

Transparency in Predictive Algorithms: A Judicial Perspective

Md. Abdul Malek
Judicial Officer,
Bangladesh Judicial Service, Bangladesh
Correspondence: abmalek.b@gmail.com

ABSTRACT:

Notwithstanding the apparent hyperbole about AI promises for judicial modernization, there arise deep concerns that span from unfairness, privacy invasion, bias, discrimination to the lack of transparency and legitimacy, etc. Likewise, critics branded their application in the judicial precincts as ethically, legally, and technically distressing. Accordingly, whereas there is already an ongoing transparency debate on board, this paper attempts to revisit, extend and contribute to such simmering debate with a particular focus from a judicial perspective. Since preserving and promoting trust and confidence in the judiciary as a whole appears to be imperative, it uses a searchlight to explore how and why justice algorithms ought to be transparent as to their training data, methods, and outcomes. This paper also ends up delineating the tentative paths to do away with black-box effects and suggesting the way out for the use of algorithms in high-stake areas like the judicial settings.

KEYWORDS: Predictive algorithms, algorithmic transparency, and predictive justice.

INTRODUCTION:

In the advent of the increased availability of data and increased computing power, intelligent machines are making a sheer shift in all walks of society by leaps and bound. The expansion of the use of big data analytics and recent improvements in deep learning methodology and machine learning algorithms also influences justice sectors, commonly criminal courts. Decisions about parole, probation, bail, and sentencing decisions are now being made in the aid of the risk assessment algorithms. Such ‘algorithmic governmentality’¹ is mostly found in the United States of America. For example, the American Law Institute also recommends a broader use of risk-assessment tools in the most recent version of the highly influential Model Penal Code.²

Despite the facts as such, the typical black-box manner of the highly successful machine learning and artificial intelligence models attracts strict scrutiny from the viewpoint of law and justice when they are applied in the judicial settings. Consequently, AI-assistive risks give rise to deep concerns from unfairness, privacy breach, bias, discrimination, and lack of legitimacy to lack of transparency and explainability, etc. Critics also condemn that using

¹See Antoinette Rouvroy, and Thomas Berns, “Algorithmic Governmentality and Prospects of Emancipation Disparateness as a Precondition for Individuation through Relationships?” (2013) 177 *Réseaux* 163 <https://www.cairn-int.info/article-E_RES_177_0163--algorithmic-governmentality-and-%20prospect.htm#>.

²See also Sonja B Starr, “Evidence-Based Sentencing and the Scientific Rationalization of Discrimination.” (2014) 66 *Stanford Law Review* 815.

such algorithmic risk assessment violates the principle of due process, individual rights, and deviates from the principle of fairness in the court proceedings. More importantly, as the black-box model of artificial intelligence provides no information about what exactly makes them arrive at their predictions, “the development of methods for visualizing, explaining and interpreting deep learning models has recently attracted increasing attention”.³

Likewise, critics and researchers have already raised their sufficient voices and deep concerns against the indiscriminate use of this algorithmic assessment in the judicial domain thanks to their inherent opacity and black-box effects which stand as “dark secret at the heart of AI”.⁴ As such, they also opined that such algorithmic practices are thus “very dangerous, unreliable”⁵ and “constitutionally, technically, and morally troubling”.⁶ It is so contended because risk assessment algorithms are proprietary black-box tools, over which there is little regulatory oversight, and transparency about how and what they work.

A number of researches also show that opaque algorithms can undercut people’s sense of fairness and trust—particularly when used in the judicial setting. In order to identify biases and errors in the algorithm and affording a defendant’s right to defense, “there are lately increasing demands for introducing transparent, expandable and responsible AI systems for the use of high-stakes policy spaces; and judicial use is no exception at all.”⁷ Hence, such concerns are not ignoble as it appears to be an antithesis to some established legal concepts of criminal procedure law, and significantly impactful on individuals’ rights, adjudicatory values, judicial fairness and integrity, and legitimacy. As such, predictive algorithms appears to be more consequential nowhere than in the criminal justice system.

Consequently, the judicial use of such predictive algorithms has already faced challenges, even in the Supreme Court of the USA in the case of ‘State v Loomis’.⁸ This case again robustly sparks the debate even in new height and gives birth afresh a set of legal and ethical questions as to their algorithms implications. Hence, this paper argues that the judicial decision-making process fairly requires more transparent, explainable, and open algorithmic models, alongside a sufficient level of fairness, and accuracy.

³Wojciech Samek, Thomas Wiegand and Klaus-Robert Müller, “Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models” [2017] ITU Journal: ICT Discoveries, Special Issue No. 1 <<https://arxiv.org/abs/1708.08296>> accessed September 9, 2020.

⁴Knight W, “The Dark Secret at the Heart of AI” (*MIT Technology Review*, 2017) <<https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/>> accessed December 12, 2020

⁵Hany Farid, “The Danger of Predictive Algorithms in Criminal Justice | Hany Farid | TEDxAmoskeagMillyard” <<https://www.youtube.com/watch?v=p-82YeUPQh0>>.

