



## **International Virtual Colloquium on Multi-disciplinary Research Impact (2<sup>nd</sup> Series)**

Organised by Research Nexus UiTM (ReNeU)  
Office of Deputy Vice Chancellor (Research and Innovation)  
Universiti Teknologi MARA 40450 Shah Alam, Malaysia, 15 June 2022



# **Artificial Intelligence-Powered Criminal Sentencing in Malaysia: A conflict with the rule of law**

**Nurus Sakinatul Fikriah Mohd Shith Putera, Hartini Saripan,  
Mimi Sintia A. Bajury, Syazni Nadzirah Ya'cob**

Faculty of Law,  
Universiti Teknologi MARA, Shah Alam, Malaysia

[nurussakinatul@uitm.edu.my](mailto:nurussakinatul@uitm.edu.my), [hartinisaripan@uitm.edu.my](mailto:hartinisaripan@uitm.edu.my), [mimisintia@uitm.edu.my](mailto:mimisintia@uitm.edu.my), [syazni@uitm.edu.my](mailto:syazni@uitm.edu.my)  
Tel: +60195778006:

### **Abstract**

Despite its promising benefits, Artificial Intelligence-powered sentencing may infringe the fundamental principle of due process, presents unacceptable risks of error and implicit bias, and reliance on AI to predict recidivism which forms significant components of the rule of law. The present research, therefore, examines technical characteristics and legal theoretical consequences of the technology used in court processes. Employing jurisprudential analysis as the method of research, this research explores an adjudicatory paradigm that prefers standardisation over discretion, leading in the waning of the notion of rule of law pertinent to the justice system.

Keywords: Artificial Intelligence, Artificial Intelligence in Courts, Artificial Intelligence in Criminal Justice

*eISSN: 2398-4287© 2021. The Authors. Published for AMER ABRA cE-Bs by e-International Publishing House, Ltd., UK. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). Peer-review under responsibility of AMER (Association of Malaysian Environment-Behaviour Researchers), ABRA (Association of Behavioural Researchers on Asians/Africans/Arabians) and cE-Bs (Centre for Environment-Behaviour Studies), Faculty of Architecture, Planning & Surveying, Universiti Teknologi MARA, Malaysia.  
DOI: <https://doi.org/10.21834/ebpj.v7i1S17.3813>*

### **1.0 Introduction**

Technological innovations throughout history have consistently brought a sea of change in the daily lives of millions. Increased efficiency, productivity, and quality are just a few of the advantages that are associated with emerging technologies. The rapid growth of technological advancements such as AI, for instance, has changed the landscape of many fields, including, but not limited to, the legal realm. Traditional problems faced by the courts, such as high caseloads, backlogs, and a lack of resources, exert pressure on the court system (Contini, 2020) and impedes access to justice (Reiling, 2009). Reiling (Reiling, 2009) observed that the use of technology in the courtroom could lead to a much-needed improvement on access to justice, reduce corruption and restore the integrity and the impartiality of the judicial institution. While there are many dimensions to the problems faced by the judiciary, technological developments have contributed significantly towards the improvement and alleviation of difficulties.

In this setting, Artificial Intelligence (AI) is a continuously developing concept, and with each breakthrough, the traditional computer system becomes increasingly capable of performing a series of complex tasks and functions. The advent of Artificial Intelligence (AI) is already increasingly changing the legal practice and justice system. AI technologies have been proven to be very useful in the legal context as they are able to take over tasks formally performed only by humans (Campbell, 2020). Zalnieriute (Zalnieriute, 2021) posits that the technology could provide the key to eliminating prejudices and increasing transparency in the courtroom. Gipson Rankin (Gipson Rankin, 2021), observed that AI's comprehensive benefits allow for fewer administrative and repetitive tasks, leading to more efficiency. The use of AI in the sphere of justice, for instance, has the potential to transform it by, among other things, speeding up judicial procedures,

*eISSN: 2398-4287© 2021. The Authors. Published for AMER ABRA cE-Bs by e-International Publishing House, Ltd., UK. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). Peer-review under responsibility of AMER (Association of Malaysian Environment-Behaviour Researchers), ABRA (Association of Behavioural Researchers on Asians/Africans/Arabians) and cE-Bs (Centre for Environment-Behaviour Studies), Faculty of Architecture, Planning & Surveying, Universiti Teknologi MARA, Malaysia.  
DOI: [https://doi.org/10.21834/ebpj.v7i1S17%20\(Special%20Issue\).3813](https://doi.org/10.21834/ebpj.v7i1S17%20(Special%20Issue).3813)*

