaces to identify the various "words." However, in Japanese or Chinese, many documents do not use this convention. Thus, a phrase in these languages could be erroneously considered as a single word. In addition to being language-specific, these computational text analysis techniques could also be designed for a particular domain (Perry and Benoit 2017). A tool designed for one field (e.g., transactional law) might produce erroneous results for another (e.g., litigation).

A more prominent problem involves the quantity, quality, and scope of data. The oft-repeated phrase "garbage in, garbage out" encapsulates this concept. Biased data could cause the tools to produce erroneous result (Chau 2022). Generative AI, for example, often

fabricate fictitious text based on the patterns that it learns from incomplete datasets (Siontis et al. 2023). Because of this limitation, it is necessary to involve a human in the loop to check and appropriately interpret the algorithmically generated outputs.

Notwithstanding these challenges, computational text analysis techniques are quickly becoming indispensable parts of an empirical legal scholar's toolkit. Used in conjunction with traditional legal research methodologies, these techniques promise to open new avenues of research that could revolutionize the study of law.

References

Adler, Rachel F., Andrew Paley, Andong L. Li Zhao, Harper Pack, Sergio Servantez, Adam R. Pah, Kristian Hammond, and SCALES OKN Consortium. 2023. "A user-centered approach to developing an AI System analyzing U.S. federal court data." *Artificial Intelligence and Law* 31 (September): 547–70. https://doi.org/10.1007/s10506-022-09320-z.

Aletras, Nikolaos, Dimitrios Tsarapatsanis, Daniel Preoţiuc-Pietro, Vasileios Lampos. 2016. "Predicting judicial decisions of the European Court of Human Rights: a natural language processing perspective." *PeerJ Computer Science* 2:e93 (October 24). https://doi.org/10.7717/peerj-cs.93.

Agrawal, Garima, Tharindu Kumarage, Zeyad Alghami, and Huan Liu. 2023. "Can Knowledge Graphs Reduce Hallucinations in LLMs?: A Survey." *arXiv preprint* arXiv:2311.07914. https://doi.org/10.48550/arXiv.2311.07914.

Ashley, Kevin D. 1989. "Toward a computational theory of arguing with precedents." In *Proceedings of the 2nd international conference on Artificial intelligence and law*, 93–102. New York: Association for Computing Machinery.

Alschner, Wolfgang and Damien Charlotin. 2021. "Data Mining, Text Analytics, and Investor-State Arbitration". Forthcoming in *International Arbitration and Technology,* edited by Pietro Ortolani et al. Wolters Kluwer, Ottawa Faculty of Law Working Paper. No. 2021-17. https://dx.doi.org/10.2139/ssrn.3857127.

Barysé, Dovilé and Roee Sarel. 2023. "Algorithms in the court: does it matter which part of the judicial decision-making is automated?" *Artificial Intelligence and Law* 32 (January 8): 117–46. https://doi.org/10.1007/s10506-022-09343-6.

Bench-Capon, Trevor J. M., ed. 1991. *Knowledge-Based Systems and Legal Applications*. San Diego, CA: Academic.

Bench-Capon, Trevor J. M., principal author. 2012. "A History of AI and Law in 50 Papers: 25 Years of the International Conference on AI and Law." *Artificial Intelligence and Law* 20 (September 29): 215–319. https://doi.org/10.1007/s10506-012-9131-x.

Blei, David M. 2012. "Probabilistic topic models." *Communications of the ACM* 55, no. 4 (April 1): 77–84. https://doi.org/10.1145/2133806.2133826.

Blei, David M. and John D. Lafferty. 2007. "A correlated topic model of science." The *Annals of Applied Statistics* 1, no. 1 (June): 17–35. https://doi.org/ 10.1214/07-AOAS114.

Bowie, Jennifer, Ali S. Masood, Elisha C. Savchak, Natalie Smith, Bianca Wieck, Cameron Abrams, and Meghna Melkote. 2023. "Lower Court Influence on High Courts: Evidence from the Supreme Court of the United Kingdom." *Journal of Law and Courts* 12, no. 1 (published online September 4): 1–22. https://doi.org/10.1017/jlc.2023.18.