⁶Sonja B Starr, “Evidence-Based Sentencing and the Scientific Rationalization of Discrimination.” (2014) 66 *Stanford Law Review*, 870.

⁷In the UK, for example, such calls have come from the House of Lords AI Committee, which argued that “the development of intelligible AI systems is a fundamental necessity if AI is to become an integral and trusted tool in our society. The EU’s high-Level Group on AI has called for further work to define pathways to achieving explainability; and in the US, the Defence Advanced Research Projects Agency supports a major research programme seeking to create more explainable AI. See “Explainable AI: The Basics POLICY BRIEFING” (2019) <<https://royalsociety.org/-/media/policy/projects/explainable-ai/AI-and-interpretability-policy-briefing.pdf>>.

⁸Han-Wei Liu, Ching-Fu Lin and Yu-Jie Chen, “Beyond State v Loomis: Artificial Intelligence, Government Algorithmization and Accountability” (2019) 27 *International Journal of Law and Information Technology*, 122.

But then again, another question comes off to what extent and what type of transparency is required for informing the judicial decision-making; why are they so important and different otherwise than other public spaces like police and regulatory bodies. The argument of this paper thus encompasses the periphery of algorithmic transparency, and also includes some allied concepts like explainability, interpretability, or comprehensibility within the spotlights of discussion; simply because a justice system already has some intrinsic epistemological aphorism beyond the technological ones.

On that account, this paper also reflects that there goes and grows a simmering discussion and debate on whether a jurisdiction should either hail or celebrate this rampant algorithmization in justice sectors, or want a pause as flawed tools, or demand for more research from diverse perspectives. In the existence of such rigorous dialogical ambiance, this paper trends to synthesize, visit and extend this ongoing debate on the criminal courts' practices on algorithmic risk assessment in terms of transparency, and then, after suggesting alternatives for the current perplexing position relating thereto, the paper concludes reaffirming that there is still a pressing need for interdisciplinary research priorities and harm minimization initiatives.

THE USE OF PREDICTIVE ALGORITHMS IN JUDICIAL SETTINGS

Predictive risk assessment algorithm has made a vivid shift in the criminal court while deciding upon who is likely to fail to appear at their court hearing, and who is likely to re-offend in the future',⁹ which is now widely understood as predictive justice.¹⁰ Such criminal justice predictive algorithms are also called 'risk assessments tools' or 'evidenced-based methods'.¹¹ Predictive tools are in widespread use across the criminal justice systems today, mostly in the USA, and an Asian country, Malaysia,¹² but the specific tools differ by country.¹³ However, in recent years, the field of risk assessment within the criminal justice system is facing wider scrutiny for lacking algorithmic transparency and accountability.

⁹Walter L Perry, et al., "Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations," (*Rand.org*, 2013) <https://www.rand.org/pubs/research_reports/RR233.html>.

¹⁰In the United States, for example, judges use software to assess a suspect's likelihood of reoffending. The term 'predictive' in AI jargon is contextually linked to the possibility of predicting future results through inductive analysis which identifies correlations between input data and output data. See Dr Jaspal Kaur Sadhu Singh, "Ethical Questions, Risks of Using AI in 'predictive Justice' | New Straits Times" (*NST Online*, February 16, 2020) <[https://www.nst.com.my/opinion/columnists/2020/02/565890/ethical-questions-risks-using-ai-predictive-justice%20\(last%20visited%20Oct%202020](https://www.nst.com.my/opinion/columnists/2020/02/565890/ethical-questions-risks-using-ai-predictive-justice%20(last%20visited%20Oct%202020)> accessed January 4, 2021.

¹¹ Sonja Starr, for example, refers to the incorporation of risk assessment into sentencing as "evidence-based sentencing" (EBS). Sonja B Starr, "Evidence-Based Sentencing and the Scientific Rationalization of Discrimination." (2014) 66 *Stanford Law Review* 803, 809.

¹²It is reported that Chief Justice for Sabah and Sarawak Tan Sri David Wong acclaimed the momentous occasion for the Malaysian judicial system to use such technology to recommend sentences based on previous cases that would be more consistent, and efficient. "We are in the process of doing it for the civil side — awarding damages for injuries in car accidents. See Julia Chan, "In a Local First, Sabah Court Gives out Sentence Assisted by AI, Malay Mail" (2020)

<<https://www.malaymail.com/news/malaysia/2020/02/19/in-a-local-first-sabah-court-gives-out-sentence-assisted-by-ai/1838906>> accessed September 16, 2020.

¹³Julia Chan, "In a Local First, Sabah Court Gives out Sentence Assisted by AI, Malay Mail" (2020) <<https://www.malaymail.com/news/malaysia/2020/02/19/in-a-local-first-sabah-court-gives-out-sentence-assisted-by-ai/1838906>> accessed September 16, 2020.

