The Potential for Jurisdictional Challenges to AI or LLM Training Datasets

Chris Draper [1] and Nicky Gillibrand [2]

[1] *Indiana University, 107 S Indiana Ave, Bloomington, IN 47405, USA*
[2] *University College Dublin, Belfield, Dublin 4, D04 V1W8, Ireland*

Abstract
Large language model (LLM) tools used in AI powered access to justice (A2J) systems experience systemic bias when their training datasets do not reflect their communities. Such bias arguably indicates that the LLM should see the validity of its legal underpinnings challenged on jurisdictional grounds. Since ChatGPT has the capacity to pass an American Bar Exam, this provides hope that LLM tools can be trained to perform the work of a legal professional at the direction of a lay person, to the perceived benefit of the underserved litigant. However, significant challenges arise when reviewing the source of the datasets in terms of adherence to legal sovereignty, rule of law and quality of outcome. While privacy and data security will often focus data sovereignty on the geographic location where the data is held, the A2J community should also be mindful of extra-jurisdictional contributions to LLM training datasets that dispute the generally accepted norm of legal sovereignty, and as a result skew its application of law to be outside the acceptable boundaries of the impacted community. To better represent the challenges posed by LLM tools a novel quadripartite theory of informational sovereignty is offered, encompassing concerns regarding population, territory, recognition and regulation of borders.

This paper will therefore examine and call into question claims that LLM is a perceived enabler of A2J. Discussion will involve how avoidance of jurisdictional challenges, such as traditional legal sovereignty, through a myopic focus on data sovereignty circumvents the risks of training data skewedness often displayed in bias, before considering how jurisdictionally defined training data limitations could impact outcome quality and the reformulation of the traditional role of the lawyer in the legal process. Finally, we will explore the dangers of failing to sufficiently address these far-reaching challenges – impacting all levels from the community to constitutional - in light of contemporary concerns and litigation.

Keywords
Validation of the Legal Underpinnings of Systems; LLM; Large Language Models; Sovereignty; Rule of Law; Jurisdiction; Bias; AI Risk; Pragmatics of Adoption; Self-Represented Litigants; Panel Discussions; Guided Discussions; Works in Progress

1. Introduction

Large language model (LLM) tools used in AI powered access to justice (A2J) systems experience systemic bias when their training datasets do not reflect their communities. Such bias arguably indicates that the LLM should see the validity of its legal underpinnings challenged on jurisdictional grounds. Since ChatGPT has the capacity to pass an American Bar Exam, this provides hope that LLM tools can be trained to perform the work of a legal professional at the direction of a lay person, to the perceived benefit of the underserved litigant. However, significant challenges arise when reviewing the source of the datasets in terms of adherence to legal sovereignty, rule of law and quality of outcome. While privacy and data security will often focus data sovereignty on the geographic location where the data is held, the A2J community should also be mindful of extra-jurisdictional contributions to LLM training datasets that dispute the generally accepted norm of legal sovereignty, and as a result skew its application of law to be outside the acceptable boundaries of the impacted community. To better represent the challenges posed by LLM tools a novel quadripartite theory of informational sovereignty is offered, encompassing concerns regarding population, territory, recognition and regulation of borders.

This paper will therefore examine and call into question claims that LLM is a perceived enabler of A2J. Discussion will involve how avoidance of jurisdictional challenges, such as traditional legal sovereignty, through a myopic focus on data sovereignty circumvents the risks of training data skewedness often displayed in bias, before considering how jurisdictionally defined training data limitations could impact outcome quality and the reformulation of the traditional role of the lawyer in the legal process. Finally, we will explore the dangers of failing to sufficiently address these far-reaching challenges – impacting all levels from the community to constitutional - in light of contemporary concerns and litigation.

2. The Current State of Legal AI

Due in no small part to the rising accessibility and the proliferation of use of AI, considerable literature on the topic continues to emerge at a rapid pace. AI itself is becoming increasingly newsworthy, particularly in the wake of ChatGPT's rise to prominence and its related controversies such as its ban in Italy,[2] amongst other notable headlines such as its ability to pass the Uniform Bar Examination in the US.[3] Whilst much of the existing literature on the role of AI in the law to this point stems from a place of hope that it may eventually have a positive impact on A2J, enabling those who cannot afford a legal professional to use accessible technology that can technically attain the level of a trained professional,[4] with some going as far as to state that AI is a prerequisite for social justice.[5] A significant volume of work also puts forward that we should remain cautious of the sudden rise of AI usage, with it holding the potential to exacerbate structural inequities inherent in society.[6]

Failure to regulate the use of AI in the legal profession remains a significant problem, with jurisdictions focusing primarily on the regulation of AI in case of autonomous vehicles and for the use of national defence.[7] The value of government regulation cannot be understated as the rolling out of an AI tool as a means to facilitate A2J can contribute to sociopolitical disparities where those who can only afford AI may be receiving low quality legal services compared to those who have the funds to engage legal professionals. Furthermore, AI broadly defined cannot constitute an appropriate answer to enhance A2J as the newest LegalTech will remain cost prohibitive to underserved members of the public, whilst high street lawyers representing less wealthy members of society will also be squeezed by LegalTech,[8] therefore a significant gulf will remain between profit and not-for-profit AI systems.[9]