harmonising case law, expanding access to court, and boosting cost-efficiency (Dymitruk, 2020). Furthermore, Gipson Rankin (Gipson Rankin, 2021) notes that AI tools can be used to assist in more strategic legal duties such as problem-solving with its ever-developing sophisticated capabilities. AI tools have been proven useful in the area of legal research as they can assist lawyers in finding relevant cases and scouring for new materials round the clock (Campbell, 2020). Campbell (Campbell, 2020) notes the use of predictive analytics version AI by lawyers in order to predict the outcome of cases more accurately. It is opined that the employment of these AI based tools significantly changes how materials are presented before judges and how client risk is assessed (Hao, 2019). A growing body of literature has highlighted that AI could assist judges and the court process in numerous ways. This assistance could, in turn, result in less bias and an increase in transparency and efficiency. For example, electronic case allocation technology could reduce bias as judges are assigned random cases rather than being given the option to choose up to their own liking (Zalnieriute, 2021). In addition to this, the employment of an electronic case management system can help detect any irregularities and improve overall accountability (Zalnieriute, 2021). Another AI tool that has been lauded for its ability to improve the court's transparency is speech-to-text technology. As Contini (Contini, 2020) puts it, this particular technology ensures a speedier and more accurate hearing records and enables easier cross-checking along with document drafting. These types of AI tools help enhance transparency and efficiency without much interference with the due process of law as they are generally not involved with the judge's decision-making. At this background, a more controversial aspect of the use of AI in the legal field as highlighted by existing literature involved its current and potential future use in judicial decision making. Figure 1 below shows the cycle of different algorithm/technology used in the criminal justice system.

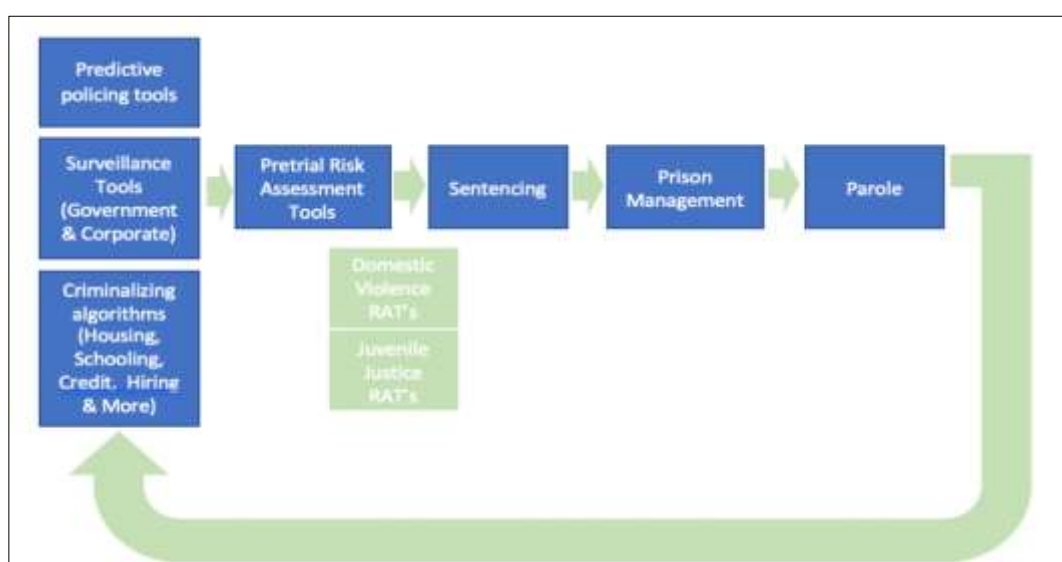


Figure 1: The Cycle of Different Algorithm/Technology Used in the Criminal Justice System  
 Source: (Electronic Privacy Information Centre, 2020)

As AI technology reaches new heights, its capabilities would gradually expand beyond sifting through vast quantities of data. For example, AIs are already developing cognitive abilities and learning from previous experiences and circumstances. The process of judging and deliberating outcomes of a case now can be supplemented or supported by AI based tools (Sourdin, 2018). This promptly leads to the debate of whether the role of the human judge will be fully replaced by intelligence-based machine in the future (Niler, 2019). After all, humans are more prone to biases, and the justice system, with a human adjudicator at the forefront, can be suspect to such. The so called 'AI Judge' in fact have already been making its way into the courtroom albeit they play a more supportive role rather than taking over the judge task itself (Sourdin, 2018). Although this is not to say that a full-fledge 'AI judge' is impossible. Estonia, for example, has undertaken an initiative to introduce a system where small claim disputes can be adjudicated by a robot (Dymitruk, 2020). In the United States, there is a noticeable presence of AI tools within the criminal justice process. States are increasingly encouraging the use AI-based predictive tools to assist judges in their decision-making (McMullan, 2019). The Correctional Offender Management Profiling for Alternative Sanctions or COMPAS, calculates recidivism risk by using a combination of static and dynamic factors (Jackson & Mendoza, 2020) and is one of the more popular versions of the tools. COMPAS has since been employed in numerous stages of the criminal justice process, and although it can be employed for purposes beyond sentencing, a number of states, including Wisconsin, Florida, and Michigan utilize it for sentencing purposes (Kehl et al., 2017). With more and more jurisdictions opting to utilize the technology, the legal, moral and ethical implications flowing from such decision are now a popular debate among scholars and experts alike. AI first made waves in the courts of Sabah and Sarawak, where it was deployed to assist in criminal sentencing. The AI Sentencing System (AISS) was developed in a collaboration between the judiciary and SAINS, Sabah and Sarawak's e-court systems. The Malaysian judiciary sent shockwaves on February 19, 2020, when it sentenced two accused persons under the Dangerous Drugs Act 1952, after months of deliberation. Following the trend, the Office of the Chief Registrar of the Federal Court of Malaysia recently issued a press statement on July 22, 2021, announcing that AI-based sentencing guidelines would be adopted in Peninsular Malaysia's Sessions Court and Magistrates Court. The implementation proceeds in three stages. This system requires the magistrates to enter the details of the accused persons, such as their age, citizenship, profession,