[The complaint is lack of consistency meaning that why is one sentence for a year while another is a fine only. Using the machine to analyze all data will achieve consistency; it speeds up the process a little because the magistrate doesn't have to look up previous cases while listening to submissions].

Apart from researchers' questions, Eric Loomis in a case of the Wisconsin Supreme Court, *Loomis v. Wisconsin* (2016), made a point on the absence of transparency in his risk assessment as a plea of defense, along with other concerns. He questioned the use in his sentencing decision of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), which is such decision support software that uses algorithms to assess the potential risk of recidivism, calculated by taking into account both an interview with the defendants and information from their criminal history. Such algorithms rely on plenty of risk factors; and predict the offenders' risk of recidivism, or risk of failing to appear when on bail, by way of scoring them as low, medium, or high. Such risk scoring is used while making the criminal court decisions on pre-trial detention, post-trial sentencing, and bond amounts or parole.¹⁴

However, it does not predict the specific likelihood of an individual offender's recidivism; rather it provides predictive outputs based on a comparison of information about the individual to a similar data group¹⁵ by algorithmic processing of the troves of data.¹⁶As aforesaid, Loomis in his case¹⁷ argued that he was not allowed to assess or the algorithm used in the COMPAS that prevents him from challenging the accuracy and scientific validity of the risk assessment; and thus, violates his constitutional right to due process for a trial court; but the latter court of late ruled against Loomis reasoning that knowledge of the algorithm's output was a sufficient level of transparency.¹⁸ Here goes simmering debate thereon for the case is that 'nowhere is the intersection of algorithms and human thinking more consequential than in the criminal justice system, where software predicts the likelihood defendants will

¹⁴As for predictive justice, there are nowadays reportedly 'more than 200 risk assessment tools available in criminal justice and forensic psychiatry, which are widely used to inform sentencing, parole decisions, and post-release monitoring. See Katalin Ligeti, "AIDP-IAPL International Congress of Penal Law: Artificial Intelligence and Criminal Justice" (2019) <http://www.penal.org/sites/default/files/Concept%20Paper_AI%20and%20Criminal%20Justice_Ligeti.pdf> accessed October 2, 2020.

¹⁵ Katalin Ligeti, "AIDP-IAPL International Congress of Penal Law: Artificial Intelligence and Criminal Justice" (2019) <http://www.penal.org/sites/default/files/Concept%20Paper_AI%20and%20Criminal%20Justice_Ligeti.pdf> accessed October 2, 2020.

¹⁶Robin A Smith, "Opening the Lid on Criminal Sentencing Software" (*Duke.edu*, July 19, 2017) <<https://today.duke.edu/2017/07/opening-lid-criminal-sentencing-software>> accessed November 2020, 20AD.

¹⁷881 N.W.2d 749 (Wis. 2016), cert. denied, 137 S.Ct. 2290 (2017); this was a Wisconsin Supreme Court case that was appealed to the United States Supreme Court but denied on June 26, 2017, See "*Loomis v. Wisconsin*, No. 16-6387 (U.S. Oct. 5, 2016)," (*SCOTUSblog*, 2017) <<https://www.scotusblog.com/case-files/cases/loomis-v-wisconsin/>> accessed September 18, 2020. [In February 2013, Eric Loomis was found driving a car that had been used in a shooting. He was arrested and found guilty. In determining his sentence, the trial judge looked at his criminal record as well as a score assigned by a tool called COMPAS. Loomis answered a series of questions that were then entered into Compas, a risk-assessment tool developed by a privately held company and used by the Wisconsin Department of Corrections. The case alleged that using such software in sentencing violates the defendant's right to due process because it prevents the defendant from challenging the scientific validity and accuracy of such test, and the system in question (COMPAS) violates due process rights by taking gender and race into account. The Supreme Court denied the writ of certiorari, thus declining to hear the case, on June 26, 2017.]

¹⁸See "*Loomis v. Wisconsin*, No. 16-6387 (U.S. Oct. 5, 2016)," (*SCOTUSblog*, 2017) <<https://www.scotusblog.com/case-files/cases/loomis-v-wisconsin/>> accessed September 18, 2020.

skip bail or commit other offenses.¹⁹ Hence, the opacity and incomprehensibility of predictive algorithms again gain sufficient momentum in search of accountability and fairness.

TRANSPARENCY IN PREDICTIVE ALGORITHMS USED IN THE CRIMINAL COURT:

As aforesaid, data science should provide transparency: to be clarified, indisputable, trustworthy, inferable, and interpretable; simply because it magically transforms data into value. As it is vividly argued that the journey from raw data to meaningful conclusions involves multiple steps and actors; the key factors for transparency thus encompasses the ideals of accountability, comprehensibility, and fairness.²⁰ Thus, there are found in recent time ever-increasing calls for bringing about meaningful transparency, explainability, accountability, and responsibility in the algorithms being used in public domains. Numerous studies show that such calls seemingly seem justified.