As AI datasets, if poorly constructed, are capable of providing incorrect information and being subject to considerable bias,[10] infringing the rights of individuals and groups with certain characteristics.[11] If used in sentencing, such bias can ultimately result in a deprivation of one's liberty based on these characteristics.[12] As such, warnings have arisen that AI datasets must not only be bigger, but also of better quality, which is generally described as the dataset being unbiased and less expensive whilst most importantly remaining legally compliant,[13] in turn assisting the cultivation of more predictable outcomes.[14] Therefore quality of datasets is paramount to AI fulfilling any sort of function and cultivating public trust as an alternative to traditional services.[15] Perhaps most prohibitively of all, those

who are unable to use computers or are without the necessary technology cannot make use of AI tools regardless, furthering social inequalities.[16]

Whilst the bulk of the literature focuses on how a failure to properly regulate AI can impact the public at an individual level, there is considerably less on the wider impact to the state's jurisdiction and constitutional architecture. Of these, it is said to be pivotally important for the societies to have control over the source code of the AI datasets before it is ceded to private tech corporations who may ultimately regulate AI and subsequently impact the rule of law.[17] The rule of law is said to be challenged in three ways by AI: the aforementioned blurring of the private-public regulatory sphere on fundamental rights; the subsequent failure to demarcate legal certainty within this framework; the lack of transparency and accountability of the mechanisms of decision-making.[18] By challenging the rule of law, one challenges potentially centuries of constitutional tradition that forms the basis of civilised society. As such, the implications may be widespread, with theorists stating that there requires a substantive reconfiguration of the relationship between law, technology and legal culture in order to incorporate algorithmic rationality.[19] If, therefore, LLMs gain a significant role in the legal profession and fail to be representative of legal culture, synonymous to some with the rule of law,[20] this can result in declining public sentiment towards the legal system more generally which is insurmountably detrimental to the wider functioning of the state.

These discourses are also significantly related to our concerns regarding the impact of LLMs and their datasets on jurisdictional sovereignty which remain largely unaddressed. It is, therefore, of utmost importance to exercise caution when considering the role of LLM tools in the law and consider any substantive advancements for its capacity through the lens of sovereignty discourses, both of the traditional and digital variety, in order to fortify the probability of representative outcomes for communities.
.

3. Framing Access To Justice

How to deliver access to justice (A2J) within society is broadly debated. Among laypersons this debate typically revolves around philosophical definitions of justice. Yet among the legal community the debate typically revolves around

the definition of access. The National Center for Access to Justice defines A2J as "when people encounter life challenges they are able to understand their rights under the law, protect those rights, obtain a fair result, and enforce that result to fully realize its value."[21] This definition frames justice as accessible through sufficient understanding and fair application of the law, with organizations like the United States Department of Justice seeing its role as helping "the justice system efficiently deliver outcomes that are fair and accessible to all, irrespective of wealth and status"[22] or the American Bar Association seeing A2J as "access to pro bono and low-cost legal services for vulnerable persons."[23] These views of justice as being attainable through greater access to the legal system have resulted in many A2J efforts focusing on the following solutions, inter alia:

- Open data initiatives - Governments and legal organizations are increasingly embracing open data initiatives, making legal information more freely available to the public. By providing access to legislation, case law, and other legal resources, these initiatives enable individuals to better understand their legal rights and obligations.

- Legal Aid Apps, Chatbots, and Self-Help Portals - Various mobile applications and chatbots have been developed to provide legal assistance and guidance to individuals who cannot afford or access traditional legal services. These tools offer information about legal rights, procedures, and resources, helping people navigate legal issues more effectively, including interactive guides, video tutorials, and legal document templates. These resources empower individuals to handle legal matters on their own, reducing the need for costly legal representation.

- Non-lawyer representation - Some legal sandbox initiatives in the United States are allowing non-lawyers to provide legal guidance on various topics.

- Pro Bono Resource Matching -Online platforms have emerged that connect individuals in need of legal assistance with volunteer lawyers willing to provide pro bono services. These platforms use

technology to match individuals with appropriate legal professionals, expanding access to free legal help.

- Remote Court Access -The adoption of remote court proceedings has accelerated in recent years, especially during the COVID-19 pandemic. Virtual courtrooms and video conferencing technologies have allowed individuals to participate in legal proceedings without the need for physical presence, saving time and reducing logistical barriers.

- Alternative and Online Dispute Resolution - Face-to-face mediation and arbitration have long been viewed as options for reducing court backlogs, with online dispute resolution (ODR) rising to prominence in the justice system following its rapid growth as a solution for resolving eCommerce disputes outside of the traditional justice system.

Each of these solutions have the potential to expand access by making legal information and resources more accessible, walking laypeople through the steps that must be taken, reducing the time it takes to find meaningful support or representation, or decreasing the time and cost for a case to be heard. In theory, the more these tools can operate without the oversight or intervention of human experts, the further barriers to access will drop.

This is where much promise is seen in AI. As examples, open data initiatives mean AI datasets could become more complete. AI chatbots could understand a layperson's issues, select the most appropriate process, any relevant forms needed, and even fill out or file those forms on their behalf. The productivity of non-lawyer and pro bono experts could leverage AI-supported intake interviews, document drafting, or meeting scheduling. Remote hearings, ADR, or ODR could be facilitated by digital clerks or neutrals. Yet all this promise is contingent on the ability to appropriately understand and act upon often murky human intention.

4. AI As A Tool for Procedural Justice

AI systems powered by LLM tools are seen as potentially transformative when framed through a procedural view of justice. Procedural justice refers to the fairness and impartiality of the processes and procedures used to resolve disputes, allocate resources, or make decisions. It emphasizes the importance of ensuring that the procedures used to make decisions are perceived as fair and just by those affected by them, regardless of the outcome.