the weight of the drugs, marital status, and whether this was their first offence. The recommended sentences for the offences would later be displayed on a screen. This was accomplished using an AISS algorithm that analysed data from instances recorded in Sabah and Sarawak between 2014 and 2019, identifying patterns that may be applied to the current case. The AISS is now being tested on two offences, drug possession and rape. These offences were selected due to the large volume of data accessible by the Sabah and Sarawak courts. In determining the appropriate sentence, a wide range of considerations are given by the judge (Sourdin, 2018) and concerns have arisen on the issue of fair trial with the use of AI algorithms in the mix. AI technologies have also been used to ensure consistency in decisions made by the courts (Campbell, 2020). While AI is becoming more enmeshed in the criminal justice system, a serious threat to the rule of law is evidently looming from its applications in a domain that was once led by humans. As large strides are made in the academic discipline of AI, this technology is set to make its way into digital decision-making systems and is in effect replacing human decision-makers. However, in many instances, this technology is a "black box" due primarily to its complexity and also because it is protected by the law (Chatziathanasiou, 2022). A prominent objection to the employment of AI systems in adjudication and other contexts is that, in Fuller's words, they can appear to the person on the receiving end of their classifications to be a "masterpiece of obscurity" (Pasquale 2015). The reward function of an automated forecasting tool is encoded into its source code. It is not necessarily or readily available for examination by those subject to its regime. Even if the code were available, neither the "ordinary citizen" nor the "trained attorney" would find it informative. This lack of transparency and the limited capacity to comprehend the inner working of AI, which are increasingly utilised by the structures of governance, pose a challenge to the traditional concepts underpinning the rule of law. The absence of a "written code outlining the norms to be used in future disputes" is one of the ways in which Fuller sees the rule of law as failing. This is especially so in relation to concepts especially associated with the rule of law, such as transparency, fairness and explainability (Bagaric et al., 2022). This research, therefore, explores an adjudicatory paradigm that prefers standardisation over discretion, leading in the waning of the notion of rule of law pertinent to the justice system.

## 2.0 Literature Review

Past literature is replete with legal issues surrounding AI such as algorithmic transparency (Andreotta et al., 2021), cybersecurity vulnerabilities (Sherer et al., 2019), unfairness, bias and discrimination (Bagaric et al., 2022), lack of contestability, legal personhood issues (Abbott & Sarch, 2019), intellectual property issues, adverse effects on workers, privacy and data protection issues (Todoli-Signes, 2019), liability for damage and lack of accountability (Hallevy, 2018) – all of which may or may not have direct connection with the rule of law. The discussions pertaining AI and the conflict on the rule of law is evidently scarce, with only a few scholars establishing the legal impediments of AI under the ambit of the rule of law. For example (Greenstein, 2021) and (Huq, 2020) analyzed the use of AI to assist judges in making judicial decisions and how it challenges the traditional notions underpinning the rule of law, however a pragmatic way forward received insignificant consideration. In the Malaysian context specifically, the discourse on the interplay between AI and the legal realm is merely developing with little to no attention is given to the conflict on the rule of law within the frame of AI-powered criminal sentencing. Therefore, this research fills in the gap by consolidating the instances of which the use of AI-powered criminal sentencing in Malaysia defies the constructs of the rule of law. It further advances the discussion by proposing the adoption of the Explainable AI in the criminal justice system in shaping the nature and form of AI in the future.

The criminal justice system has sought to capitalise the latest capabilities in data analysis and processing — machine learning — to the sentencing task. The notion of the rule of law on the other hand, extends beyond making laws transparent, equitable, and accessible. Adherence to the principle entails holding the government and the business sector accountable. It implies that citizens should have access to fair and impartial conflict resolution and be able to anticipate that the mechanisms through which laws are developed, administered, and enforced are accessible and fair. These principles are particularly valuable for defining our understanding and thinking about the governance of AI. In light of the rule of law, it is argued that the AI-powered criminal sentencing tool used in the criminal justice system must aid due process or procedural fairness and judges must be informed of the caveats and inherent risks of such technology besides it being justifiable (Barra et al., 2018). Evidently, the deployment of AI-powered criminal sentencing ought to be supplemented with commitment to negate the reproduction of systemic discrimination. However, AI-driven outcomes in the context of criminal sentencing remain impenetrable (Kehl et al., 2017). This relates to the algorithm being proprietary or "black boxed" in nature - only the owners, and to a limited degree the purchaser, can understand the delivery of decisions by the software. The transparency issue of AI is two-fold (Washington, 2019). First, the evaluation for accuracy and bias for researchers and external experts is arguably a hurdle. The know-how of the system weighing the different factors and if some of them are more important than others are thus, indefinite.

The lack of information as to the internal operation of AI impede the utilisation of its application in criminal sentencing - criminal defendants cannot be certain whether or how questionable factors like gender or racial proxies may have impacted the risk assessment score or the judge's ultimate sentencing decision (Sukhodolov & Bychkova, 2018). The transparency challenges of AI have real life consequences, evident in *Loomis v State*; where the non-transparency inherent in a proprietary system has led the defendant to argue his inability to independently verify the accuracy of the tool. In *Loomis*, for example, the court rejects the gender claim because the sentencing judge did not directly mention it when explaining his decision - a distinction which seems to overlook the fact that a judge may never explicitly state a factor like gender when it is discreetly incorporated into an opaque risk score rather than considered openly in the presence investigation report or at a hearing (Liu et al., 2019) Hence, predicting the outcome of the algorithm is onerous, hampering the defence case. Secondly, while predictive algorithms in courtrooms are considerably "black boxes" to outsiders and are vulnerable to opacity issues, the proprietary tools developed by profit-oriented corporations pose distinct challenges (Stobbs et al., 2017). To remain competitive, these companies possess greater interest in keeping their algorithms away from public scrutiny. In contrary, academic researchers and

governments are more inclined to make the details of their algorithms publicly available and ensure that they are subject to appropriate scrutiny and oversight (Dressel & Farid, 2018).