Brown, Adam R. 2022. *The Dead Hand's Grip: How Long Constitutions Bind States*. Oxford: Oxford University Press.

Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhari-wal, Arvid Neelakantan, Parav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Akec Radford, Ilya Sutskever, and Dario Amodei. 2020. "Language models are few-shot learners." *Advances in neural information processing systems* 33: 1877–901.

Bryan, Amanda C. and Eve M. Ringsmuth. 2016. "Jeremiad or weapon of words?: The power of emotive language in Supreme Court dissents." *Journal of Law and Courts* 4, no. 1 (Spring): 159–85. https://doi.org/10.1086/684788.

Budziak, Jeffrey, Matthew P. Hitt, and Daniel Lempert. 2019. "Determinants of Writing Style on the United States Circuit Courts of Appeals." *Journal of Law and Courts* 7, no. 1 (Spring): 1–28. https://doi.org/10.1086/701128.

Busch, Marc L. and Krzysztof J. Pelc. 2019. "Words matter: how WTO rulings handle controversy." *International Studies Quarterly* 63, no. 3 (September): 464–76. https://doi.org/10.1093/isq/sqz025.

Carlson, Keith, Michael A. Livermore, and Daniel Rockmore. 2016. "A quantitative analysis of writing style on the U.S. Supreme Court." *Washington University Law Review* 93, no. 6: 1461–510.

Chau, Bao Kham. 2022. "Governing the Algorithmic Turn: Lyft, Uber, and Disparate Impact." *William and Mary Center for Legal and Court Technology* (June 30).

Chen, Daniel L. 2019. "Judicial analytics and the great transformation of American Law." *Artificial Intelligence and Law* 27 (March): 15–42. https://doi.org/10.1007/s10506-018-9237-x.

Cheruvu, Sivaram. 2019. "How do institutional constraints affect judicial decision-making? The European Court of Justice's French language mandate." *European Union Politics* 20, no. 4 (July 12): 562–83. https://doi.org/10.1177/1465116519859428.

Choi, Jonathan H. 2023. "How to use large language models for empirical legal research." *Journal of Institutional and Theoretical Economics (Forthcoming)*.

Choi, Jonathan H. 2024. "Measuring Clarity in Legal Text." *University of Chicago Law Review (Forthcoming).*

Coglianese, Cary. 2004. "E-rulemaking: information technology and the regulatory process." *Administrative Law Review* 56, no. 2 (Spring 2004): 353– 402.

Contos, George, John Guyton, Patric Langetieg, and Melissa Vigil. 2011. "Individual taxpayer compliance burden: the role of assisted methods in taxpayers response to increasing complexity." In *IRS Research Bulletin: Proceedings of the IRS Research Conference 2010*, ed. Martha E. Gangi, Alan Plumley, 191–220. Washington, DC: Intern. Revenue Serv.

Corley, Pamela C. and Justin Wedeking. 2014. "The (dis)advantage of certainty: The importance of certainty in language." *Law & Society Review* 48, no. 1 (March): 35–62. https://doi.org/10.1111/lasr.12058.

Costa, Yuri D. R., Hugo Oliveira, Valério Nogueira Jr., Lucas Massa, Xu Yang, Adriano Barbosa, Krerley Oliveira, and Thales Vieira. 2023. "Automating petition classification in Brazil's legal system: a two-step deep learning approach." *Artificial Intelligence and Law*. https://doi.org/10.1007/s10506-023-09385-4.

Cromwell, Johnathan R., Jean-François Harvey, Jennifer Haase, and Heidi K. Gardner. 2023. "Discovering Where ChatGPT Can Create Value for Your Company." *Harvard Business Review* (June 9).

Dadgostari, Faraz, Mauricio Guim, Peter A. Beling, Michael A. Livermore, and Daniel N. Rockmore. 2021. "Modeling law search as prediction." *Artificial Intelligence and Law* 29 (March): 3–34. https://doi.org/10.1007/s10506-020-09261-5.

Daniels, Jody J., and Edwina L. Rissland. 1997. "Finding legally relevant passages in case opinions." In *Proceedings of the 6th International Conference on Artificial Intelligence and Law*, 39–46. New York: Assoc. Comput. Mach.

