Biased Algorithmic Risk Assessment in criminal justice settings: How is COMPAS fraying the fabric of the right to a human decision-making in criminal procedure law

Avaliação algorítmica discriminatória em processo penal: como o COMPAS está a erodir as bases fundacionais do direito a uma decisão humana em processo penal

Hugo Luz dos Santos[1]
City University of Macau, Macau, China
hugo.miguel.luz@gmail.com
http://orcid.org/0000-0003-0297-3546

Abstract: Artificial Intelligence-embedded technologies represent a blazingly new path, a foray into uncharted territory that holds a wealth, and a treasure trove, of scorching challenges that should not leave anyone lukewarm. When it comes to cutting-edge technology, you better not try your luck and the let the chips fall where they may. Artificial Intelligence-embedded technology, albeit a seemingly bedazzling gift to humankind, is not without multifarious caveats. Amongst which stands a slew of perceived perils arising out of biased algorithm risk assessment in criminal justice settings that is shaping up to topple the foundations upon which stands algorithmic fairness. Hardly any bewildering surprise stems from the fact that Artificial Intelligence-embedded technology holds the blueprint of our destiny and the disruptive power it carries. This is about the unfolding of newer and higher forms of intrusiveness in our daily lives that extend beyond our collective grasp. Building upon this, a couple of vexing questions crop up with bursting unease: should relevant stakeholders (i.e. judges, prosecutors, and lawyers) rely

[1] PhD in Law and University Professor at City University of Macau, China/Fellow of the Royal Society of Arts of the United Kingdom "in recognition of his outstanding contributions to the field of justice, rule of law and policy worldwide".

upon algorithmic risk assessment (i.e. COMPAS) in the purview of criminal justice? On the flip side, should machines be entrusted with the task of making high-stake decisions in the realm of criminal procedure law settings? Bearing these burning (research) questions very firmly in mind, this paper adamantly contends that machines should not wield the power of making high-stake decisions that may bear heavily on citizens' fundamental rights (i.e. the right to a due process, a central tenet of criminal procedure law) or inordinate harms may otherwise occur (i.e. the loss of functional reputation of the system of administration of justice).

Keywords: Artificial Intelligence-embedded technologies; COMPAS; biased algorithmic risk assessment; algorithmic fairness; criminal procedure law; machine learning algorithms; right to a human decision-making; predictive parity; robot-judges.

Resumo: *A inteligência artificial representa um caminho inteiramente novo, uma incursão fulgurante em territórios nunca antes mapeados, o que, em si mesmo tomado, encerra uma pletora de candentes desafios que não deverá deixar ninguém indiferente. No que tange a tecnologia de ponta, não se deve confiar plenamente na sorte – sequer no bambúrrio. A inteligência artificial, apesar das suas múltiplas vantagens, não está isenta de uma miríade de preocupantes inconsistências e de inultrapassáveis insuficiências. De entre as quais se respiga a discriminação algorítmica que resulta da utilização de algoritmos preditivos discriminatórios no âmbito do processo penal, o que poderá derruir as bases fundacionais em que se estriba a justiça algorítmica. Pode afirmar-se, sem surpresa de monta, que a Inteligência Artificial avoca, nas suas mãos tremeluzentes, o destino da humanidade, atento o potencial disruptivo que encerra. Está fundamentalmente em causa a emergência de novas – e profundamente disruptivas – formas de intrusão nas nossas vidas quotidianas que são, em si mesmo tomadas, insuscetíveis de ser acurada e proficientemente antecipadas no momento atual que nos interpela. É com este pano de fundo que florescem as candentes questões: deverão os operadores judiciários (i.e., juízes, procuradores e advogados) confiar plenamente na avaliação algorítmica do risco em processo penal? Por outro lado, deveremos confiar à Inteligência Artificial o múnus de realizar decisões que podem afectar seriamente o núcleo essencial dos direitos fundamentais dos cidadãos? Tendo em linha de conta as aporias acima assinaladas, este texto conclui que à inteligência artificial não deverá atribuído o poder de decidir questões de supina importância em processo penal (i.e. o direito a um processo justo*

e equitativo, um esteio essencial do processo penal moderno), sob pena de consequências danosas irreversíveis (e.g. a perda de reputação funcional do sistema de administração de justiça penal).