### 3.0 Methodology

This research correlates its approaches on the premise of a legal interdisciplinary interest, a concept pivotal to the development of legal arguments by obtaining input from other discipline of knowledge, specifically for this research, the field of Artificial Intelligence-Powered Criminal Sentencing (Naudé Fourie, 2015). In a sphere of this research, the doctrinal analysis is utilized to integrate various statutory provisions, regulatory principles, interpretive guidelines, and framework (Hutchinson & Duncan, 2014) relevant to AI in the context of Artificial Intelligence-Powered Criminal Sentencing framework. The primary purpose of this method is also scrutinized in elucidating normative critics for the formulation of proposals for the future endeavours. In addition, the sources for the doctrinal analysis approach are gathered in the form of written sources such as statutory legislation, case law, regulation guideline documents, journal articles and reports retrieved from a library-based search. The library-based search was facilitated by the UiTM Online Public Access Catalog (OPAC) system to obtain primary data such as the Penal Code (Act 574), Criminal Procedure Code (Act 593), Subordinate Courts Act 1948, Criminal Justice Act 1953, and law cases (Ghapheryc & White, 2012). The secondary data, on the other hand, inclusive of journal articles and reports were acquired from browsing law databases such as Hein Online, the Current Law Journal, the Malayan Law Journal, and other databases namely Springer, Emerald, SAGE, ScienceDirect and others.

Data analysis perspectives embraced by this research include the comparative method and interpretive method. For comparative analysis, this research leverages on this method in producing suggestions to improve the existing legal position on AI as a powered criminal sentencing (Nelken, 2016). In general, comparative analysis observes a systematic perusal of rules, procedures, institutions, implementations within a single or multiple legal system based on objective comparative assessments of similarities, differences, and their repercussions (Van Hoecke, 2016). A few jurisdictions are selected as the country comparison for it provides more legal certainty and a better response to a particular event (Von Mehren et al. 1988). For this research, the Canada, United States, and the European Union are selected as the country comparison given the formulation of the research question which emphasizes on the adequacy of the Malaysian AI adjudication regulation in governing the risk that AI posed to the legal system. In this setting, the researchers' prior knowledge on Estonia, which has introduced a system where small claim disputes can be adjudicated by a robot (Dymitruk, 2020). Whilst in the United States, the AI tools were becoming more applicable in handling the criminal justice process. States are increasingly encouraging the use AI-based predictive tools to assist judges in their decision-making (McMullan, 2019). Interpretive analysis on the other hand, associate itself with the significance of doctrinal research in formulating legal doctrines through the analysis of legal rules. Unlike scientific research of which the validity of the research findings relies heavily on empirical investigation, the process of analysis by which doctrines are derived gravitate towards more idiosyncratic and argument-based approaches (Chynoweth, 2006).

### 4.0 Findings and Discussion

The research findings indicate that the advancements in the field of AI are a looming threat to the concept of rule of law as evidenced by its potential usage in the criminal justice system. As a legal concept, the rule of law is so elusive that the more one seeks to define it, the more diffuse it appears to become. The definition of the rule of law spans a wide range; thus, in addressing the harms associated with AI in relation to the rule of law, a common denominator that stands out is AI systems' lack of transparency which may escape our capacities for intuitive explanation (Williams et al., 2019).

#### 4.1 AI's Transparency Paradox and the the Notion of Accessibility to the Law

The concept of the rule of law has a wide variety of interpretations, which ranges anything from its utility in preventing abuse of power to its function as a mechanism for expressing the necessary characteristics towards building a more equitable society (Greenstein, 2021). A primary characteristic of the rule of law is that it incorporates the concept of reciprocity between the government and the people. Those in the position of power or authority must respect the confines of laws; meanwhile, the people must abide by the established legal norms which treats everyone equally. It has been opined that the introduction of AI into the judicial institution could potentially disrupt this very balance. Fundamentally, when those who are governed are left in the dark about the technologies used developed by private entities which have or could potentially replace the government or state. If the citizens are kept in the dark, the law remains a 'black-box' and there is lack of consistency (due to the machine learning nature of AI), then citizens will no longer have a moral duty to obey these law notes. Further, the rule of law revolves around the idea of a transparent and accountable government that respects the tenets of laws. Gowder (Gowder, 2016) notes that transparency necessitates public knowledge of how the state operates and allows easy access to legal norms and administrative decisions (Gowder, 2016). This provides the people with the opportunity to know and understand what is expected of them, subsequently giving them the agency to choose on how they conduct their daily affairs. Additionally, predictability and consistency in the laws will promote obedience as the people understand the expectation and consequences that could arise from non-compliance. The issue of transparency and accountability was raised in numerous literatures as a main cause of concern when it comes to using AI in the judicial decision-making process. For instance, the primary concern is whether an automated system such as a computer programme has the legal authority to make judgements about cases instead of a human judge. Justice Perry raises questions on who determines and who has the authority to make such a judgement in relation to an automated system that creates administrative rulings. Which comes first: the automated system itself, the computer programmer, the policymaker, or the decision-maker? (Sourdin, 2018).

