



# From Data to Justice: How Machine Learning Advances Legal Analytics

**Dahlak Daniel Solomon**<sup>1</sup> 

Taipei Medical University, Taiwan

**Renu Pal Sood**<sup>2</sup>,  **Malvika Gupta**<sup>3</sup> 

Shoolini University, India

**Rajat Singh**<sup>4</sup> 

Ministry of Labour and Employment, Government of India, India

## Abstract

Machine Learning (ML) is revolutionizing legal analytics by facilitating a more evidence-based, efficient, and transparent approach to legal decision-making. This paper examines the application of machine learning (ML) in legal practice through fundamental methodologies, including supervised and unsupervised learning, natural language processing (NLP), and the interpretability of trained models, as well as practical applications and use cases in areas such as case prediction, contract review, compliance monitoring, and sentiment analysis. The paper also explores important ethical, legal, and operational challenges associated with ML in the legal space, namely bias, privacy, occlusion, and liability. By calling for an interdisciplinary lens, the paper highlights emerging approaches, such as fairness-aware and privacy-preserving ML, while noting the idea of having both governance frameworks in place and a common language for responsibly deploying ML. In the end, the analysis emphasizes ethically-informed, transparent ML systems that advance legal integrity and ultimately access to justice.

Keywords: Machine learning, Legal analytics, Predictive analytics, Ethical AI, Data privacy, Algorithmic bias, Legal technology, Transparency, Accountability.

<sup>1</sup> IIPP Intern, Taipei Medical University, Taiwan. ORCID ID: <https://orcid.org/0000-0003-1836-733X>

<sup>2</sup> Associate Professor (Law), Shoolini University, Solan, Himachal Pradesh, India. Email: [renupalsood@shooliniuniversity.com](mailto:renupalsood@shooliniuniversity.com) ORCID ID: <https://orcid.org/0009-0003-4005-2120>

<sup>3</sup> Research Scholar, Shoolini University, Solan, Himachal Pradesh, India. Email: [advmalvika60@gmail.com](mailto:advmalvika60@gmail.com) ORCID ID: <https://orcid.org/0009-0003-8608-3133>

<sup>4</sup> Enforcement Officer, Employees' Provident Fund Organisation, Ministry of Labour and Employment, Government of India. ORCID ID: <https://orcid.org/0009-0008-1996-6928>

## 1. Introduction

The legal profession has undergone a substantial shift in the past few years, fuelled by the rapid rise of digital data and evolving computational technologies. Data analytics has emerged as one of the foremost technological disruptors shaping how legal practitioners access, structure, and respond to the world of legal knowledge—as well as the decisions of legal strategy itself—by digital records of widespread digitization of legal records, court opinions, contracts, and regulatory documents, large-scale legal data, is not only available, but more often than not, available for more advanced forms of computational analysis (Brooks et al., 2020; Greenleaf et al., 2018; Krishna, 2023). This transition has led away from intuition and precedent based reasoning to one that is informed by data analytics, which increases the overall levels of transparency, promotes efficiency, and assists with legal foresight (Caserta, 2020).

Simultaneously, the use of machine learning (ML) has created unprecedented opportunities within legal analytics to derive insights from legal datasets that are often complex and huge in size. Machine learning is a subset of artificial intelligence whereby systems can learn from data to be able to identify patterns and develop predictive outputs. With ML, we can automate many different matters that would take a long time to ask human lawyers and have significant thought put into them; we can now have the answers provided to us, without the long wait for calculating by humans or the considerable cost associated with using human labor.

Additionally, through ML, we can utilize the predictive qualities that the tech ML brings, conduct litigation trend analysis, automate review of documents and more, enhance our review of contracts, and accomplish more tasks in lower time and cost outputs. ML technology is changing the scope and scale of all legal services (Mowbray et al, 2020; William, 2023) and, importantly, developing better efficiencies and better ways to meet complex legal issues in the process.

Nevertheless, ML in the legal space is not without risk and limitations. As machine learning has the potential to improve accuracy, lower costs and provide access to justice (Markou & Deakin, 2019), it also introduces serious ethical and regulatory implications especially for: algorithmic bias originating from training data as these biases could be built into the model, opacity of “black box” ML models (not knowing how conclusions are reached), and who is accountable for legal decisions made by automated processes (Milaninia, 2020; Schmidt et al., 2020). In addition, sensitivity about legal data for model training must comply with strict privacy and data protection regulations. As such, ethical and socially responsible adoption of ML in legal practice will require navigating an evolving technological and legal environment (Soukupová, 2021).