The concept of procedural justice is rooted in the belief that people have a fundamental need to be treated fairly and with respect, and that the procedures used to make decisions can have a significant impact on how they perceive the fairness of those decisions. If disputes are resolved through a process that the community agrees is "fair," then the outcome of that process should be "just."

The concept of what constitutes "fairness" with respect to the processes that make up the justice system grew out of communities' norms and values. Historically, communities established their own rules and systems for resolving disputes and administering justice arising from their distinct legal culture. These systems were based on the norms, values, and customs of the community and were designed to reflect the unique needs and characteristics of that community.

For example, in many Indigenous communities, the concept of restorative justice was and still is an important part of their justice system.[24] In this system, the focus is on healing relationships and restoring balance, rather than on punishment or retribution. This approach is grounded in the values of community, respect, and harmony.

Similarly, in many small communities, disputes were often resolved through mediation or negotiation rather than through formal legal proceedings. These informal methods of dispute resolution were based on a sense of community and mutual respect, and often involved the participation of respected community members or elders.[25]

As communities grew and became more complex, the need for more formal systems of governance and justice arose. However, the underlying values and principles of fairness and equity remained an important part of these systems. The legal system that evolved from these community-based systems is built upon the principles of due process, impartiality, and the rule of law, as circumscribed by jurisdictional boundaries.

5. Justice Through The Rule of Law

The role of the rule of law within legal systems cannot be understated. The rule of law cemented its place as a foundational principle of constitutional law centuries prior, continuing to predominate until the present day. The rule of law acts as a safeguard against arbitrary power and a maintainer of public order.[26] Also within this, it acts as a bedrock for the formation of laws as the principal consideration on lawfulness on public legal action. In order to protect the rule of the law, a practical restriction exists in terms of each state having responsibility to maintain the quality of the rule of law. Responsibility for this substantially befalls the legal system and to a degree, the system of government. Both of these are impacted by public values to some extent, the law must adhere to the concerns of public policy and legal culture whilst the careers of many of those in the governmental sphere rests firmly upon public opinion.

The rule of law is said to be challenged in three ways by AI: the blurring of the private-public regulatory sphere on fundamental rights; the subsequent failure to demarcate legal certainty within this framework; the lack of transparency and accountability of the mechanisms of decision-making.[27] All of the above add a layer of obfuscation to a system that is already subject to unintelligibility at the level of a layperson. The result of this would be a more significant gap between the public and those in the legal profession thus causing a disengagement and a subsequent decline in legal culture.

Within the discussion of jurisdictions, a heavier usage of AI LLMs in their current form would result in an incremental decrease in representative legal outcomes. The absence of clear direction would subsequently culminate in a decline in legal culture being the primary source of law as it has previously been in common law systems. To uproot a primary source of law particularly through the backdoor, perhaps the one source that the public are undeniably aware of, is incredibly problematic from a democratic perspective. The legal system does not exist in a vacuum thus it is incontrovertible that an attempt to remedy the A2J crisis should not contravene democracy and the foundations of a community.

6. The Role of Jurisdiction

Jurisdictional boundaries are geographic or legal limits that define the authority of courts and other legal institutions to hear and decide cases. They represent an important component of the justice system, as they help to ensure fairness and impartiality by preventing conflicts of interest and promoting consistency and predictability in legal outcomes.

One way that jurisdictional boundaries support fairness in the justice system is by ensuring that cases are heard in a neutral and impartial venue. By establishing clear rules for which court or jurisdiction has authority over a particular case, jurisdictional boundaries help to prevent conflicts of interest and ensure that cases are heard in a forum that is independent and unbiased. Despite this, jurisdictional contestation is commonplace within private international legal cases where foreign laws may contravene the public policy interests of the lex fori thus transgressing the interests of the community in question.[28] The additional layer of complication formed by AI that exists outside of jurisdictional boundaries can be reasonably expected to add further complexity to the legal system by blurring the jurisdictional lines between legal precedents.

Appropriate jurisdictional boundaries that protect fairness as interpreted by the communities within those boundaries promote consistency and predictability in legal outcomes. This is achieved by establishing clear rules originating from community norms for which jurisdiction or court has authority over a particular type of case, legal institutions can ensure that cases are decided in a manner that is consistent with established legal principles and precedents in line with the principle of parity.

The nature of precedents themselves can create significant challenges within a jurisdiction and for AI machine learning, particularly when jurisdictional contestation is already a considerable problem. Whilst it is significant to ensure that an AI only applies the dataset applicable to the community in question in the application of law, it is often the case that a state may make reference to another jurisdictions legal precedent. For instance, the common law legal system of Ireland often makes reference to the precedents of other common law jurisdictions such as the legal system of England and Wales to assist in determining appropriate outcomes.

Rather than binding precedent, this is merely persuasive precedent. As such, teaching LLMs to differentiate between the use of other jurisdiction's law as persuasive precedent rather than the basis of another community's law which would largely be unrepresentative of that community's sentiment will pose a significant challenge to the effective use of AI in law, requiring considerably more nuance than LLM's provide in their current form.

Yet these precedents, and even sometimes the principles underpinning those precedents, are not permanent. These changes in precedent or principles are driven by the fact that community input and court decisions are intertwined. As court decisions can be influenced by community input, most often provided by lawyers or other legal practitioners, community input is also shaped by court decisions. When a court makes a decision in a particular case, based on how the community it serves argues the law before it, the decision sets a precedent for future cases that involve similar legal issues. Precedent is important because it ensures that the law is applied consistently over time, and it allows individuals and organizations to rely on the law and predict legal outcomes.