Now we may begin with demystifying the demand for transparency in AI algorithms. Transparency is a core social or constitutional value and a core part of systems of accountability for powerful actors.²¹ Most of today's machine learning algorithms operate as black boxes as is it not sufficiently understandable how and why the algorithms reach particular decisions, recommendations, or predictions. So it is time to see what does transparency means in our context, and then, what kind of transparency in algorithms is demanded for judicial use? Transparency is simply understood as openness of the datasets, and communication of the analyzed data and the mechanisms used by the models. Virginia Dignum (2019) in her book 'Responsible Artificial Intelligence' stated that algorithmic transparency as a principle demands that the factors influencing the decisions made by algorithms should be 'visible and transparent' to the people who use, regulate and are impacted by those algorithms.²² Kemper, J. and Kolkman, D. (2019) also put that transparency is the 'understandability of a specific model' found into open standards for government by way of the publication of datasets and other open-access schemes.²³

To demystify, Annany, M. and Crawford, K. (2016) clarify that transparency is thus not simply 'a precise end state in which everything is apparent,' but a system of observing and knowing that promises a form of

¹⁹Samuel Greengard, "Algorithms in the Courtroom," (*Communications of the ACM*, 2020) <<https://cacm.acm.org/news/244263-algorithms-in-the-courtroom/fulltext>> accessed September 19, 2020.

²⁰See Jakko Kemper and Daan Kolkman, "Transparent to Whom? No Algorithmic Accountability without a Critical Audience" (2019) 22 *Information, Communication & Society* 2081 <<https://doi.org/10.1080/1369118X.2018.1477967>>.

²¹"Explainable AI: The Basics POLICY BRIEFING" (2019) <<https://royalsociety.org/-/media/policy/projects/explainable-ai/AI-and-interpretability-policy-briefing.pdf>>.

²² Virginia Dignum, *Responsible Artificial Intelligence : How to Develop and Use AI in a Responsible Way* (2019) 59 .

²³Jakko Kemper and Daan Kolkman, "Transparent to Whom? No Algorithmic Accountability without a Critical Audience" (2019) 22 *Information, Communication & Society* 2081 <<https://doi.org/10.1080/1369118X.2018.1477967>>.

control.²⁴ From a predictive justice perspective, such demystification is considered instrumental for bringing in fairness and credibility in the court proceedings. So, when machine learning is used for assessing one's risk of recidivism, it is desirable that an open, apparent, and understandable dataset should be chosen and used in algorithmic models. In this context, since the current algorithms and machine learning techniques are in the need of improvement, improving algorithmic transparency would then be synonymous with advancing openness of the algorithm's source code as well as inputs and outputs that are used to make relevant algorithmic decisions.²⁵

The 2016 'FAIR Guiding Principles for scientific data management and stewardship' applies equally for both humans, and suggests that transparency is essential, for the importance of data accessibility and the need to contextualize data.²⁶ Although these guiding principles revolve around data management in academia, their impact extends well beyond the realm of universities, funding bodies, and publishers.²⁷ This stance is also applicable in the courtroom context. In the making of judicial decisions, more regulation may not bring transparency in algorithms. The improvement of algorithmic transparency may also not mean to have a quantitative increase of information, but higher information quality.

However, the available discourses on the very notion then reveal primarily three types of queries regarding algorithmic transparency; e.g. to know specifically *what* is predicted, and *how* and *why* is that predicted; and specific *to whom*; transparency. These categories in this connection *prima facie* justify a fair call for total transparency, that is, for the specifics of all phases to be opened up to the general public, *i.e.* total transparency. Paul B. de Laat (2017) pertinently posits that total transparency seems to be the perfect recipe for restoring accountability for algorithmic systems; for example, whenever parties and/or institutions have to be called to account, the raw data of the whole process are to be available for inspection.²⁸ However, Zarsky (2013) shed some light on the rebuttal of the total transparency argument in the consideration of privacy, perverse effects, competition, and inherent opacity.²⁹ Keeping that in mind, Paul B. de Laat (2017) concludes with the argument that only oversight bodies should enjoy full transparency.³⁰

²⁴M Annany and K Crawford, "Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability" (2016) 20 *New Media & Society* 973.

²⁵Bruno Lepri, et al., "Fair, Transparent, and Accountable Algorithmic Decision-Making Processes" (2017) 31 *Philosophy & Technology* 611.

²⁶In 2016, the 'FAIR Guiding Principles for scientific data management and stewardship' were published in *Scientific Data*. The authors intended to provide guidelines to improve the Findability, Accessibility, Interoperability, and Reuse of digital assets. The FAIR principles find their origin in a workshop held at the Lorentz Centre in Leiden, the Netherlands. See Wilkinson et al., "The FAIR Guiding Principles for Scientific Data Management and Stewardship" (2016) 3 *Scientific Data* <DOI-10.1038/sdata.2016.18> accessed October 21, 2019.

²⁷Jakko Kemper and Daan Kolkman, "Transparent to Whom? No Algorithmic Accountability without a Critical Audience" (2019) 22 *Information, Communication & Society* 2081 <<https://doi.org/10.1080/1369118X.2018.1477967>>.