This article originated from the recognition that machine learning is set to become an integral part of the next phase of legal analytics because of its ability to process, analyze and extract insights from large amounts of unstructured legal data. This may change the way we perceive and execute justice. ML stands to streamline processes, reveal latent patterns in case law, and produce predictive models, subsequently impacting the quality, speed, and consistency of legal decision-making. While all this is promising, the achievement of these outcomes hinges on developing and applying ML systems with proper ethical frameworks and regulatory mechanisms.

This paper seeks to fulfil two main objectives. The first objective is to demonstrate the increasing importance of data analytics in the current era of law by exploring how legal practitioners have adopted data in the practice of law, focusing on its importance for strategic planning, resource allocation, and compliance with law. Secondly, this paper will present a thorough history of machine learning as a technology that has the potential to be transformative in legal analytics, including an overview of the capabilities of ML and its limitations, while incorporating a critical review of the ethical, regulatory, and technical factors underpinning expressed capabilities. To this end, the paper examines the basic machine learning methods relevant to legal analytics and demonstrates their application in legal reasoning, predictive outcomes, and data analysis.

## **2. Foundations of Machine Learning in Law**

Before discussing applications of machine learning (ML) in legal analytics, it is important first to define some of the central principles supporting machine learning. At an essential level, machine learning is the development of algorithms that allow computer systems to learn from past data and to make predictions or decisions without programming a computer for every task (Mitchell, 1997; Bell, 2022). The ability of machine learning to “learn,” in which it adapts and improves its performance over time, is particularly applicable in static environments in which labels, precedents, and statutes compile into large corpora, like the typical data a legal analyst would work with.

There are numerous paradigms of learning in machine learning, but two, supervised and unsupervised learning, provide the primary foundation. For a supervised learning model, the algorithm is trained on a labeled dataset, which consists of data that pairs input features with known results or classifications (Hastie, Tibshirani, Friedman, 2009). Thus, supervised learning would be well suited for predicting legal case outcomes, assessing compliance risk, or using past annotated data for tasks like automated document classification (Katz, Bommarito, Blackman, 2014).

Conversely, unsupervised learning is performed using data that is not tagged. In this context, we may be attempting to identify hidden structures in the data or patterns (Gentleman & Carey, 2008). A few examples relevant to a legal context would be whole populated legal documents clustered, anomaly detection on record keeping and compliance, or identifying topic structures with large sets of case law or legislation (Chalkidis et al., 2022).

Semi-supervised learning integrates both types of learning. This means that a labeled dataset is small, but there is a larger volume of unlabeled data. This can be useful in legal contexts because it's possible to get a single amount of labeled datasets, as they can be confidential or complex, but generally, large amounts of legal text available remain unlabeled (Engelen & Hoos, 2020; Hady & Schwenker, 2013).

Another well-established paradigm is reinforcement learning. Reinforcement learning is a type of learning where an algorithm learns a desirable behaviour through trial and error, and feedback appears in the form of rewards or punishment (Sutton & Barto, 2018). While the application of reinforcement learning in the legal domain is new, it has potential for areas such as negotiation modeling or optimizing legal strategies during litigation.

### **2.1 Natural Language Processing in Legal Contexts**

Natural Language Processing (NLP) is an important subfield of ML related to legal analytics, and NLP examines how machines can understand and process human language (Kang, 2020). NLP is a Base Technology underlying many legal tech applications, including document summarization, named entity recognition from legal contracts, sentiment analysis of judicial opinions, and automated extraction of relevant clauses (Lein et al., 2022). As legal information is constituted mainly of textual representations, governmental or statistical documents, this makes NLP a necessary component for creating systems that can analyze statutory law, case law, and legal argumentation.

### **2.2 Data Preparation and Feature Engineering**

For a given ML system to function effectively, it is half manifested in the choice of algorithm used, but is just as critically influenced by the quality of the data provided. In the domain of actionable legal data, raw data is rarely very structured, structured text is noisy or heterogeneous, and requires careful pre-processing and validation. Feature engineering—the process of selecting and transforming variables of interest from raw data—also involves the selection of the information that improves model performance (Zebari, 2020). In legal analytics, this might include recognition of key phrases in case law, citation patterns from a case series, or tagging a semantic entity from regulatory documents.

### **2.3 Evaluation and Interpretability**

One of the most important steps in any ML deployment, especially in areas with sensitive consequences such as law, is model evaluation. Measurement of ML models typically involves standard metrics such as accuracy, precision, recall, and the F1 score (Raschka & Mirjalili, 2018). However, while a model's technical performance can be measured, it lacks deeper meaning in law. Model interpretability and accountability become the most important forms of evaluation in law, which means the ability to report back to the end-user how a model arrived at its decision and why it did so (Koh & Liang, 2017; Chen, 2023). For a model to uphold its legal validity and to support trust by institutions and confidence by the public, end-users must be able to ascertain how the model arrived at a specific decision in something like a risk assessment tool or outcome prediction model.