The lack of transparency involved in the use of AI tools in the court process could arguably undermine the judicial institution authority and subsequently the rule of law itself. Critics have suggested that AI systems should at least be explainable, interpretable, or transparent when used as a decision-making tool (Williams et al., 2019). The European Commission for the Efficiency of Justice, for instance, had taken the initiative to govern the use of AI in the courtroom by publishing a charter which emphasized on the principles of non-discrimination, transparency, and respect for fundamental rights (McMullan, 2019). Others have argued that the algorithms themselves should be compelled to generate automated explanations for why they made the decisions they did. The threat to the rule of law, according to Greenstein (Greenstein, 2021), stems from the fact that most of the decision-making systems are complex 'black-boxes'. He posited that the lack of insight, transparency, fairness and explainability makes them more prone to undermining the integrity of the law and its traditional elements as we know it (Greenstein, 2021). The COMPAS tool, as adopted by a growing number of United States' jurisdictions to assist courts in the area of criminal justice, including sentencing (Kehl et al., 2017), serves as an example of some of the problems that may be encountered when the courts employ an intelligent system. COMPAS uses algorithms (most often machine learning) which predict the risk of relapse into criminal activities and rate defendants based on the likelihood that they will commit a new crime if released. Since the algorithms have not been disclosed, the methods they use to estimate recidivism risks remain unknown (Wang, 2020).

Although defendants are informed of the variables used to measure their recidivism risk in the COMPAS tool, arguably, this offers little to no comfort as they still do not know how the variables are weighed (Lightbourne, 2017). It has been argued that mere information does not amount to an explanation, and this would prove to pose a problem if they were ever to challenge the decision made (Wang, 2020). Reiling (Reiling, 2020) suggests that in order to properly utilize AI in the judicial process, the AI in question must be able to provide clarification on how it arrived at the conclusion it did. Justifications for decisions made are important to the notion of the rule of law as they could reveal any shortcoming or errors, such as whether the adjudicators in question have acted arbitrarily or outside its jurisdiction and serve as evidence to support any case against the initial decision (Kemper, 2020). In addition to this, any technological reforms with the promise to improve the justice system must also include the judges. It is vital for judges to understand to a certain extent how the technology that is supposed to assist them works. In *Loomis v Wisconsin*, it has been noted that the judge presiding the case constantly inquiring both parties on how the technology operates (Kehl et al., 2017). Surden (Surden, 2019) notes that the issue of interpretability of AI systems and transparency around how AI systems are making their decisions may further complicates its relationship with the legal realm and raises concerns with regards to the principle of rule of law. It is opined that AI systems are built in such a manner that even the programmers behind them would have an issue in deciphering their underlying mechanism. (Christian, 2020). Using them to assist decision-making in areas such as criminal justice where the stakes are high, has been considered dangerous due to the insufficient understanding of the technologies in question (Chiao, 2019). Chiao notes the machine-learning techniques employed in the AI tools, which can evolve and learn new things on its own over time, something similar to on-the-job training for humans (Jain, 2019), poses a distinct set of concerns. Its ability to evolve could lead to a situation where the same inputs at two distinct point of time results in different outcomes (Villasenor & Foggo, 2020). What was initially programmed by the creator may have changed due to machine-learning techniques, which renders very few people who could understand the AI being used. Chiao (Chiao, 2019) distinguished AI tools from other complex technologies such as airplanes noting that with the latter, even if laymen or the user of the technology itself does not completely understand its inner workings, at least the experts in the field do whereas an AI tool with machine learning could change unpredictably over the course of time.

#### 4.2 The Demand for Explainable Artificial Intelligence

Due to the ways algorithms calculate their "reasoning," the mechanisms that give rise to the recommendations given by predictive algorithms are now unknown and undiscoverable. AI in this form is often referred to as "black-box" AI. The "neural networks" that power black-box AI algorithms are built to imitate the human brain, allowing them to self-teach, make decisions, and respond accurately. Despite the technology's ability to offer accurate responses and aid in improving the cost, access, and quality of the legal service, the algorithms by which it operates may "become less understandable to users and even the developers who initially designed the technology." This means that AI will be unable to demonstrate how it arrived at its judgments. Transparency in AI systems is challenging to define since it encompasses numerous distinct considerations. For some, transparency may mean "discoverable" - another ill-defined phrase - while for others, it translates into meeting certain standards. Others may argue that transparency merely requires consumers to have access to information concerning training metrics. Another level of complexity to address, is the question of who a system must be visible to - the user. - and whether this transparency should represent merely explainability (the "how" it works) or understandability (the "what" it does) (the "why" it delivers a given answer). Even if the algorithm is precisely coded within the machine-learning process (a step-by-step approach for solving a problem) and treats everyone equally, it can still give discriminatory results (Dueno, 2020). Duncan added that even in an ideal world, when data scientists meticulously construct the algorithm to avoid prejudice on the basis of race, ethnic origin, religion, or any other socioeconomic element, the underlying data may be biased, producing skewed conclusions. Even for the researchers who built the systems, the operations of machine learning approaches are opaque, and while this may not be a problem in many areas of applied machine learning, AI systems must be transparent when used in judicial settings, where the explainability of decisions and the transparency of reasoning are of significant—even civilizational—importance (Završnik, 2020). A decision-making process that is hidden and difficult to understand is not regarded legitimate or autocratic. The new techniques deployed in criminal justice situations may thus be at odds with fundamental rights due to the intrinsically opaque character of these AI systems.