As society and values change over time, legal principles and precedents must also change. New societal norms must be reflected in new court decisions that establish new legal interpretations in order for the community to continue interpreting the justice system as just.

7. Lawyers as the Voice of the Community

Lawyers play a critical role in shaping legal principles and interpretations through their advocacy on behalf of clients. This role is so vital that nearly every jurisdiction enforces significant penalties when individuals, or computers in some jurisdictions, are seen to be engaged in the unauthorized practice of law.[29] In court cases, lawyers argue for a particular interpretation of the law that they believe best serves their client's interests. This interpretation can influence the court's decision and can also shape future legal precedent. It is these arguments made by lawyers that, in aggregate, represent the norms of the community.[30]

These arguments have the potential to change accepted legal principles or precedent through two primary strategies. First is through arguing for a more expansive or limited interpretation of a statute or legal principle. For example, a lawyer may argue that the First Amendment's protection of freedom of speech includes certain forms of expression that the government is trying to restrict. Second, lawyers can argue that existing legal principles or precedents should be changed or modified in light of changing societal values or as a matter of public policy. This argument is often based on a claim that a particular legal principle or precedent is outdated or does not adequately address current issues. Both of these strategies are heavily dependent upon community acceptance that the lawyer is correctly understanding both the law and the community it is serving.

To ensure lawyers are taking actions that have the potential to change the law from a position of understanding regarding the current law, jurisdictions typically have a set of rules and regulations in place to ensure that lawyers representing clients in front of the court are competent. These rules and regulations are designed to ensure that lawyers have the necessary education, training, and ethical standards to represent clients effectively.

For instance, nearly every jurisdiction requires bar admission as the primary way of ensuring competency. Lawyers must meet certain educational and character requirements to be admitted to the bar and practice law in a particular jurisdiction. For example, in the United States, lawyers must graduate from an accredited law school, pass a bar exam, and meet certain character and fitness standards to be admitted to practice law. Once admitted, most jurisdictions require lawyers to engage in ongoing education and training to maintain their competence. Lawyers may be required to complete a certain number of continuing legal education (CLE) credits each year to stay up-to-date on changes in the law and legal practice. In addition to education and training requirements, jurisdictions may also have rules and regulations in place to ensure ethical conduct and professional responsibility. For example, lawyers must adhere to rules of professional conduct that govern their behavior and ensure that they act in the best interests of their clients. Failure to comply with these rules can result in disciplinary action, including suspension or revocation of the lawyer's license to practice law.[31]

All of these rules are in place for protecting the authenticity with which the community, through the voice of those lawyers who represent members

of the community and the judges who preside over court actions, is accurately represented through a continually modifying justice system.

8. The Role of Lawyers in AI Legal Systems

The promise of access to justice tools that employ AI is rooted in the idea that such tools could eliminate the need for lawyers. If appropriately implemented, advocates believe general citizens could interact with an AI powered dispute resolution tool through the development of LLM-driven systems that direct participants through the procedures of justice towards an accepted resolution filed with the courts.[32] In this system, it is not correct to think that lawyers would just disappear. Lawyers, in terms of all parties with an influencing role in the outcome of case, therefore, will be subject to a vastly different role in the legal system. This is despite their role as trained professionals who have undertaken many years of training to attain their level of competence. Although not free from criticism, the public are considerably more forgiving and empathetic to human error rather than computational error which is expected to be faultless.[33] While AI is technically able to attain the level of a legal professional given its proven ability to pass the Uniform Bar Exam with a score within the 90th percentile,[34] raw legal prowess is an insufficient indicator of appropriate observance of legal norms. Where lawyers are subject to mechanisms of accountability which forms a core administrative legal principle, AI systems are unable to bear significant repercussions for their shortcomings and violations of ethics or proper legal procedure, but rather run the risk of being placed as a liability shield.[35] As such the retention of lawyers as a human in the loop remains a necessity in order to protect core legal principles at risk of AI overreach.[36] Therefore, lawyers would manifest themselves in a different manner: through the arguments they have made, the decisions they were party to, or the precedents they caused to be set are contained in the AI training data. LLM-based access to justice tools will require training on vast amounts of textual data representing community interests through the arguments made by the lawyers representing the community. These models use machine learning techniques to identify patterns in the data and develop a set of rules or patterns that can be used to make predictions or generate new text. These predictions and generated text represent the arguments and decisions that would be made or arrived at by the community, so long as the dataset was generated by the community.

As with any other computer system, an LLM operates solely based on the data to which it has been exposed. These datasets are used to "teach" the model how to recognize patterns and make predictions. But the very nature of modern AI/ML systems means they typically reflect the average of the dataset's opinions expressed in their training data and struggle to identify special circumstances or edge cases. As such, it is of utmost importance that there is large datasets of multiple cases in order to accurately automate legal predictions and have general applicability.[37] Yet even with large datasets this gravitation to the norm is a feature of the neural networks these tools are built upon, making them incapable of accurately applying specific logical processes or account for edge cases without them being directly coded into the system.[38]

If the outputs of the LLM are to be appropriate for a jurisdiction, they must be so on three grounds. The LLM training data must reflect the community bounded by that jurisdiction, meaning the model inputs should only be generated by individuals who have met the standards required of representing the community within that jurisdiction. Second, the datasets must be substantial enough to result in generalisable and predictable outcomes based upon that community's law without reference to law from other jurisdictions that would not ordinarily be cited in traditional legal precedents. And lastly, operational logic reflecting procedure specific to a jurisdiction must be directly encoded for instances when the law clearly requires a known cause to procures a specific effect.