In the last years, the emerging field of explainable AI (XAI) processes (e.g., SHAP values, LIME, etc.) has developed methods to support 'black-box' models in being more transparent and interpretable to end-users, particularly in high-risk legal domains (Doshi-Velez & Kim, 2017). Without some way to provide interpretability, legal actors cannot understand, accept, or challenge the algorithmic decisions of models, presenting issues of procedural fairness and possibly impacting due process.

## **3. Applications of Machine Learning in Legal Analytics**

Machine learning (ML) has made great strides in the practice of law, building off existing technologies in terms of legal analytics in civil discovery and legal information retrieval (Delgado, 2019). ML has to offer a variety of tools to be able to pull useful insights, predict the outcomes of legal matters, or improve decision-making on any type of legal product (Ahmad et al., 2020). The most disruptive aspect of this is its ability to automate laborious processes and to help find patterns in large bases of legal information, both of which can lead to more efficient and more accurate (Atkinson et al., 2020).

Below, we highlight the main areas where ML is being used to improve legal analytics.

### **3.1 Predictive Analytics for Case Results**

Predictive analytics is one of the most important uses of ML in law. By using supervised learning models that are created with historical data from prior litigation, e.g., facts, legal arguments, decisions, and the behavior of judges in context, ML systems can predict the probability of different case law results (Medvedeva, Vols, & Wieling, 2020). They also provide important context when lawyers are evaluating risk and developing strategy, particularly in the context of

negotiations (e.g., settlement or whether to go to trial). An illustrative example is JPMorgan Chase's COIN (Contract Intelligence) system that reduced the time bar taking to review loophole agreements to mere seconds from 360,000 hours (Chase, 2017). Even though the system focuses on contracts, how it functions relates closely to data-driven components of models that predict an outcome. Aletras et al. (2016) pointed out how ML models predict the outcome of the European Court of Human Rights cases, reporting an overall accuracy of 79%.

### **3.2 Contract Analysis and Management**

Contract analysis is now another distinct core use case for ML, primarily using NLP. These ML tools automate the identification of contract clauses, points of risks, conflicts, and deviations from standardized template versions (Hassan et al., 2021). ML-enabled applications like Kira Systems, Evisort, and LawGeex are being used to quickly conduct contract reviews, augmenting compliance, shortening review time, and reducing human errors. According to a Deloitte (2018) case study, reviewing contracts with AI improved review efficiency by 20–40% in due diligence and transaction workflow in mergers and acquisitions.

### **3.3 Legal Research and Document Analysis**

ML can automate legal research. ML makes it possible for systems to extract relevant information from large sets of articles, cases, statutes, administrative regulations, and scholarly commentary (Ahmad et al., 2020). Such as Westlaw Edge and ROSS Intelligence are allowing for tools able to provide contextual case suggestions, obtain analogous premises from case law, and flag variations in doctrinal interpretations in different jurisdictions (Katz, Bommarito, & Blackman, 2017). While ROSS Intelligence is no longer operational due to litigation, the firm was successful in applying IBM Watson's natural language processing capabilities to provide search, rank, and document relevant natural language legal questions and corresponding answers, indicative of opportunities for future innovation.

### **3.4 Sentiment Analysis and Social Media Analysis**

ML applications, mostly in the areas of text classification and sentiment analysis, are becoming more prevalent in analyzing public sentiment about legal issues, prominent cases, or legislation (Mahajan & Mansotra, 2021). Legal teams could monitor online dialogue, juror sentiment, and coverage for relevant cases to manage reputational risks or to prepare for jury selection. The benefits of this type of peth is evidence-based in validating promising approaches to KPI and regulatory compliance based on industry information on issues in real-time.

### **3.5 Legal Chatbots and Virtual Assistants**

ML conversational agents can help address access-to-justice gaps by providing the public with basic legal information and guidance on legal processes (Socatiyanurak, 2021). Typically trained on jurisdiction statutes and their own FAQ, chatbots can assist users with drafting legal documents, navigating court proceedings, and basic legal rights information. Examples of applications in this area include *DoNotPay*, a chatbot designed to appeal on behalf of people parking tickets and providing other legal services, *Hello Divorce*, which walks the user through the divorce process in select U.S. states.

### **3.6 Fraud Detection and Compliance Monitoring**

Anomaly detection and unsupervised learning approaches are increasingly being adopted to detect fraudulent transactions, for example, in the context of financial law. Such approaches are being employed to detect fraud or embezzlement, or compliance violations (Huang, 2024). Machine learning models laden with historical fraud data are finding subtle anomalies, or transactions that deviate from the norm, within large sets of financial data across various contexts. Exploiting anomaly detection methods is being used in a variety of compliance platforms, serving the bank, insurance, procurement, and other regulatory spaces, and is becoming incorporated and advanced in systems with the intent of helping to reduce risks in real-time.