The rise of AI has so far occurred in a regulatory vacuum (Calo, 2017). Except for few legislations regarding driverless vehicles and drones, laws or regulations that explicitly appraise the exceptional challenges posed by AI are inadequate, and virtually no courts appear to have established practises specifically addressing who should be held legally responsible if an AI causes harm (Prakken, 2017). There is a disturbing dearth of legal scholarship discussing potential regulatory approaches to AI. It does not appear that any prevailing scholarship examines AI regulation through the lens of healthcare institutional competence — that is, the issue of what type(s) of regulatory

compliance would be best equipped to confront the unique challenges presented by the rise of AI (Li et al., 2018). This primarily revolves around the explainability notion - the call for AI decisions to be transparent and interpretable to human users. This approach is expected to be regulated in the light of limiting the functionality of AI applications to those which are explainable (Gunning et al., 2019). Consequently, the development of AI adoption and its performance is predicted to be hindered by XAI, leading to Google releasing a 2019 white paper calling for government proactivity in setting out standards for explainability (Robbins, 2019). One of the earliest measures to address the issue of explainability has been the European Union's (EU) General Data Protection Regulation (GDPR) which is a 2016 EU regulation addressing data protection and privacy. The GDPR surfaced as the prominent global standard for many fundamentals of data and AI policy due to the extraterritorial reach of its provisions. GDPR has been adaptive as to its implications for AI and recently understood as requiring that AI must be explainable, however subject to interpretation (Lu et al., 2019). The precise provision, Article 15, necessitates that in AI-based decision making, relevant parties have the right to access "meaningful information about the logic involved, as well as the significance and envisaged consequences of such processing for the data subject" – or better known as the "right to an explanation" (Kim & Routledge, 2017). Irrespective of the term "right to an explanation" appearing only in the non-binding recitals, other jurisdictions including Canada have since called for mandatory explainability in AI. In Canada, however, the legal requirement addressing explainability is absent. Instead, documentation from the Treasury Board Secretariat (TBS) suggesting that AI ought to be "auditable" and that a "meaningful explanation" be provided upon request is recognised as the ruling policy (Gaon & Stedman, 2019). Canada's pending international legal obligations, such as in the United States, Mexico, Canada Agreement (USMCA) and Comprehensive Trans-Pacific Partnership (CTPP), approach differently indicating that a punitive adherence to the principle of AI explainability would be observed as an undue barrier to trade (Castelvecchi, 2016). In this perspective, these agreements severely hinder the punishment of companies for a lack of explainability in their AI. Thus far, none of these documents reflect the concluding remarks on explainability since all of the legal signalling about explainability has yet to be tested by the judicial system. Consequently, the principle of AI explainability fitting within the existing Canadian legal realm and policy is ambiguous.

## 6.0 Conclusion and Recommendations

AI is entering the courts intrusively, bringing profound impact. But the convenience of this disruptive technology comes at a cost – one that deals with the domain humans were conventionally the sole actor. To equate judging within the confinement of data processing is rather erroneous, thus a proper legal construct in demarcating the line separating that of human judge and data-driven system will help ensure that AI plays a principled and appropriate role in advancing dispute resolutions. Due to the rapid developments in this field, the challenges and prospects related to harnessing AI in judicial systems and their repercussions for human rights and the rule of law ought to be carefully deliberated among stakeholders from the judicial ecosystem. In this context, the research recommends that resources should be leveraged at venturing into the development of the principle of Explainable AI and finding ways to incorporate the principle into the criminal justice system in order to resolve the conflict on the rule of law. This principle is capable of explaining the manner of which the algorithm produces its output and is much needed to allow human users' comprehension and confidence of the output. Hence, the principle of Explainable AI helps characterise its accuracy, fairness and transparency in AI-powered decision making is possibly one of the ways in addressing the constraints affecting the rule of law discussed previously (Deeks, 2019). Though this research contributes to the pool of discussion regarding the application of the Explainable for AI-powered criminal sentencing in Malaysia, the research has yet to explore on a pragmatic way to develop rules for xAI given the nascent stage of using the system in Malaysia. Therefore, future research could consider developing a database of existing case law in light of the Xai principle, tailored to the different requirements of explanation in an array of litigation. This will eventually provide informative incentives to the AI developers as they strive to translate the present theoretical debates into operative xAI tools.

## Acknowledgement

The authors would like to express their gratitude for the financial support from Universiti Teknologi MARA under the Lex Praesta Research Grant (600-TNCPI 5/3/DDF (FUU) (001/2020) granted for this research.