Dave, Tirth, Sai Anirudh Athaluri, and Satayam Singh. 2023. "ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations." *Frontiers in Artificial Intelligence* 6 (May). https://doi.org/10.3389/frai.2023.1169595.

Di Bello, Marcello, and Bart Verheiji. 2019. "Evidence and decision making in the law: theoretical, computational and empirical approaches." *Artificial Intelligence and Law* 28 (June 22): 1–5.

Dombrowski, Quinn. 2020. "Preparing Non-English texts for computational analysis." *Modern Languages Open* (August 28). https://doi.org/10.3828/mlo.v0i0.294.

Evans, Michael, Wayne McIntosh, Jimmy Lin, and Cynthia Cates. 2007. "Recounting the courts? Applying automated content analysis to enhance empirical legal research." *Journal of Empirical Legal Studies* 4, no. 4 (December 10): 1007–39. https://doi.org/10.1111/j.1740-1461.2007.00113.x.

Fedorenko, Evelina, Anna Ivanova, Riva Dhamala, and Marina Umaschi Bers. 2019. "The language of programming: a cognitive perspective." *Trends in cognitive sciences* 23, no. 7 (July): 525–528. https://doi.org/10.1016/j.tics.2019.04.010.

Frankenreiter, Jens. 2019. "Writing style and legal traditions." See Livermore and Rockmore 2019, 153–90.

Frankenreiter Jens and Michael A. Livermore. 2020. "Computational methods in legal analysis." *Annual Review of Law and Social Science* 16 (October): 39–57. https://doi.org/10.1146/annurev-lawsocsci-052720-121843.

Geoghegan, Clara. 2023. Colorado Lawyer Cited Fake Cases in Motion Written with ChatGPT. *LawWeek Colorado* (June 21).

Glazier, Rebecca A., Amber E. Boydstun, and Jessica T. Feezell. 2021. "Self-coding: A method to assess semantic validity and bias when coding open-ended responses." *Research & Politics* 8, no. 3 (July 27). https://doi.org/10.1177/20531680211031752.

Gonçalves, Teresa and Paulo Quaresma. 2005. "Is linguistic information relevant for the classification of legal texts?" In *Proceedings of the Tenth International Conference on Artificial Intelligence and Law*, 168–76. New York: Assoc. Comput. Mach. https://doi.org/10.1145/1165485.1165512.

Gordon, Thomas F. 1993. "The Pleadings Game." *Artificial Intelligence Law* 2: 239–92.

Governatori, Guido. 2005. "Representing business contracts in RuleML." *International Journal of Cooperative Information Systems* 14, no. 02n03: 181–216. https://doi.org/10.1142/S0218843005001092 .

Guha, Neel, Julian Nyarko, Daniel E. Ho, Christopher Ré, Adam Chilton, Aditya Narayana, Alex Chohlas-Wood, Austin Peters, Brandon Waldon, Daniel N. Rockmore, Diego Zambrano, Dmitry Tallsman, Enam Hoque, Faiz Surani, Frank Fagan, Galit Sarfaty, Gregory M. Dickinson, Haggai Porat, Jason Heglan, Jessica Wu, Joe Nudell, Joes Niklaus, John Nay, Jonathan H. Choi, Kevin Tobia, Margaret Hagan, Megan Ma, Michael Livermore, Nikon Rasumov-Rahe, Nils Holzenberger, Noam Kolt, Peter Henderson, Sean Rehaag, Sharad Goel, Shang Gao, Spencer Williams, Sunny Gandhi, Tom Zur, Varun Iyer, and Zehua Li. 2023. "Legalbench: A collaboratively built benchmark for measuring legal reasoning in large language models." *arXiv preprint* arXiv:2308.11462.

Hardcastle, Gray Valerie. 2018. "Group-to-individual (G2i) inferences: challenges in modeling how the U.S. court system uses brain data." *Artificial Intelligence and Law* 28 (October 10): 51–68. https://doi.org/10.1007/s10506-018-9234-0.

Hausladen, Carina I., Marchel H. Schubert, and Elliot Ash. 2020. "Text classification of ideological direction in judicial opinions." *International Review of Law and Economics* 62:105903 (June). https://doi.org/10.1016/j.irle.2020.105903.