### **3.7 Decision Support Systems**

Decision support systems driven by machine learning can assist judges and solicitors by providing access to cited precedents along with the ability to assess legal reasoning and to present subsequent probabilities of future outcomes based on data provided on case specifics. These tools can assist with tracing consistent reasoning by judges and ensure the judicial reasoning process is visible for scrutiny through the transparency of the recommended algorithms (Završnik, 2020). Perhaps one of the most significant cases of judicial acceptance of decision support technology is in the High Court of England and Wales when allowing predictive coding in *Pyrrho Investments Ltd v MWB Property Ltd*, when a court provided for the use of technology-assisted review (TAR) in the course of electronic discovery as a means of reducing to expenditure and time if the matter were dealt with via the application of numerous lawyers.

While decision support systems are primarily advisory tools, the rate at which they are becoming more capable has led to discussions surrounding the legality of using such systems as machines, risks of using black-box systems, and bias in the algorithmic systems.

#### 4. Challenges and Ethical Considerations

There are meaningful ethical and operational issues arising from the implementation of machine learning (ML) in legal practice and analysis. The benefits of machine learning in legal practice and analysis include efficiencies, increased predictability, and improved decisions based on law and rights. There are also new variations of issues regarding bias, transparency, and rights that ML could be perpetuating. Researchers are starting to observe the complexities arising from the use of ML technologies in legal practice. Hamilton and Davison (2022) speculate that ML practices may inappropriately breach employment discrimination protections, privacy protections, and accounts when algorithms used in a legal context are dealing with data, such as professional and sensitive personal data.

Miller (2019) highlighted the ethical implications of ML, such as voter profiling and manipulation when used in the context of elections. Mohan and Dutta (2023) explain the issues with automated decision-making systems, such as facilitating opacity and preserving bias in legal reasoning. These studies suggest legal practitioners should start to reflect on the societal and normative issues related to the use of ML technologies.

Ethical considerations surrounding ML-enabled legal systems require an additional notion of a principled framework based on fairness, transparency, and accountability. The incorporation of ML within law should not violate the established legal principles of fairness, transparency, and accountability. Rather, the following challenges must be systematically addressed to ensure the responsible and fair use of ML in legal systems.

##### 4.1 Bias and Fairness

Bias in ML models is a core ethical issue in legal applications. Algorithms trained on historical legal data may essentially (re)produce existing societal and institutional biases. Milaninia (2020) presents an example of how biased training data can generate disproportionate outcomes based on bias, particularly in the areas of sentencing or predictive policing. Traditional evaluation metrics can also mask underlying disparities that may mislead practitioners about the fairness of an outcome, particularly in legal systems. Although there are strategies to mitigate bias—data pre-processing to remove sensitive features, fairness constraints at the level of the algorithm, and additive models for post-hoc auditing relying only on observable features—it does not imply that it produces fair outcomes.

In the legal context, such as a judge's ruling or agency decision, the possibility of hidden bias (and its effects) cannot be overlooked because the consequences of individual or favorite contingency decisions can be significant on the lives and freedoms of individuals. The illegitimacy of ML-supported processes of legal

response and decision-making, as well as individually unfair outcomes, may stem from bias.

#### **4.2 Interpretability and Transparency**

A further challenge to deploying ML in legal analysis is the comprehension of the interpretability of the algorithmic decisions. For many ML models (particularly deep learning models), the use of various neural networks obscures the interpretability of decisions and can make it seem like the models are “black boxes.” Carvalho et al. (2019) emphasize that this opacity makes it impossible for legal professionals to know how or why an automated decision was reached.

Given that justifying decisions is fundamental to the legal profession, such opacity heightens the relevance of this insufficiency. If stakeholders (judges, lawyers, and litigants) cannot see, critique, or contest how an algorithm produced its outputs, the likelihood of injustice increases. Various Techniques, such as explainable AI (XAI), model distillation, and rule-based approximations, are being developed to address issues of opacity, but they are limited to law’s use and require more institutional support before performance will improve.

#### **4.3 Data Privacy and Security**

Legal data, such as case files, client communications, and proprietary documents, is sensitive data in nature. Ideally, ML systems would utilize large sets of historical data to make predictions or suggestions about what legal action to take. As such, any statutory protection offered by data anonymization procedures deteriorates with the size of the dataset and the potential risk of a breach of confidentiality. Gellert (2022) notes that as the underlying datasets of a machine learning system becomes larger, the losing party risks exposing potentially irreparably sensitive information with the ML systems Third Party disclosures, specifically when the datasets or any of its component variables are not properly anonymized or, when accessing the data and consequently making the data publicly available, authorized or unauthorized.