9. Reformulating Digital Sovereignty

Protecting communities from the potential harm of AI systems often takes the framing of an outside force acting upon the affected population. In the legal technology vertical, this force can often be seen as anything from profit driven corporations to malevolent State actors.[39] This focus on protection from outside forces drives protection efforts towards the concepts of digital sovereignty, at whose heart is the concept of data sovereignty. While reasonable, AI-driven justice

technologies tools push us to realize that these strategies are fundamentally ineffectual.

Digital sovereignty refers to the idea that nations and individuals should have control over their own digital technologies, data, and infrastructure. The concept of digital sovereignty is based on the idea that the digital world has become a vital part of modern life, and that control over digital technologies and data is essential for maintaining national security, economic competitiveness, and personal privacy. In attempts to exert this control, the focus of digital sovereignty can be framed within the remit of traditional geopolitical sovereignty which has been subject to centuries of prior discourse.[40] Here, Krasner's quadripartite conception of sovereignty can be reworked as a basis to incorporate the challenges presented by an increasing use of AI in the legal profession[41]:

- Population is conceptualized as control over data. Digital sovereignty emphasizes the importance of individual and national control over personal data and information. This includes data privacy, data protection, and the ability to decide how and when data is collected, used, and shared.
- Territory is conceptualized as control over digital infrastructure. Digital sovereignty also involves control over the infrastructure and systems that support digital technologies. This includes control over networks, servers, and other digital hardware and software.
- Recognition is conceptualized as control over digital governance. Digital sovereignty emphasizes the importance of national sovereignty in digital governance and regulation. This includes the ability of nations to set their own rules and regulations for digital technologies and data, and the ability to enforce those rules and regulations.
- Regulation of borders is conceptualized as protection against cyber threats. Digital sovereignty also involves protecting against cyber threats such as cyber-attacks, cyber espionage, and cyber terrorism. This includes developing robust cybersecurity measures and protocols, and collaborating with other nations to combat cyber threats.

While traditional sovereignty concepts consider the population to be human individuals, digital sovereignty considers data itself to be the population that must be protected through rigorous control.[42] When defining this data population, the concept of data sovereignty typically features two unique aspects whose reasonableness AI-driven tools directly challenge:

- Data protection laws. Many countries have implemented data protection laws that regulate the collection, storage, and use of personal data. These laws give individuals control over their personal data and require organizations to obtain consent before collecting and processing personal data, and
- Data localization. Data localization is the practice of requiring that data be stored in a specific geographic location. This allows countries to maintain control over their citizens' data and protect it from foreign governments and companies.

The focus on these two aspects of data sovereignty are typically implemented by governments through restricting what data generated by one person's existence can be copyrighted by another without the generator's consent, and restricting the jurisdiction wherein the silicon upon which the generated dataset must be physically located.

AI tools challenge the reasonableness of modern data sovereignty constructs because, although they must access the data contained on the silicon that is intended to be protected by the concepts of digital and data sovereignty, the information perceived from an AI tool is a biproduct of the appropriate relationships interpreted between the training data. For United States Citizens, this can be illustrated by the difference between an integer 123456789, a person defined by social security number 123-45-6789, and a company defined by employer identification number 12-3456789.

The data generated by an individual is an artifact of their existence and cannot recreate a projection of their existence without the context of the individual. The information associated with this contextually derived assembly of the data is what makes any AI or LLM usable. This is why concepts of data sovereignty when considering the regulation of AI for LegalTech uses require a reconfigured, more appropriate "information sovereignty" concept.

In the same way that the laws of a jurisdiction are only accepted if they reflect the community contained within the jurisdiction, and the laws of a jurisdiction are made by the legal professionals

operating within that jurisdiction, an LLM is only appropriate for use within a jurisdiction if the data is assembled in a manner that incorporates the context of the legal professionals from within that jurisdiction. The location of the silicon upon which the data that assembles that data into information, or the location of the stochastic datasets that dynamically deploy that data within an AI tool, do pose a risk in the form of model access or reliability. But the appropriateness of an AI tool is based solely on its ability to represent the information gathered through observation of the population it will serve. This requires that tool suitability is defined by the source of information that was observed through the training of the model.

The fact that any LLM is little more than a technological mimic of the observations it is fed has become more rapidly understood than possibly any comparable revelation for any other transformative technology.[43] This means that, in the same way precedent in a jurisdiction would not be accepted if it was attempted to be made by a legal practitioner who is not authorized to practice in that jurisdiction, an AI LLM that is used by a jurisdiction must be restricted to assemblies of data that are deemed appropriate because they are trained upon observations of practitioners from that jurisdiction. This rethinking of how AI tools should be jurisdictionally restricted leads to a proposal of "information sovereignty" that could be represented as:

- Population. Model training must be limited to observations or interactions with individuals from that jurisdiction.
- Territory. The jurisdiction is not geographically constrained but instead inclusive of practitioners and systems operating within its represented community.
- Recognition. System outputs must be sufficiently auditable to verify that it is consistently reflecting an appropriate representation of community accepted practitioners.
- Regulation of borders. System outputs must be sufficiently immutable to prevent modification when transferred across systems.