²⁸Paul B. de Laat, "Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?" (2017) 31 *Philosophy & Technology* 525-41.

²⁹For more, see Tal Zarsky, "Transparent Predictions" (2013) 4 *University of Illinois Law Review* 1504.

³⁰Full transparency concerning all phases of decision-making includes that the institutions involved are open about how data have been obtained, make these data available to anyone interested, disclose the ranking algorithm in use (down to the very specifics), and specify whether decision-making (*i.e.*, the actual

Algorithmic sentencing is constitutionally, technically, and morally troubling;³¹ For example, persons assessed by COMPAS have little to no transparency about the process, the data quality, or how the data are being used in their cases. Opacity, the black-box effects in Machine Learning, is thus often stated as one of the main impediments for transparency in artificial intelligence systems. Algorithmic transparency is about how the algorithm learns a model from the data and what kind of relationships it can learn. Molnar (2019) argued that algorithm transparency only requires knowledge of the algorithm and not of the data or learned model.³² Deep learning approaches are less well understood and thus, considered less transparent.

Lack of clarity in the construction and production of algorithms and the ‘inner workings of these tools are largely hidden from public view’³³ are cardinal of them. They are unexplainable algorithmic tools; because ‘the engineer could not explain the software’s decision-making process’;³⁴ and are subject to property rights on algorithms (*e.g.* IP protection as trade secret); which hinders to inspection of either the algorithms or the training data sets that take out from government oversight and accountability. Because such algorithms are also proprietary in nature, they are not subject to state or federal open government laws in the USA, for example.³⁵

However, there are pronounced limits on the degree of transparency that can be attained. In the race of transparency debate, Annany and Crawford (2016) opined that transparency cannot be a characteristic of an algorithmic model. Rather the opacity of algorithms should be considered with a sensitivity for the contexts of their use; transparency is performed by socio-technical assemblages of algorithms and people.³⁶ Further, the justice systems are not an ordinary field, and the government is not an ordinary firm. Rather, every government incurs some legal obligations like respecting citizens’ rights and dignity, upholding equal protection, myriad court precedents and established jurisprudential principles. So, the choice of algorithms should not in defiance of the notions of equity, predictability, and transparency, which are ‘at the heart of what our justice system strives to provide’.³⁷

rankings that become public) proceeds in a semi-automated or fully automated fashion. See Paul B. de Laat, “Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?” (2017) 31 *Philosophy & Technology* 528.

³¹Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66:803 *Stan. L. Rev.*, 870, (2014).

³²Christoph Molnar, *Interpretable Machine Learning : A Guide for Making Black Box Models Interpretable* (Lulu 2019) <<https://christophm.github.io/interpretable-ml-book/index.html>> accessed December 25, 2020.

³³Electronic Privacy Information Center, “EPIC - Algorithms in the Criminal Justice System: Pre-Trial Risk Assessment Tools” (*Epic.org*, 2014) <<https://epic.org/algorithmic-transparency/crim-justice/>>.

³⁴Jason Tashea, “Courts Are Using AI to Sentence Criminals. That Must Stop Now,” (*Wired*, April 17, 2017) <<https://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/>> accessed September 18, 2020.

³⁵Electronic Privacy Information Center, “EPIC - Algorithms in the Criminal Justice System: Pre-Trial Risk Assessment Tools” (*Epic.org*, 2014) <<https://epic.org/algorithmic-transparency/crim-justice/>>.

³⁶M Annany and K Crawford, “Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability” (2016) 20 *New Media & Society* 973.

³⁷See Caleb Watney, “When It Comes to Criminal Justice AI, We Need Transparency and Accountability” (*R Street*, December 1, 2017) <<https://www.rstreet.org/2017/12/01/when-it-comes-to-criminal-justice-ai-we-need-transparency-and-accountability/>> accessed January 3, 2021.

Machine-learning techniques, artificial neural networks, in particular, are a form of the machine-learning algorithm with a structure roughly based on that of the human brain. At least as far as the present context and time are concerned, it is the case that neural networks, a deep learning algorithm³⁸ meant to act like the human brain, cannot be transparent because of their very nature. Like other kinds of machine-learning algorithms, they can solve problems on a trial and error basis without being explicitly programmed with rules to follow.³⁹ So lack of clarity about how algorithms exercise their power over humans as they are largely black-boxed and beyond query or question.