Shope (2021) offers a similar warning, stating that even legal datasets that are properly sanctioned and technically anonymous, and have been anonymized, could always be reverse-engineered and the sensitive data re-exposed. It is imperative for lawyers to fully comprehend the importance of using the highest threshold of privacy-preserving techniques to protect HHPC, such as differential privacy, encryption (encryption), and access controls to ensure that they are prescribing to adequately protecting client information, and to respect the statutory legal obligations of the legal profession, including data protection regulations such as the GDPR and HIPAA.

#### **4.4 Model Robustness and Reliability**

The legal system is inherently complex, variable, and ever-changing, which raises serious concerns about the viability and reliability of ML models. An ML algorithm that is trained on a narrow data set could provide unacceptable outcomes for a different jurisdiction, different developing legal rules, or even a completely new type of case. As a result, whenever ML systems are required to provide reliable support in a particular legal context, the situation requires thorough validation and testing.

This usually includes a thorough validation plan that is developed and implemented by both data scientists and legal experts to evaluate if the model outputs are aligned with their doctrinal principles and procedural fairness (Zarsky, 2016). If the project team does not contain legal experts working alongside the data scientists, at some point, there will be a considerable likelihood that the ML tools, although they may produce technically valid outputs, may also lead to legally nonsensical implications.

#### **4.4 Ethical Use of Predictive Analytics**

While predictive analytics may be valuable in terms of resource management or risk assessment, such as in terms of algorithms to forecast recidivism, or predicted outcomes of legal cases, or breaches of contract, there may be ethical challenges regarding autonomy, consent, and the presumption of innocence. Golbin (2020) poses the question that when you assess a future legal decision based on hypothesized probabilities, there is a risk that you can take legal action based on inferred trends in behaviour, which invokes a failure of due process and an individual's rights.

Legal professionals must be conscious of these risks and ensure that any predictive system, regardless of merit, is present in an assistive capacity rather than a substitutive capacity for human judgment. To ensure that predictive or similar analytics optimization represents a realistic and ethical consideration and is comparable to the legal rules from case law instead of trying to displace them, it should be accompanied by well-defined standards, a supervisory system tracing each action taken in connection with the predictive system, and meaningful community engagement.

#### **4.5 Regulatory Compliance**

As legal analytics tools develop based on ML and intersect multiple areas of law, there are regulatory frameworks applicable to specific areas of law. Webley (2019) articulates that lawyers must ensure that any technology they work with complies with legislation around privacy, intellectual property, discrimination, and professional rules and conduct. For example, ML systems operating as a form of

legal research must comply with existing copyright laws governing the use of case databases, while systems used for client analytics must be compliant with attorney–client privilege and confidentiality rules. Regulating compliance matters because non-compliance exposes legal practitioners to sanctions but also undermines trust in the potential to productively use technological solutions in legal work. Thus, compliance should be viewed as an ongoing process embedded in the ML system development cycle.

#### **4.6 Human–AI Collaboration**

Machine Learning should not substitute for human legal intelligibility but should have a complementary position. Effective implementation of Machine Learning into legal workflows requires strong frameworks for human–AI collaboration. Noll et al. (2021) indicate that hybrid systems with human and algorithmic agents require clearly described roles, responsibilities, and decision-making authority. Hybrid systems must describe elements such as the situations that call for human intervention, the process for circumventing questionable outcomes, and the documentation to maintain accountability chains. Legal institutions should formulate best practices and offer training around human–AI collaboration to reinforce the precedence of human expertise in legal decision making.

### **5. Future Directions and Opportunities**

The legal field stands to change substantially given the rapid advancements in machine learning. New developments in machine learning research offer considerable opportunities for innovation in legal analytics and create novel fields of practice that may enhance legal reasoning, decision support, and access to justice. Achieving the maximum potential of machine learning in law will require a cross-disciplinary approach and research focus on specific subfields where machine learning can design solutions to persistent legal problems. The ensuing list proposes some possibilities for potential avenues of development.

#### **5.1 Advanced Predictive Modeling**

Predictive analytics has already started to change the way legal professionals look at case outcomes, litigation risk, and judge behavior. Hopefully, the next step will be through even better methodologies such as deep neural networks, transfer learning, and reinforcement learning (Surden, 2019). Using these methodologies bestows a high level of accuracy by allowing the construction of deep contextual models that predict more than just outcomes of legal decisions. For example, deep learning models could be trained using multimodal legal data (text, metadata, and prior precedents) to potentially predict how the judge came to their reasoning

or how a contract would be executed. Ultimately, if implemented in a justified manner, these predictive analytics could help increase strategic planning, resource allocation, and dispute resolution in all legal systems.