In following this structure, AI could be used in such a way that it does not harm the democratic foundations of a community nor lead to unfounded or unrepresentative outcomes. Since LLMs are not at the stage where they can appropriately respond to concerns expressed by the legal community, sufficiently considering these four tenets would go a significant way to addressing these concerns and fortifying trust in AI. Until this is the case, it would be improper to consider LLMs as a sufficient device to contribute meaningfully towards access to justice on more than just a superficial level. Those who cannot afford traditional legal services still deserve representative legal outcomes and rights to due process. Where a case may hinge on a fine technicality, AI is unlikely to yet have the appropriate level of nuance to effectively respond. Whilst this remains the case, this variety of technology has not yet sufficiently evolved into a trusted legal tool.

10. The Risks of Doing Nothing

Shifting industry focus from one of digital sovereignty to information sovereignty will likely be a significant effort. In the meantime, the A2J community will have to grapple with the risks posed by current tools and weigh potential impacts. Doing this requires examining some of the prevalent comforts, fears, or mitigating strategies when considering appropriate strategies with respect to AI integration without information sovereignty protections into A2J systems. For instance, consider the following scenarios:

- "Drafting demand letter or communications can be done safely because it will always be reviewed before they go anywhere." As the world recently observed in Mata v. Avianca, Inc.,[44] even lawyers who are paid their full rate may have a tendency to rely too heavily on a technology that convincingly mimics intelligence. In Mata v Avianca, Inc., a brief filed with the court contained multiple citations that were invented by ChatGPT by combining fragments of real training data. The likelihood of AI generated drafts being given a less than appropriate review significantly increases when a case is being handled pro bono. When an AI system is so convincing and the outputs are not jurisdictionally constrained, A2J is depending on a pro bono attorney becomes effectively an on-

the-loop, active safeing system that must perform the labor intensive job of verifying facts in a document that appears correct.

- "Selecting appropriate forms or appropriate citations can be done now." Correct, form selection or citation reference when using appropriate search criteria can be successfully completed today. In theory, AI should be able to speed up these processes by requiring fewer less informed inputs from a user to find the most correct result faster. However, unless the system is using details other than those communicated by the user to the system through a language-based search, the model interpreting those search inputs must be built to accurately reflect the context of that jurisdiction. For example, damage value and circumstance play a significant role in understanding where a case can be filed, with that decision often varying by jurisdiction. If the AI model is not trained in a manner that accounts for such nuance, the expert system finding the right form with the wrong context could result in justice being denied.

- "Providing legal information through tools like Chatbots is a straightforward exercise that poses little risk." Apart from the fears of bias and inaccuracy that have been well documented in legal chatbot use cases,[45] the experience of New Jersey Courts in building its Judiciary Information Attendant (JIA) demonstrated that unanticipated questions could require up to 70% of inquiries be responded to by human attendants.[46] Where the JIA design sent inquiries to a call center when answers fell outside of rigid parameters, the nimbleness of an AI-powered chatbot could allow the system to more often believe it is fully understanding the inquiry in a manner that leads to a false response.

- "A poorly written brief poses little risk and will not be precedent setting." The risk posed by poorly cited or constructed arguments is often dismissed based on the idea that there are enough people in the system to catch any errors before they produce an impact. Yet in the same way it has been demonstrated that judges will too often ignore analytics in favor of their own biases a majority of the time when looking at pretrial diversion programs supposed by AI-enabled risk evaluations,[47] a judge is not infallible when spotting unsupported arguments that could become precedent setting. In cases where invalid arguments are accepted within the system, the threat of the judicial system's public acceptance rapidly grows. However, from an A2J perspective, the court's rejection of an invalid argument developed by a layperson is likely more immediately damaging because their access to fairness has been denied due to the AI system misdirecting them in the development of the brief presented to the court. In both scenarios, acceptance or rejection, public confidence is eroded either slowly or rapidly.

- "The model can just be finetuned to be safer." ChatGPT has proven that any system which is probabilistically assembling responses to prompts can easily produce erroneous answers. While many of these answers may seem to provide information that goes beyond what is contained in the training data, this interpretation is the technological equivalent of observing dinosaurs in the clouds. Since these erroneous answers are partly due to the inappropriateness of the dataset, finetuning the dataset through weighting or censoring is not a sufficient solution. Controlling a probabilistic system by reducing a probability does not eliminate its potential to emerge, which is why tools like ChatGPT can still believe the 2+3 could equal 87.[48] AI tools for A2J applications will not only need to have clear acceptability boundaries more akin to expert systems than ChatGPT-style AI, these protections need to be jurisdictionally bounded with logical relationship appropriate to a jurisdiction included in their evaluative structure in

order to be sure that any result accurately reflects a valid outcome.

11. Concluding Remarks

Whilst ostensibly the use of AI tools presents significant opportunities, at present it is plagued with risks and inconsistencies that would further jeopardize A2J in the long term if left unaddressed. By permitting an undeveloped system to act in lieu of the services of a legal professional, those who cannot afford a lawyer are directly disadvantaged with the less than normative creation of further barriers to A2J. As such, the improper use of AI tools as a replacement for conventional legal services has far-reaching implications, impacting the individual, their community and the traditional conception of the state. It is posited this will transpire primarily through jurisdictional overreach of AI tools that pose the substantial risk of blurring the delimitations of community law through datasets that fail to differentiate along jurisdictional boundaries.