Hayes-Roth, Frederick, Donald A. Waterman, and Douglas B. Lenat. 1983. *Building Expert Systems*. Boston: Addison-Wesley Longman Publishing.

Herron, Felix, Keith Carlson, Michael A. Livermore, and Daniel N. Rockmore. 2024. "Judicial Hierarchy and Dynamic Discursive Influence in the U.S. Courts." *Philosophical Transactions of the Royal Society* (forthcoming).

Hinkle, Rachael K. 2016. "Strategic anticipation of en banc review in the US courts of appeals." *Law & Society Review* 50, no. 2 (June): 383–414. https://doi.org/10.1111/lasr.12199.

Hutchinson, Terry C. and Joanne Moran. 2005. "The Use of Research Assistants in Law Faculties: Balancing Cost Effectiveness and Reciprocity." In *Proceedings Faculty of Law Research Interest Group* 1–17.

Katz, Daniel Martin, Michael James Bommarito, Shang Gao, and Pablo Arredondo. 2023. "Gpt-4 passes the bar exam." *Available at*: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4389233.

Keppens, Jeroen. 2011. "On extracting arguments from Bayesian network representations of evidential reasoning." In *Proceedings of the thirteenth international conference on artificial intelligence and law*, 141–50. New York: ACM Press.

Kotsoglou, Kyriakos N. 2019. "Proof beyond a context-relevant doubt. A structural analysis of the standard of proof in criminal adjudication." *Artificial Intelligence and Law* 28 (March 18): 111–33. https://doi.org/10.1007/s10506-019-09248-x.

Law, David S. 2018. "The global language of human rights: a computational linguistic analysis." *Law Ethics Hum. Rights* 12, no. 1 (June 21): 111–50. https://doi.org/10.1515/lehr-2018-0001.

Law, David S., David Zaring. 2010. "Law versus ideology: the Supreme Court and the use of legislative history." *William & Mary Law Review* 51, no. 5: 1653–747.

Lee, Hyunsoo, Jin-Kook Lee, Seokyung Park, and Inhan Kim. 2016. "Translating building legislation into a computer-executable format for evaluating building permit requirements." In *Automation in Construction* 71, edited by Mikko Malaska and Rauno Heikkilä, 49–61. Amsterdam: Elsevier Science Publishers.

Liu, Ye, Yi Li, Shang-Wei Lin, and Rong Zhao. 2020. "Towards automated verification of smart contract fairness." In *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundation of Software Engineering*, 666–677. New York: Association for Computing Machinery.

Livermore, Michael A. and Bao Kham Chau. 2024. "Studying Judicial Behaviour with Text Analysis." In *Oxford Handbook of Comparative Judicial Behaviour*. Oxford: Oxford University Press.

Livermore, Michael A., Felix Herron, and Daniel Rockmore. 2024. "Language Model Interpretability and Empirical Legal Studies." *Virginia Public Law and Legal Theory Research Paper*. Journal of Institutional and Theoretical Economics (forthcoming).

Livermore, Michael A., Peter Beling, Keith Carlson, Faraz Dadgostari, Mauricio Guim, and Daniel N. Rockmore. 2021. "Law search in the age of the algorithm." *Michigan State Law Review* 1183 (April): 1183–240.

Livermore, Michael A. 2020. "Rule by rules." In *Computational Legal Studies: The Promise and Challenge of Data-Driven Research,* edited by Ryan Whalen, 238–264. Edward Elgar.

Livermore, Michael A. and Daniel N. Rockmore. 2019. *Law as Data: Computation, Text, and the Future of Legal Analysis*. Santa Fe: Santa Fe Institute Press.

Livermore, Michael A., Allen B. Riddell, and Daniel N. Rockmore. 2017. "The Supreme Court and the judicial genre." *Arizona Law Review* 59, no. 4: 837–901.

Merigoux, Denis, Nicolas Chataing, Jonathan Protzenko. 2021. "Catala: A Programming Language for the Law." *Proceedings of the ACM on Programming Languages* 5, no. ICFP: 77:1–29.