### **5.2 Explainable AI and Interpretability**

Due to the emphasis on justification in legal decision-making, it is necessary for legal analytics tools to attain an elevated level of explainability. In more general terms, a situation may arise where syntactic understanding alone cannot support substantive debate or judicial review over the decisions made by the algorithm. Therefore, future ML legal research will need to prioritize explainable artificial intelligence (XAI), which includes efforts towards developing technology that provides explanations to individuals about how complex models reach decisions (Carvalho et al., 2019). There are many ways to understand, if not explain, an algorithm, including using feature attribution, extracting rules, and developing counterfactual explanations so that, upon output, we can identify why an output was produced and determine proper accountability. We must embed interpretability into the development processes of legal ML systems to engage with due process and the rule of law.

### **5.3 Fairness-aware Machine Learning**

Fairness is a fundamental principle of the legal process and must be included in algorithmic procedures. Future research must address how to quantify, evaluate, and mitigate algorithmic bias in or on legal data. Techniques for building fairness-aware ML systems, such as fairness constraints, adversarial debiasing, and disparate impact testing and analysis, can be used (Mehrabi et al., 2021). Additionally, legal scholars and technologists will need to work collaboratively to define context-specific fairness metrics, such as equal opportunity or calibration, that match existing legal definitions of fairness and equity for context. Given that the algorithmic architecture shapes the nature and extent of deliberation in the project, fairness should be built into the algorithmic architecture to ensure that machine learning systems work to support rather than undermine the ethical foundations of law.

### **5.4 Privacy-Preserving Machine Learning**

With legal documents and client information, there's a high importance on protecting privacy. There are developments in privacy-preserving machine learning, and early gains from federated learning, homomorphic encryption, and differential privacy will help engage in safe and compliant analysis of federated legal datasets (Shokri & Shmatikov, 2015). Privacy-preserving methods allow ML models to learn from data without being exposed, keeping the potential for privacy

violations and data protection regime ramifications low. While privacy governance is becoming more stringent and enforceable, it is envisaged that privacy-preserving ML will be crucial to stimulate innovation and development, while also preserving individual rights.

### **5.5 Natural Language Understanding and Legal NLP**

Legal documents contain dense, formalized language and other characteristics that offer challenges to natural language processing (NLP). However, as the NLP community transitions to using transformer architectures (e.g., BERT, GPT), new opportunities for NLP in law and other domains may emerge as a result of fine-tuning the domain-specific models for tasks such as automated contract review, legal document summarization, precedent extraction, and statutory interpretation (Chalkidis et al., 2020). These developments may potentially reduce cognitive load for legal practitioners and increase the efficiency and accessibility of legal services across marginalized jurisdictions.

### **5.6 Ethical and Regulatory Frameworks**

As AI systems are more entrenched in legal work processes, there is an increasing need for meaningful ethical and regulatory governance frameworks. These frameworks are necessary to provide guidance for developers, practitioners, and regulators that is much more robust than vague principles. Some of the inherent components of these frameworks include protocols for algorithmic auditing, impact assessments, and stakeholder oversight processes (Morley et al., 2021). In addition, frameworks need to allow for the rapid evolution of all emerging technologies to allow fungibility of the law and the continued assessment and iteration of existing law and policy. Developing such standards will require ongoing engagement from interdisciplinary stakeholders and civil society organizations, as well as courts, bar associations, and other actors.

## **Conclusion**

Integrating machine learning (ML) with legal analytics creates a significant opportunity and a complicated challenge. There is potential, for example, that ML tools can aid legal professionals by analyzing significant amounts of legal data, predicting future behaviour, and informing strategic decisions. If this sounds like using more credible evidence and a more effective way to conduct legal practice, this is a development in legal practice. However, the promise of ML will only be realized in ways that we find meaningful if we apply and embed our primary legal values of fairness, transparency, and accountability into the integration.

ML is also promising for helping to improve access to justice, especially for marginalized communities. From automating simple, routine legal tasks to making legal information more easily available and digestible, this technology could help to mitigate the issues faced by many individuals who struggle to secure representation and to democratize access to legal knowledge (Remus & Levy, 2016; Susskind, 2019). Nevertheless, the viability of ML disrupting this space depends heavily on embedding equity and inclusion in the very design and implementation of such systems.

The risks of algorithmic opacity, bias, and misuse of data are particularly significant in legal contexts due to the principles of procedural fairness and due process. Ethical usage of algorithmic legal reasoning requires more than technical solutions - normative approaches that draw on legal ethics and human rights (Zarsky, 2016) are required. ML systems must be transparent, auditable, and accountable to ensure that they build, rather than erode, public trust in the institutions of law.