The proposed starting point for a solution is set forth as a new conception of informational sovereignty to act as a bulwark for the protection of democracy and the individual. This is based upon the importance of limiting the model's training to observing individuals from the *population* in question, including the practitioners and systems operating with that *territory*, providing accountability through the *recognition* of reflecting the outputs of practitioners within that community whilst in doing so providing sufficiently immutable outputs to prevent modification outside *regulated borders*.

Although shifting the focus from digital to informational sovereignty will be subject to incremental change, this adapted criteria for the training of LLM's would be appropriate reassurance for communities to consider the use of AI tools in such a manner that would accelerate rather than inhibit A2J. In the meantime, mitigation of the risks is paramount given the invention of false evidence by LLM tools like ChatGPT, the lack of predictability and accuracy in outcomes and bias that threaten due process and A2J in legal systems.

[2] H. Ruschemeier, 'Squaring the Circle' https://verfassungsblog.de/squaring-the-circle/ (last accessed 8th May 2023)

[3] ABA Journal – D. Cassens Weiss, 'Latest Version of ChatGPT Aces Bar Exam With Score Nearing 90th Percentile' https://www.abajournal.com/web/article/latest-version-of-chatgpt-aces-the-bar-exam-with-score-in-90th-percentile (last accessed 9th May 2023)

[4] J. Villasenor, 'How AI Will Revolutionize the Practice of Law' https://www.brookings.edu/blog/techtank/2023/03/20/how-ai-will-revolutionize-the-practice-of-law/ (last accessed 8th May 2023)

[5] A. Buccella, ''AI For All' Is A Matter of Social Justice' (2022) AI and Ethics

[6] H. Kanu, 'Artificial Intelligence Poised to Hinder, Not Help Access to Justice' https://www.reuters.com/legal/transactional/artificial-intelligence-poised-hinder-not-help-access-justice-2023-04-25/ (last accessed 8th May 2023)

[7] Law Library: Library of Congress, 'Regulation of Artificial Intelligence in Selected Jurisdictions' https://tile.loc.gov/storage-services/service/ll/llglrd/2019668143/2019668143.pdf (last accessed 8th May 2023) p. 1-2

[8] A. Telang, 'The Promise and Peril of AI Legal Services to Equalize Justice' https://jolt.law.harvard.edu/digest/the-promise-and-peril-of-ai-legal-services-to-equalize-justice (last accessed 8th May 2023)

[9] A. Reichman and G. Sartor, 'Algorithms and Regulation' within 'Constitutional Challenges in the Algorithmic Society' eds H-W. Micklitz, O. Pollicino, A. Reichman, A. Simoncini, G. Sartor and G. De Gregorio (Cambridge University Press, 2022) p. 157

[10] C. Gans-Combe, 'Automated Justice: Issues, Benefits and Risks in the Use of Artificial Intelligence and Its Algorithms in Access to Justice and Law Enforcement' within 'Ethics, Integrity and Policymaking: The Value of the Case Study' eds D. O'Mathuna & R. Iphofen (Springer, 2022) p. 175

[11] R. Rodrigues, 'Legal and Human Rights Issues of AI: Gaps, Challenges and Vulnerabilities' (2020) Journal of Responsible Technology 4 100005

[12] United Nations Office on Drugs and Crime, 'Artificial Intelligence: A New Trojan Horse for Undue Influence on Judiciaries'

https://www.unodc.org/dohadeclaration/en/news/2019/06/artificial-intelligence_-a-new-trojan-horse-for-undue-influence-on-judiciaries.html (last accessed 9th May 2023)

[13] J. Soh Tsin Howe, 'Building Legal Datasets' https://datacentricai.org/neurips21/papers/74_CameraReady_building-legal-datasets-CamReady.pdf p. 1-2

[14] S. Wolfram, 'What Is ChatGPT Doing… and Why Does it Work?' (Wolfram Media, 2023)

[15] M. Kusak, 'Quality of Data Sets That Feed AI and Big Data Applications Enforcement' (2022) ERA Forum 23 p. 209

[16] Law Society Gazette, 'Will LawTech Extend Justice or Deepen the Digital Divide?' https://www.lawsociety.ie/gazette/top-stories2/will-lawtech-increase-access-to-justice-or-deepen-the-digital-divide (last accessed 8th May 2023)

[17] S. Rosengrun, 'Why AI is a Threat to the Rule of Law' (2022) Digital Society 1(10) p. 10

[18] O. Pollicino & G. De Gregorio, 'Constitutional Law in the Algorithmic Society' within 'Constitutional Challenges in the Algorithmic Society' eds H-W. Micklitz, O. Pollicino, A. Reichman, A. Simoncini, G. Sartor and G. De Gregorio (Cambridge University Press, 2022) p. 7

[19] M. Catanzariti, 'Algorithmic Law: Law Production by Data or Data Production by Law?' within 'Constitutional Challenges in the Algorithmic Society' eds H-W. Micklitz, O. Pollicino, A. Reichman, A. Simoncini, G. Sartor and G. De Gregorio (Cambridge University Press, 2022) p. 89

[20] R. Michaels, 'Legal Culture' available at: https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=3012&context=faculty_scholarship p. 1

[21] National Centre for Access to Justice, 'What is Access to Justice?' https://ncaj.org/what-access-justice, (last accessed May 27, 2023)

[22] Office for Access to Justice, 'About ATJ' https://www.justice.gov/atj/about-atj (last accessed 1st June 2023)

[23] ABA, 'Access to Justice' www.americanbar.org/topics/access/ (last accessed 1st June 2023)