McCarty, Thorne L. and N.S. Sridharan. 1981. "The Representation of an Evolving System of Legal Concepts: II. Prototypes and Deformations." In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, 246–53. San Francisco: Morgan Kaufmann Publishers.

Medvedeva, Masha, Michel Vols, and Martijn Wieling. 2020. "Using machine learning to predict decisions of the European Court of Human Rights." *Artificial Intelligence and Law* 28 (June): 237–66. https://doi.org/10.1007/s10506-019-09255-y.

Merken, Sara. 2023. "New York lawyers sanctioned for using fake ChatGPT cases in legal brief." *Reuters* (June 26).

Nyarko, Julian and Sarath Sanga. 2022. "A Statistical Test for Legal Interpretation: Theory and Applications." *The Journal of Law, Economics, and Organization* 38, no. 2 (July): 539–69. https://doi.org/10.1093/jleo/ewab038.

Olsen, Henrik Palmer and Aysel Kuҫuksu. 2017. "Finding hidden patterns in ECtHR's case law: on how citation network analysis can improve our knowledge of ECtHR's Article 14 practice." *International Journal Discrimination and the Law* 17, no. 1 (February 28)*:* 4–22. https://doi.org/10.1177/1358229117693715.

Osenga, Kristen. 2012. "The Shape of Things to Come: What We Can Learn from Patent Claim Length." *Santa Clara High Technology Law Journal* 28, no. 3: 617–56.

Patton, Dana and Joseph L. Smith. 2017. "Lawyer, interrupted: Gender bias in oral arguments at the US Supreme Court." *Journal of Law and Courts* 5, no. 2 (Fall): 337–61. https://doi.org/10.1086/692611.

Perlman, Andrew. 2023. "The Implications of ChatGPT for Legal Services and Society." *The Practice: Generative AI In the Legal Profession* (March/April).

Prakken, Henry. 2020. "A new use case for argumentation support tools: supporting discussions of Bayesian analyses of complex criminal cases." *Artificial Intelligence and Law* 28 (March): 27–49. https://doi.org/10.1007/s10506-018-9235-z.

Quinn, Kevin M., Burt L. Monroe, Micheal Colaresi, Michael H. Crespin, and Dragomir R. Radev. 2010. "How to analyze political attention with minimal assumptions and costs." *American Journal of Political Science* 54, no. 1 (January): 209–28. https://doi.org/10.1111/j.1540-5907.2009.00427.x.

Radford, Alec, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. "Improving language understanding by generative pre-training." *OpenAI* (June 11).

Renton, Alastair. 2007. "Seeing the point of politics: exploring the use of CSAV techniques as aids to understanding the content of political debates in the Scottish Parliament." *Artificial Intelligence and Law* 14: 277–304. https://doi.org/10.1007/s10506-007-9040-6.

Rice, Douglas. 2019. "Measuring the issue content of Supreme Court opinions." *Journal of Law and Courts* 7, no. 1 (Spring): 107–27. https://doi.org/10.1086/701130.

Rice, Douglas, Jesse H. Rhodes, and Tatishe Nteta. 2019. "Racial bias in legal language." *Research & Politics* 6, no. 2 (May 14). https://doi.org/10.1177/2053168019848930.

Rice, Douglas and Christopher Zorn. 2019. "Corpus-based dictionaries for sentiment analysis of specialized vocabularies." *Political Science Research and Methods* 67: 1–16. https://doi.org/10.1017/psrm.2019.10.

Rissland, Edwina L., Kevin D. Ashley, and R.P. Loui. 2003. "AI and Law: A Fruitful Synergy." *Artificial Intelligence* 150, no. 1–2 (November): 1–15. https://doi.org/10.1016/S0004-3702(03)00122-X.

Ruger, Theodore W., Pauline T. Kim, Andrew D. Martin, and Kevin M. Quinn. 2004. "The Supreme Court forecasting project: legal and political science approaches to predicting Supreme Court decision making." *Columbia Law Review* 104, no. 4 (May): 1150–209.

Ruggieri, Salvatore, Dino Pedreschi, and Franco Turini. 2010. "Integrating induction and deduction for finding evidence of discrimination." *Artificial Intelligence and Law* 18 (March): 1-43. https://doi.org/10.1007/s10506-010-9089-5.