## References

- Abbott, R. B., & Sarch, A. F. (2019). Punishing Artificial Intelligence: Legal Fiction or Science Fiction. *SSRN Electronic Journal*, 323–384. <https://doi.org/10.2139/ssrn.3327485>
- Andreotta, A. J., Kirkham, N., & Rizzi, M. (2021). AI , big data , and the future of consent. *AI & SOCIETY*, 6(2019), 1–14. <https://doi.org/10.1007/s00146-021-01262-5>
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*.
- Bagaric, M., Svilar, J., Bull, M., Hunter, D., Rev., N. S.-A. C. L., & 2022, undefined. (2022). The Solution to the Pervasive Bias and Discrimination in the Criminal Justice System: Transparent and Fair Artificial Intelligence. *HeinOnline*. [https://heinonline.org/hol-cgi-bin/get\\_pdf.cgi?handle=hein.journals/amcrlml95&section=7](https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/amcrlml95&section=7)
- Barra, S., Bessler, C., Landolt, M. A., & Aebi, M. (2018). Testing the validity of criminal risk assessment tools in sexually abusive youth. *Psychological Assessment*. <https://doi.org/10.1037/pas0000590>

- Campbell, R. W. (2020). Artificial Intelligence in the Courtroom : the Delivery of Justice in the Age of Machine Learning. *Colorado Technology Law Journal*, 18(2), 323–350.
- Castelvecchi, D. (2016). The Black Box 2.0. *Nature*, 538(7623), 20–23. <http://www.nature.com/news/can-we-open-the-black-box-of-ai-1.20731>
- Chatziathanasiou, K. (2022). Beware the Lure of Narratives: “Hungry Judges” Should Not Motivate the Use of “Artificial Intelligence” in Law. *German Law Journal*, 23(4), 452–464. <https://doi.org/10.1017/glj.2022.32>
- Chiao, V. (2019). Fairness, accountability and transparency: Notes on algorithmic decision-making in criminal justice. *International Journal of Law in Context*, 15(2), 126–139. <https://doi.org/10.1017/S1744552319000077>
- Christian, G. (2020). Predictive Coding: Adopting and Adapting Artificial Intelligence (AI) In Civil Litigation. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3530039>
- Contini, F. (2020). Artificial Intelligence and the Transformation of Humans, Law and Technology Interactions in Judicial Proceedings. *Law, Technology and Humans*, 2(1), 4–18. <https://doi.org/10.5204/lthj.v2i1.1478>
- Diego Carreras, A. de. (2020). The Moral (Un)Intelligence Problem of Artificial Intelligence in Criminal Justice: A Comparative Analysis under Different Theories of Punishment. *UCLA Journal of Law & Technology*, 25(1), 1–38.
- Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*. <https://doi.org/10.1126/sciadv.aao5580>
- Dueno, T. (2020). Racist Robots And The Lack Of Legal Remedies In The Use Of Artificial Intelligence In Healthcare. *Connecticut Insurance Law Journal*, 27(1), 337–358.
- Dymitruk, M. (2020). Artificial Intelligence as a Tool to Improve the Administration of Justice? *Acta Universitatis Sapientiae Legal Studies*, 8(2), 179–189. <https://doi.org/10.47745/ausleg.2019.8.2.03>
- Electronic Privacy Information Centre. (2020). *AI and Human Rights: Criminal Justice System*. Electronic Privacy Information Centre. <https://epic.org/ai/criminal-justice/index.html>
- GAON, A., & STEDMAN, I. (2019). ACALL TO ACTION:MOVING FORWARD WITH THE GOVERNANCE OF ARTIFICIAL INTELLIGENCE IN CANADA. *Alberta Law Review*, 56(4), 1–8.
- Gipson Rankin, S. (2021). Technological Tethereds: Potential Impact of Untrustworthy Artificial Intelligence in Criminal Justice Risk Assessment Instruments. *Washington and Lee Law Review*, 78(2), 647–724. <https://doi.org/10.2139/ssrn.3662761>
- Gowder, P. (2016). The Rule of Law in the Real World. *The Rule of Law in the Real World*, 1–275. <https://doi.org/10.1017/CBO9781316480182>
- Greenstein, S. (2021). Preserving the rule of law in the era of artificial intelligence (AI). In *Artificial Intelligence and Law* (Issue 0123456789). Springer Netherlands. <https://doi.org/10.1007/s10506-021-09294-4>
- Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2019). XAI-Explainable artificial intelligence. *Science Robotics*. <https://doi.org/10.1126/scirobotics.aay7120>
- Halley, G. (2018). The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control. *Akron Intellectual Property Journal*, 2(2), 2016. <https://doi.org/10.1016/j.gecco.2019.e00539><https://doi.org/10.1016/j.foreco.2018.06.029>[http://www.cpsg.org/sites/cbsg.org/files/documents/Sunda\\_Pangolin\\_National\\_Conservation\\_Strategy\\_and\\_Action\\_Plan\\_28LoRes29.pdf](http://www.cpsg.org/sites/cbsg.org/files/documents/Sunda_Pangolin_National_Conservation_Strategy_and_Action_Plan_28LoRes29.pdf)<https://doi.org/10.1016/j.forec>
- Hao, K. (2019). *AI is sending people to jail—and getting it wrong | MIT Technology Review*. MIT Technology Review. <https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/>
- Huq, A. Z. (2020). *Artificial Intelligence and the Rule of Law (forthcoming in The Routledge Handbook of the Rule of Law)*. 1964, 1–14. <https://ssrn.com/abstract=3794777>
- Jackson, E., & Mendoza, C. (2020). Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not. *Harvard Data Science Review*, 2(1), 1–15. <https://doi.org/10.1162/99608f92.1b3dadaa>
- Jain, P. (2019). *Artificial Intelligence for sustainable and effective justice delivery in India*.
- Julia Angwin, Jeff Larson, Surya Mattu, & Lauren Kirchner. (2016). Machine bias. *ProPublica*.
- Kehl, D., Guo, P., & Kessler, S. (2017). Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing. *Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School*.
- Kemper, C. (2020). Kafkaesque AI? Legal Decision-Making in the Era of Machine Learning. *Intellectual Property and Technology Law Journal*, 24(2), 251–294.
- Kim, T. W., & Routledge, B. R. (2017). Why a Right to an Explanation of Algorithmic Decision-Making Should Exist: A Trust-Based Approach. *International Data Privacy Law*, 32(4), 1–25. <https://doi.org/10.2139/ssrn.3716519>
- Lightbourne, J. (2017). Damned Lies & Criminal Sentencing Using Evidence-Based Tools. *Duke Law & Technology Review*.
- Liu, H. W., Lin, C. F., & Chen, Y. J. (2019). Beyond state v loomis: Artificial intelligence, government algorithmization and accountability. *International Journal of Law and Information Technology*. <https://doi.org/10.1093/ijlit/ez001>
- Loeb, J. (2018). Algorithms on the beat: AI and the future of policing. *Engineering & Technology*. <https://doi.org/10.1049/et.2018.0401>
- Lu, J., Lee, D. (DK), Kim, T. W., & Danks, D. (2019). Good Explanation for Algorithmic Transparency. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3503603>