As such, it is not a new claim that the stakes in the criminal justice system are too high to blindly trust in black-box algorithms that haven't been properly tested against available alternatives.⁴⁰ So if you don't know exactly how the proprietary algorithm works; it is the fact is that you're in prison, but you don't know why, and even you can't argue. In fact, you only know the assessment score it spits out, which judges may consider at sentencing.⁴¹ Consequently, One strand of criticism demands that they should be built from publicly available data sets using open-source software; and disclose the details of their algorithms, so that anyone can inspect the data fed into them. Accordingly, risk assessments should be impermissible unless both parties get to see all the used data. It should be an open, full-court adversarial proceeding.⁴²

Other than that, the public authorities like police or court typically do not write their own algorithms; they buy them from private sectors, such as Google, Amazon, Microsoft, and the like. That's being the fact; the algorithmic decision-making in the public sector has 'black-box' effects. Accordingly, this process is hidden and always changing, not explicitly programmed, which runs the risk of limiting a judge's ability to render a fully informed decision, and a defendant's ability to adequately defend his case in an explicit way. Hence, one might easily place his eye-brow towards the opacity of the AI technology that put a hindrance to ensure fairness in the court process; and then, the rise of algorithm machines as new forms of algorithmic power that are reshaping how the court works.

Keeping that in mind, the critics poignantly opined that since the methodology behind the assessment algorithms is a trade secret,⁴³ and their

³⁸Deep Learning is a sub-field of machine learning in Artificial intelligence (A.I.) that deals with algorithms inspired by the biological structure and functioning of a brain to aid machines with intelligence. Jojo John Moolayil, "A Layman's Guide to Deep Neural Networks" (*Medium*, July 24, 2019) <<https://towardsdatascience.com/a-laymans-guide-to-deep-neural-networks-ddcea24847fb>>

³⁹"Neural Networks, Explained" (*Physics World*, July 9, 2018) <<https://physicsworld.com/a/neural-networks-explained/>>.

⁴⁰Robin A Smith, "Opening the Lid on Criminal Sentencing Software" (*Duke.edu*, July 19, 2017) <<https://today.duke.edu/2017/07/opening-lid-criminal-sentencing-software>> accessed November 2020, 20AD.

⁴¹Ellora Thadaney Israni, "Opinion | When an Algorithm Helps Send You to Prison" *The New York Times* (October 26, 2017) <<https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html>> accessed July 23, 2020.

⁴²Julia Angwin, et al, "Machine Bias" (*ProPublica*, May 23, 2016) <<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>> accessed December 12, 2020.

⁴³Wexler Rebecca, "Opinion | When a Computer Program Keeps You in Jail," (*The New York Times*, June 13, 2017) <<https://www.nytimes.com/2017/06/13/opinion/how-computers-are-harming-criminal-justice.html>> accessed September 18, 2020.

proprietary nature prevents the disclosure of how risk scores are calculated; meaning that they cannot be examined by the public and the aggrieved persons, which may be a violation of the due process of law and a deviation from the principle of fairness in the court proceeding. So corporate secrecy offers is no public accountability, no transparency, and no oversight of how AI systems operate; which tantamount to precluding scrutiny of how the systems work or how they could go haywire.⁴⁴

Even, 'how these algorithms change the way justice is administered is largely unknown'.⁴⁵ In addition, the lack of methodological transparency from data assessment companies has made the study difficult.⁴⁶ Consequently, courts cannot evaluate how the risk scores are determined or how the factors are weighed. So, seeing something in an algorithm may lead to control over it, and liberal democracy's openness may ultimately create security. Hence, any information used in the court is required to be made easily discernible and legible; so that competent audiences involved would be able to comprehend that information. In this way, as Mackenzie (2008) puts it, transparency becomes performative: it does work, casts systems as knowable and, by articulating inner workings, it produces understanding.⁴⁷

Considering such aspects, that tendency also seems to be troubling in private sector contexts, as commercial torts may be committed with impunity thanks to the opacity of ranking and rating systems. Even in the context of voting, authorities have been sluggish about demanding software that is auditable and understandable by outsiders. In view of this, the task of criminal sentencing should be a bridge too far for conscientious judges, not for a machine. This is why; 'sending someone to jail thanks to the inexplicable, unchallengeable judgments of a secret computer program is too 'black mirror' for even hardened defenders of corporate privileges'.⁴⁸

PATHWAYS TO TAKING OUT BLACK BOX EFFECTS

A report of the Council of Europe on Artificial Intelligence (2018) reiterates that all involved stakeholders should promote a culture of transparency enshrining the disclosure of the used AI applications, a description of their logic as well as the access to the structure of algorithms and to the introduced datasets.⁴⁹ Again, on the 'black box' puzzle, Virginia Dignum (2019) artfully suggests that removing the algorithmic black box by means of

⁴⁴Wojciech Samek, Thomas Wiegand and Klaus-Robert Müller, "Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models" [2017] ITU Journal: ICT Discoveries, Special Issue No. 1 <<https://arxiv.org/abs/1708.08296>> accessed September 9, 2020.

⁴⁵TOM SIMONITE, "Algorithms Should've Made Courts More Fair. What Went Wrong?" (*Wired*, 2019) <<https://www.wired.com/story/algorithms-shouldve-made-courts-more-fair-what-went-wrong/>> accessed September 19, 2020.

⁴⁶Julia Angwin, et al, "Machine Bias" (*ProPublica*, May 23, 2016) <<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>> accessed December 12, 2020.