Interdisciplinary engagement is vital to realize ethical uses of algorithmic systems in legal contexts. Legal scholars, technologists, ethicists, and regulators must develop robust governance frameworks, enforceable standards, and timely oversight mechanisms together (Morley et al., 2021). Ultimately, machine learning could be more than an inherently self-serving pursuit of efficiency, but rather a commitment to justice, the integrity of law, and the public interest. Only then can ML make meaningful contributions to the aims of law reform in an increasingly digital society.

### **Disclosure statement**

No potential conflict of interest was reported by the author(s).

### **Funding Information**

No funding was received for this research. The study was conducted independently, with institutional support from the Taipei Medical University, Shoolini University, and the Government of India.

### **Ethical and Originality Statement**

The Author(s) declare that this work is original and has never been published in any form or any other media, nor is it under consideration for publication in any journal, and all sources cited in this work refer to the basic standards of scientific citation.

## References

- Ahmad, M., Yousaf, M., & Khalid, A. (2020). Machine learning in legal text analytics: A review. *IEEE Access*, 8, 213782–213798.
- Aletras, N., Tsarapatsanis, D., Preoțiu-Pietro, D., & Lampos, V. (2016). Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. *PeerJ Computer Science*, 2, e93. <https://doi.org/10.7717/peerj-cs.93>
- Atkinson, K., Bench-Capon, T., & Bollegala, D. (2020). Explanation in AI and law: Past, present, and future. *Artificial Intelligence*, 289, 103387. <https://doi.org/10.1016/j.artint.2020.103387>
- Bell, J. (2022). What is machine learning? In *Machine learning and the city* (pp. 207–216). Wiley. <https://doi.org/10.1002/9781119815075.ch18>
- Brooks, C., Gherhes, C., & Vorley, T. (2020). Artificial intelligence in the legal sector: Pressures and challenges of transformation. *Cambridge Journal of Regions, Economy and Society*, 13(1), 135–152. <https://doi.org/10.1093/cjres/rsz026>
- Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics*, 8(8), 832. <https://doi.org/10.3390/electronics8080832>
- Caserta, S. (2020). Artificial intelligence in the judiciary: A comparative overview. *AI & Society*, 35(3), 453–465. <https://doi.org/10.1007/s00146-020-00999-7>
- Chalkidis, I., Fergadiotis, M., Malakasiotis, P., & Aletras, N. (2022). LexGLUE: A benchmark dataset for legal language understanding in English. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- Chalkidis, I., Fergadiotis, M., Malakasiotis, P., & Androutsopoulos, I. (2020). LEGAL-BERT: The muppets straight out of law school. *arXiv preprint arXiv:2010.02559*. <https://arxiv.org/abs/2010.02559>
- Chase, J.P. Morgan. (2017). *JPMorgan Chase unveils Contract Intelligence Platform (COIN)* [Press release].
- Delgado, F. A. (2019). Machine learning in legal practice: Notes from recent history. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 557–558. <https://doi.org/10.1145/3306618.3314324>
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*. <https://doi.org/10.48550/arXiv.1702.08608>
- Engelen, F., & Hoos, H. (2020). Semi-supervised learning in legal analytics: Potentials and pitfalls. *Artificial Intelligence and Law*, 28(3), 271–294. <https://doi.org/10.1007/s10506-020-09265-1>