Schepers, Iris, Masha Medvedeva, Michelle Bruijn, Martijn Wieling, and Michel Vols. 2023. "Predicting citations in Dutch case law with natural language processing." *Artificial Intelligence and Law* (June 28): 1–31. https://doi.org/10.1007/s10506-023-09368-5.

Schreiner, Maximilian. 2023. "GPT-4 architecture, datasets, costs and more leaked." *The Decoder* (July 11).

Sergot, Marek J., Fariba Sadri, Robert A. Kowalsi, Frank R. Kriwaczek, Peter Hammond, and H. T. Cory. 1986. "The British Nationality Act as a logic program." *Communications of the ACM* 29, no. 5 (May 1): 370–86. https://doi.org/10.1145/5689.5920.

Siontis, Kostantinos C., Zachi I. Attia, Samuel J. Asirvatham, and Paul F. Friedman. 2023. "ChatGPT hallucinating: can it get any more humanlike?" *European Heart Journal* 45, no. 5 (December 13): 321–23. https://doi.org/10.1093/eurheartj/ehad766.

Soled, Jay A. 1996. "Computers, Complexity, and the Code: Dawn of a New Era." *Tax Notes Today* 73:471.

Susskind, Richard. 2013. *Tomorrow's Lawyers: An Introduction to Your Future*. Oxford: Oxford University Press.

Taylor, Amelia V. and Eva Mfutso-Bengo. 2023. "Towards a machine understanding of Malawi legal text." *Artificial Intelligence and Law* 31 (March): 1–11. https://doi.org/10.1007/s10506-021-09303-6.

Varga, Dávid, Zoltán Szoplák, Stanislav Krajci, Pavol Sokol, and Peter Gurský. 2021. "Analysis and Prediction of Legal Judgements in the Slovak Criminal Proceedings."

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez , Lukasz Kaiser, and Illia Polosukhin. 2017. "Attention is all you need." *Advances in neural information processing systems*. https://doi.org/10.48550/arXiv.1706.03762.

Vicinanza, Paul, Amir Goldberg, and Sameer B. Srivastava. 2023. "A Deep-Learning Model of Prescient Ideas Demonstrates that They Emerge from the Periphery." *PNAS Nexus* 2, no. 1 (January): 1–11. https://doi.org/10.1093/pnasnexus/pgac275.

Wolfram, Stephen. 2016. "Computational Law, Symbolic Discourse, and the AI Constitution." *Wired.com* (12 October).

Wyner, Adam Z., Trevor J. M. Bench-Capon, and Katie M. Atkinson. 2011. "Towards formalising argumentation about legal cases." In *Proceedings of the 13th International Conference on Artificial Intelligence and Law*. New York: Association for Computing Machinery. https://doi.org/10.1145/2018358.2018359.

Zheng, Zibin, Shaoan Xie, Hong-Ning Dai, Weili Chen, Xiangping Chen, Jian Weng, and Muhammad Imran. 2020. "An overview on smart contracts: Challenges, advances and platforms." In *Future Generation Computer Systems* 105 (April), 475–491. Amsterdam: Elsevier Science Publishers.

Zhou, Yulin, Lijuan Liu, Yanping Chen, Ruizhang Huang, Yongbin Qin, and Chuan Lin. 2023. "A novel MRC framework for evidence extracts in judgment documents." *Artificial Intelligence and Law* 32 (January). https://doi.org/10.1007/s10506-023-09344-z.

Zubrod, Alivia, Lucian Gideon Conway III, Kathrene R. Conway, and David Ailanjian. 2020. "Understanding the Role of Linguistic Complexity in Famous Trial Outcomes." *Journal of Language and Social Psychology* 40, no. 3 (September 13): 354–77. https://doi.org/10.1177/0261927X20958439.

Zufall, Frederike, Marius Hamacher, Katharina Kloppenborg, Torsten Zesch. 2022. "A Legal Approach to Hate Speech: Operationalizing the EU's Legal Framework against the Expression of Hatred as an NLP Task." *arXiv preprint arXiv:2004.03422*. https://doi.org/10.48550/arXiv.2004.03422.