⁴⁷Donald Mackenzie, "An Engine, Not a Camera How Financial Models Shape Markets" (2008) <<https://uberty.org/wp-content/uploads/2015/02/MacKenzie-An-Engine-Not-a-Camera.pdf>> accessed December 16, 2020.

⁴⁸Frank Pasquale, "Secret Algorithms Threaten the Rule of Law," (*MIT Technology Review*, July 1, 2017) <<https://www.technologyreview.com/2017/06/01/151447/secret-algorithms-threaten-the-rule-of-law/>> accessed September 18, 2020.

⁴⁹Council of Europe, Report on Artificial Intelligence, TPD (2018) 09 Rev, Strasbourg, 3 December 2018.

providing meaningful means to inspect and evaluate the algorithms used would not suffice.⁵⁰

Rather, ‘transparency may then be better served by ‘openness and control’ over the ‘whole learning and training process’ than by just removing the algorithmic black box. Trust in the system will improve if we can ensure openness of affairs in all that is related to the system, she adds.⁵¹ Her arguments are practically justifiable in the sense that only the inspection and evaluation of the algorithm in use may infringe property rights and business models of those who develop the algorithms, and also the ‘code’ used therein would not make much sense to most users.

Openness narrative demands that the used training datasets must be open for everyone. If a training set has a bias, let people know those things. Dignum (2019) thus underscores the observance of ethical principles and putting human values at the core of system design. She also articulates the need for rethinking the optimization criteria for machine learning; and calls for researchers and developers’ mind-shift towards improving transparency rather than performance, leading to a new generation of algorithms that can turn machine learning into valuable learning.⁵²

As aforesaid, in the regulatory sense, Paul B. de Laat (2017) arguably put an alternative by positing that ‘full transparency for oversight bodies alone is the only feasible option; extending it to the public at large is normally not advisable. This type of full transparency is to be observed in data collection (*e.g.*, whether data are appropriate to the questions being asked or whether the datasets are free of bias), model construction (whether an appropriate model gets constructed that best fits the data), and model use.⁵³ Then, if any objection to transparency (four kinds of objections- privacy, perverse effects, competition, and inherent opacity) is raised, give an opportunity of being heard in furtherance of the golden principles of natural justice.

Mazzotti (2017) also advocates for a standard of transparency to elucidate the technical choice of certain expert groups; but (Kemper and Kolkman, 2019) counters him putting that ‘the reality of algorithms today is that the precise operations and potential ramifications of an assemblage of algorithms remain obscure even for these experts; rather underscores the notion of transparency and approaches it from a glitch studies perspective; and demonstrates that transparency of algorithms can only be attained by virtue of an interested-critical audience.⁵⁴ They justified arguing that even if an algorithmic model is made entirely transparent, not all of its potential effects and faculties can be inferred; and criticize those concepts by drawing on glitch

⁵⁰Virginia Dignum, *Responsible Artificial Intelligence : How to Develop and Use AI in a Responsible Way* (2019) 59 59.

⁵¹See also Virginia Dignum, “On Bias, Black-Boxes and the Quest for Transparency in AI,” (*Delft Design for Values Institute*, January 26, 2018) <<https://www.delftdesignforvalues.nl/2018/on-bias-black-boxes-and-the-quest-for-transparency-in-artificial-intelligence/>> accessed April 22, 2020.

⁵²Virginia Dignum, *Responsible Artificial Intelligence : How to Develop and Use AI in a Responsible Way* (2019) 59; Virginia Dignum, “On Bias, Black-Boxes and the Quest for Transparency in AI,” (*Delft Design for Values Institute*, January 26, 2018) <<https://www.delftdesignforvalues.nl/2018/on-bias-black-boxes-and-the-quest-for-transparency-in-artificial-intelligence/>> accessed April 22, 2020.

⁵³ Paul B. de Laat, “Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?” (2017) 31 *Philosophy & Technology* 525-41.

⁵⁴Jakko Kemper and Daan Kolkman, “Transparent to Whom? No Algorithmic Accountability without a Critical Audience” (2019) 22 *Information, Communication & Society* 2081 <<https://doi.org/10.1080/1369118X.2018.1477967>>.

studies and introducing the dimensions of speed, critical audiences, and complexity and irreducibility.⁵⁵

Moreover, so far as courtrooms' decisions are in context, it too seems to be relational that only transparency requirement is not viable, we reasonably need for trustworthiness. It is likewise argued that 'trustworthiness demands transparency, but not just 'fishbowl' transparency in which huge amounts of information are provided in indigestible form; it also does not necessarily provide explainability'.⁵⁶ Rather, a truly trustworthy algorithm does require honesty, competence, and reliability; which is called Onora O'Neill's idea of 'intelligent transparency'- meaning that information should be intelligible, accessible, usable, and assessable.⁵⁷ Hence, it will not be verbose to opine that such trustworthy algorithms seem to be judicially befitting.