[24] B. Jarrett & P. Hyslop, 'Justice for All: An Indigenous Community-Based Approach to Restorative Justice in Alaska' (2014) Northern Review 38 p. 239

[25] J. Folberg, 'A Mediation Overview: History and Dimension of Practice' (1983) Mediation Quarterly 1 p. 5

[26] J. Raz, 'The Rule of Law and its Virtue' within 'The Authority of Law: Essays on Law and Morality' (Oxford University Press, 1979) p. 210

[27] O. Pollicino & G. De Gregorio, 'Constitutional Law in the Algorithmic Society' within 'Constitutional Challenges in the Algorithmic Society' eds H-W. Micklitz, O. Pollicino, A. Reichman, A. Simoncini, G. Sartor and G. De Gregorio (Cambridge University Press, 2022) p. 7

[28] F. Ghodoosi, 'The Concept of Public Policy in Law: Revisiting the Role of the Public Policy Doctrine in the Enforcement of Private Legal Arrangements' (2016) Nebraska Law Review 94(68) p. 690

[29] M. Rotenberg, "Stifled Justice: The Unauthorized Practice of Law and Internet Legal Resources" (2012). Minnesota Law Review. 347 p. 731

[30] Judicature – D. F. Levi, D. Remus & A. Frisch, 'Reclaiming the Role of Lawyers as Community Connectors' https://judicature.duke.edu/articles/reclaiming-the-role-of-lawyers-as-community-connectors/ (last accessed 15th May 2023)

[31] American Bar Association, 'Model Rules of Professional Conduct – Table of Contents' https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/model_rules_of_professional_conduct_table_of_contents/ (last accessed 15th May 2023)

[32] A. Buccella, ''AI For All' Is A Matter of Social Justice' (2022) AI and Ethics

[33] A. Reichman & G. Sartor, 'Algorithms and Regulations' within 'Constitutional Challenges in the Algorithmic Society' eds H-W. Micklitz, O. Pollicino, A. Reichman, A. Simoncini, G. Sartor and G. De Gregorio (Cambridge University Press, 2022) p. 161

[34] ABA Journal – D. Cassens Weiss, 'Latest Version of ChatGPT Aces Bar Exam With Score Nearing 90th Percentile' https://www.abajournal.com/web/article/latest-version-of-chatgpt-aces-the-bar-exam-with-score-in-90th-percentile (last accessed 9th May 2023)

[35] J. J. Bryson, M. E. Diamantis & T. D. Grant, 'Of, for, and by the people: the legal lacuna of synthetic persons' (2017) Artificial Intelligence Law 25 p. 287

[36] A. Reichman & G. Sartor, 'Algorithms and Regulations' within 'Constitutional Challenges in the Algorithmic Society' eds H-W. Micklitz, O. Pollicino, A. Reichman, A. Simoncini, G. Sartor

and G. De Gregorio (Cambridge University Press, 2022) p. 174

[37] A. Reichman & G. Sartor, 'Algorithms and Regulations' within 'Constitutional Challenges in the Algorithmic Society' eds H-W. Micklitz, O. Pollicino, A. Reichman, A. Simoncini, G. Sartor and G. De Gregorio (Cambridge University Press, 2022) p. 157

[38] S. Wolfram, 'What Is ChatGPT Doing… and Why Does it Work?' (Wolfram Media, 2023) p. 99

[39] S. Rosengrun, 'Why AI is a Threat to the Rule of Law' (2022) Digital Society 1(10) p. 9

[40] T. Hobbes, 'Leviathan' (Harvard Classics, 1651) Chapter 13 Para 10; W. A. Dunning, 'Jean Bodin on Sovereignty' (1896) Political Science Quarterly 11(1) p. 92

[41] S. Krasner, 'Sovereignty: Organised Hypocrisy' (Princeton University Press, 1999) within this work Krasner sets out four variants of sovereignty: domestic (exercise of authority within a territory), interdependence (control over cross-border flow), international legal (recognition of territory by other territories) and Westphalian (non-intervention by others in the affairs of a territory)

[42] L. Amoore, 'Cloud geography: Computing, data, sovereignty' (2018) Progress in Human Geography 42(1) p. 16

[43] Boost.AI 'What are Large Language Models and How Do They Work?' https://www.boost.ai/blog/llms-large-language-models (last accessed 16th May 2023)

[44] E. Volokh, 'A lawyer's filing 'is replete with citations to non-existent cases' Thanks, Chat GPT?' https://reason.com/volokh/2023/05/27/a-lawyers-filing-is-replete-with-citations-to-non-existent-cases-thanks-chatgpt/ (accessed 28th May 2023)

[45] A. Asher-Schapiro & D. Sherfinski, 'Analysis: Chatbots in U.S. Justice System Raise Bias, Privacy Concerns' https://www.reuters.com/legal/litigation/chatbots-us-justice-system-raise-bias-privacy-concerns-2022-05-10/ (last accessed 28th May 2023)

[46] Joint Technology Committee, 'Introduction to AI for Courts' https://www.ncsc.org/__data/assets/pdf_file/0013/20830/2020-04-02-intro-to-ai-for-courts_final.pdf (28th May 2023)

[47] American Constitution Society, 'Roadblock to Reform' https://www.acslaw.org/wp-content/uploads/2018/11/RoadblockToReformReport.pdf p. 3 (last accessed 28th May 2023)

[48] S. Wolfram, 'What Is ChatGPT Doing… and Why Does it Work?' https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/ (last accessed 1st June 2023)