- Gellert, R. (2022). Understanding the notion of risk in the General Data Protection Regulation. *Computer Law & Security Review*, 38, 105433. <https://doi.org/10.1016/j.clsr.2020.105433>
- Gentleman, R., & Carey, V. J. (2008). Unsupervised machine learning. In *Bioconductor Case Studies* (pp. 137–157). Springer. [https://doi.org/10.1007/978-0-387-77240-0\\_10](https://doi.org/10.1007/978-0-387-77240-0_10)
- Golbin, I. (2020). Responsible AI: A primer for the legal community. *Proceedings of the 2020 IEEE International Conference on Big Data*, 2121–2126. <https://doi.org/10.1109/BIGDATA50022.2020.9377738>
- Greenleaf, G., Mowbray, A., & Chung, P. (2018). Building sustainable free legal advisory systems: Experiences from the history of AI & law. *Computer Law & Security Review*, 34(2), 314–326. <https://doi.org/10.1016/j.clsr.2018.02.007>
- Hady, M. F. A., & Schwenker, F. (2013). Semi-supervised learning. In *Intelligent Systems Reference Library* (Vol. 215, pp. 215–239). Springer. [https://doi.org/10.1007/978-3-642-36657-4\\_7](https://doi.org/10.1007/978-3-642-36657-4_7)
- Hamilton, M., & Davison, R. (2022). Legal informatics and algorithmic accountability in the workplace. *Yale Journal of Law & Technology*, 24(1), 1–33. <https://yjolt.org>
- Hassan, S., Malik, M. N., & Khan, M. U. G. (2021). Automating contract management using natural language processing and machine learning. *Journal of Information Science*, 47(3), 369–384. <https://doi.org/10.1177/0165551520985495>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- Kang, T. (2020). Natural language processing in the legal domain: A survey. *AI and Ethics*, 1(2), 189–202. <https://doi.org/10.1007/s43681-020-00020-5>
- Katz, D. M., Bommarito, M. J., & Blackman, J. (2014). Predicting the behavior of the Supreme Court of the United States: A general approach. *PLOS ONE*, 9(4), e92413. <https://doi.org/10.1371/journal.pone.0092413>
- Koh, P. W., & Liang, P. (2017). Understanding black-box predictions via influence functions. *Proceedings of the 34th International Conference on Machine Learning (ICML)*.
- Mahajan, S., & Mansotra, V. (2021). Sentiment analysis in legal domain: A survey and case applications. *International Journal of Advanced Computer Science and Applications*, 12(2), 342–348. <https://doi.org/10.14569/IJACSA.2021.0120245>
- Markou, C., & Deakin, S. F. (2019). *Ex Machina Lex: The limits of legal computability* [Working paper]. SSRN. <https://doi.org/10.2139/ssrn.3407856>
- Medvedeva, M., Vols, M., & Wieling, M. (2020). Using machine learning to predict decisions of the European Court of Human Rights. *Artificial Intelligence and Law*, 28(2), 237–266. <https://doi.org/10.1007/s10506-019-09255-y>

- Milaninia, N. (2020). Biases in machine learning models and big data analytics: The international criminal and humanitarian law implications. *International Review of the Red Cross*, 102(913), 199–234. <https://doi.org/10.1017/S1816383121000096>
- Miller, J. (2019). Algorithmic manipulation and electoral ethics: Regulating voter profiling. *Ethics and Information Technology*, 21(3), 183–195.
- Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.
- Mohan, A., & Dutta, K. (2023). Artificial intelligence in legal decision-making: Ethical concerns and frameworks. *Journal of Law and AI Ethics*, 2(1), 45–67.
- Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2021). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, 27(1), 1–31.
- Mowbray, A., Chung, P., & Greenleaf, G. (2020). Utilising AI in the legal assistance sector—Testing a role for legal information institutes. *Computer Law & Security Review*, 38, 105407. <https://doi.org/10.1016/j.clsr.2020.105407>
- Remus, D., & Levy, F. S. (2016). Can robots be lawyers? Computers, lawyers, and the practice of law. SSRN. <https://doi.org/10.2139/ssrn.2701092>
- Schmidt, P., Biessmann, F., & Teubner, T. (2020). Transparency and trust in artificial intelligence systems. *Journal of Decision Systems*, 29(3), 260–278. <https://doi.org/10.1080/12460125.2020.1819094>
- Shope, M. L. (2021). Lawyer and judicial competency in the era of artificial intelligence: Ethical requirements for documenting datasets and machine learning models. *Georgetown Journal of Legal Ethics*, 34, 196–220.
- Socatiyanurak, V. (2021). LAW-U: Legal guidance through artificial intelligence chatbot for sexual violence victims and survivors. *IEEE Access*, 9, 131440–131461. <https://doi.org/10.1109/ACCESS.2021.3113172>
- Soukupová, J. (2021). AI-based legal technology: A critical assessment of the current use of artificial intelligence in legal practice. *Masaryk University Journal of Law and Technology*, 15(2), 279–300.
- Susskind, R. (2019). *Online courts and the future of justice*. Oxford University Press.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.
- Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Transparent, explainable, and
- William, P. (2023). Crime analysis using computer vision approach with machine learning. *Lecture Notes in Networks and Systems*, 297, 315–332. [https://doi.org/10.1007/978-981-19-7982-8\\_25](https://doi.org/10.1007/978-981-19-7982-8_25)
- Zarsky, T. Z. (2016). The trouble with algorithmic decisions: An analytic roadmap to examine efficiency and fairness in automated and opaque decision-making.

---

*Science, Technology, & Human Values*, 41(1), 118–132.  
<https://doi.org/10.1177/0162243915605575>

Završnik, A. (2020). Algorithmic justice: Algorithms and big data in criminal justice settings. *European Journal of Criminology*, 17(4), 449–467.  
<https://doi.org/10.1177/1477370819876762>