However, since transparency works as a key enabler of accountability, as it is also argued that, only transparency and auditing do not necessarily suffice for accountability.⁵⁸ Independent certification may then potentially be a solution for bringing about the AI system's trustworthiness, reliability and transparency. In this context, principle-4 of the CEPEJ Ethical Charter (transparency, impartiality, and fairness) is worth mentioning, which mandates for external audits and data processing methods' accessibility and understandability.⁵⁹

Furthermore, while advocating for verifying and making transparent the behavior of an algorithmic decision-making process, Bruno Lepri, *et al.* (2017) opined that an auditing strategy deals with the decision process as a black box, whose inputs and outputs are visible, while inner workings are not'.⁶⁰ Hence, in the certification process, 'independent and trusted institutions would be able to validate and test algorithms, applications, and products against a set of well-defined principles, and guarantee the quality of the system'.⁶¹

In this connection, Završnik (2019) supports certification and auditing schemes by the third-party audits; for it may also increase trust in algorithms in terms of their fairness.⁶² Applying this in the context of judicial decision-making, the judicial institution in a given jurisdiction would then have the choice of what type of system would best meet the institutional requirements.

⁵⁵Jakko Kemper and Daan Kolkman, "Transparent to Whom? No Algorithmic Accountability without a Critical Audience" (2019) 22 *Information, Communication & Society* 2081 <<https://doi.org/10.1080/1369118X.2018.1477967>>.

⁵⁶David Spiegelhalter, "Should We Trust Algorithms?" (2020) 2 *Harvard Data Science Review* 1.

⁵⁷"Science as an Open Enterprise | Royal Society" (royalsociety.org) <<https://royalsociety.org/topics-policy/projects/science-public-enterprise/>> accessed January 5, 2021; and also see "Explainable AI: The Basics POLICY BRIEFING" (2019) <<https://royalsociety.org/-/media/policy/projects/explainable-ai/AI-and-interpretability-policy-briefing.pdf>>..

⁵⁸Bruno Lepri, et al., "Fair, Transparent, and Accountable Algorithmic Decision-Making Processes" (2017) 31 *Philosophy & Technology* 611-27.

⁵⁹The European Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe has adopted the first European Ethical Charter on the use of artificial intelligence in judicial systems. This first European text sets out ethical principles relating to the use of artificial intelligence (AI) in judicial systems. See European Commission for the Efficiency of Justice (CEPEJ), "CEPEJ European Ethical Charter on the Use of Artificial Intelligence (AI) in Judicial Systems and Their Environment" (*European Commission for the Efficiency of Justice (CEPEJ)*, 2018) <<https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment>>.

⁶⁰Bruno Lepri, et al., "Fair, Transparent, and Accountable Algorithmic Decision-Making Processes" (2017) 31 *Philosophy & Technology* 611-27.

⁶¹Virginia Dignum, *Responsible Artificial Intelligence : How to Develop and Use AI in a Responsible Way* (2019) 98.

⁶²Aleš Završnik, "Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings" [2019] *European Journal of Criminology* 1.

Alternatively, an expert witness could also be appointed by the judge to verify either the algorithmic process, or the neural network of a certain AI system, whenever the parties to the proceeding express their doubt about the propriety, fairness, and correctness of the automated data.⁶³ Additionally, just as judges are compelled to explain not only their decisions but also their reasoning via published opinions, programmers must shed some light on their contributions too.⁶⁴ So programmers should not be allowed to hide behind a veil of ignorance.

CONCLUSION:

A judicial perspective suggests a revolution in algorithms and AI systems for being observed from the viewpoint of the law, judicial principles, individual rights, not only from the algorithm itself or those who build algorithms. Since the current version of predictive algorithms in use remains black boxes and more consequential in the justice context, it would not be quintessential to use them in judicial forums. Considering that the justice system manifests a better path for uprooting traditional flaws and unfairness and ensuring accountability, responsibility, and public trust in the decision-making process, it becomes indispensable for a new genre of transparent algorithms. The increased stakeholders' involvement, public investment, and more research in policy space and system designs could produce far better outcomes than the status quo. When such algorithms become ready for flourishing in the judicial precincts, 'full transparency' within an open technological ecosystem may be attributed to only the oversight bodies; instead of extending the same to the public at large. In addition, complete dehumanization at the decision-making loop of the judiciary also appears to be not a viable option. Instead, human oversight and judge's discretion should always get the priority in the judicial context.

⁶³Katalin Ligeti, "AIDP-IAPL International Congress of Penal Law: Artificial Intelligence and Criminal Justice" (2019)

<http://www.penal.org/sites/default/files/Concept%20Paper_AI%20and%20Criminal%20Justice_Ligeti.pdf> accessed October 2, 2020.

⁶⁴See Ellora Israni and Evelyn Chang (ed.), "Algorithmic Due Process: Mistaken Accountability and Attribution in State v. Loomis," (*Harvard Journal of Law & Technology*, August 31, 2017)

<<https://jolt.law.harvard.edu/digest/algorithmic-due-process-mistaken-accountability-and-attribution-in-state-v-loomis-1>> accessed March 23, 2020.