- McMullan, T. (2019). *A.I. Judges: The Future of Justice Hangs in the Balance*.
- Niler, E. (2019). *Can AI Be a Fair Judge in Court? Estonia Thinks So | WIRED*. Wired. <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/>
- O'Brien, T. (2020). Compounding Injustice: The Cascading Effect of Algorithmic Bias in Risk Assessments. *Georgetown Journal of Law & Modern Critical Race Perspectives*, 13(1), 39–83. <https://doi.org/10.2139/ssrn.3694818>
- Prakken, H. (2017). On the problem of making autonomous vehicles conform to traffic law. *Artificial Intelligence and Law*, 25(3), 341–363. <https://doi.org/10.1007/s10506-017-9210-0>
- Reiling, D. (2009). Technology for Justice: How Information Technology Can Support Judicial Reform. In *Leiden University Press*.
- Reiling, D. (2020). Courts and Artificial Intelligence. *International Journal for Court Administration*, 11(2), 1–10. <https://doi.org/10.36745/IJCA.343>
- Rich, M. (2016). Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment. *University of Pennsylvania Law Review*.
- Rigano, C. (National I. of J. (2019). Using artificial intelligence to address criminal justice needs. *NIJ Journal*.
- Rizer, A., & Watney, C. (2018). Artificial Intelligence Can Make Our Jail System More Efficient, Equitable and Just. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3129576>
- Robbins, S. (2019). A Misdirected Principle with a Catch: Explicability for AI. *Minds and Machines*, 29(4), 495–514. <https://doi.org/10.1007/s11023-019-09509-3>
- Sherer, J. A., Hoffman, T. M., & Fedeles, E. R. (2019). Artificial Intelligence and Associated Clinical Data Privacy Considerations. *BMC Medical Ethics*, 22(122), 1–5.
- Sourdin, T. (2018). Judge v robot? Artificial intelligence and judicial decision-making. *University of New South Wales Law Journal*, 41(4), 1114–1133.
- Stobbs, N., Hunter, D., & Bagaric, M. (2017). Can Sentencing Be Enhanced by the Use of Artificial Intelligence? *Criminal Law Journal*.
- Sukhodolov, A. P., & Bychkova, A. M. (2018). Artificial intelligence in crime counteraction, prediction, prevention and evolution. *Russian Journal of Criminology*. [https://doi.org/10.17150/2500-4255.2018.12\(6\).753-766](https://doi.org/10.17150/2500-4255.2018.12(6).753-766)
- Surden, H. (2019). Artificial intelligence and law: An Overview. *Georgia State University Law Review*, 35(4), 1306–1337.
- Todoli-Signes, A. (2019). Algorithms, artificial intelligence and automated decisions concerning workers and the risks of discrimination: the necessary collective governance of data protection. *Transfer*, 25(4), 465–481. <https://doi.org/10.1177/1024258919876416>
- Villasenor, J., & Foggo, V. (2020). Artificial Intelligence, Due Process and Criminal Sentencing. *Michigan State Law Review*, 2, 295–354.
- Wang, N. (2020). Black Box Justice: Robot Judges and AI-based Judgment Process in China's Court System. *IEEE Xplore*, 58–65.
- Washington, A. L. (2019). How to Argue with an Algorithm : Lessons from the COMPAS- ProPublica Debate (Working Draft). *The Colorado Technology Law Journal*.
- Williams, G., Zalnieriute, M., & Moses, L. B. (2019). The Rule of Law and Automation of Government Decision-Making. *Modern Law Review*, 82(3).
- Zalnieriute, M. (2021). *Technology and the Courts : Artificial Intelligence and Judicial Impartiality Submission to Australian Law Reform Commission Review of Judicial Impartiality*. June, 1–8.
- Završnik, A. (2020). Criminal justice, artificial intelligence systems, and human rights. *ERA Forum*. <https://doi.org/10.1007/s12027-020-00602-0>