Palavras-chave: *Inteligência Artificial; COMPAS; avaliação algorítmica do risco enviesada; justiça algorítmica; processo penal; algoritmos; direito a uma decisão judicial emitida por um juiz; predição paritária; robôs-juízes.*

1. Introduction

Algorithmic risk assessment has garnered widespread attention[2]. Algorithmic risk assessment has accordingly been received as a lightning bolt of great significance in the purview of criminal justice[3]. Which caused algorithmic decision-making to swell[4]. It is true that algorithmic

---

2 See: Slobogin, Christopher, "Principles of Risk Assessment: Sentencing and Policing", *Ohio State Journal of Criminal Law*, Volume 15, 2018, pp. 583 and ff and passim (setting out foundational principles for how risk-assessment tools should be devised and deployed, specifically in the sentencing compass). See also: Mayson, Sandra G., "Bias In, Bias Out", *Yale Law Journal*, Volume 128, 2019, pp. 2218-2300. See also: Xi Chen, "Algorithmic proxy discrimination and its regulations", *Computer Law & Security Review*, *The International Journal of Technology Law and Practice*, Volume 54, September 2024, 2024, pp. 1-13.

3 Mayson, Sandra G., "Dangerous Defendants", *Yale Law Journal*, Volume 127, 2018, p. 490 ("Bail reform is gaining momentum nationwide."); Stevenson, Megan, "Assessing Risk Assessment in Action", *Minnesota Law Review*, Volume 103, 2018, p. 303 (discussing the burgeoning use of pretrial risk assessment as a mandatory component of bail decisions in Kentucky).

4 Such a set of algorithmic tools - which are currently being deployed in predicting criminal recidivism - are gaining traction and gathering pace in Italy: Gialuz, M., "Quando la giustizia penale encontra l´intelligenza artificiale: luci e e ombre di risk assessment tools tra Stati Uniti ed Europa", 2019, available at: www.penalecontemporaneo.it.d/6702-quando-la-giustizia-penale-encontra-l´intelligenz-artificiale-luci-e-ombre-di-risk-assessment-tools (access: 03.07.2025) and United Kingdom: Peeters, R./Schuilenburg, M., "Machine Justice: Governing Security through Bureaucracy of Algorithms", *Information Polity*, Volume 23, 2018, pp. 267-274; Hamilton, Melissa, "Sentencing Disparities", *British Journal of American Legal Studies*, Volume 6, Issue 2, 2017, pp. 181 ff.

risk assessment[5] has aroused heated controversy lately[6], but truer still police, prosecutors, judges, and other criminal justice stakeholders are increasingly relying upon algorithmic risk assessment to predict[7] if a defendant will engage in flagitious activities in the forthcoming future[8]. The vexing question that this essay seeks to adroitly address, and deftly tackle, is: should they?

These deep-rooted concerns are neither unwarranted, nor uncalled for, nor are they statistically unsound[9]. It hardly amounts to a bewildering finding that these cutting-edge tools are not without

---

5 See: Brennan, T., Dieterich, W., & Ehret, B., "Evaluating the predictive validity of the COMPAS risk and needs assessment system", *Criminal Justice and Behavior*, Volume 36, Issue 1, 2009, pp. 21-40.

6 But see: Eaglin, Jessica M., "Constructing Recidivism Risk", *Emory Law Journal*, Volume 67, 59, 2017, 61 n.1 (arguing that "[p]redictive technologies are spreading through the criminal justice system like wildfire"). None of this is to suggest, nor is to say, that actuarial risk assessment is blazingly new to criminal justice system. Pivotally, parole boards have long relied on risk-assessment instruments since the 1930s. By the same token, some jurisdictions used algorithms to forecast certain types of crimes, such as sex offenses, for decades past. See: Mason, Sandra G., "Bias in, Bias Out.", *cit*., pp. 2219-2298.

7 Zhu, Mirilla, "Jury, Using Artificial Intelligence to Predict Recidivism Rates", *Yale Scientific*, 2020, passim. See also: Chasemi et al, Medhi, "The application of Machine Learning to a General Risk-Need Assessment Instrument in the Prediction of Criminal Recidivism", *Criminal Justice and Behavior*, Volume 48, 2020, pp. 518-538.

8 See: Skeem, J. and C. Lowenkamp, "Using Algorithms to Address Trade-Offs Inherent in Predicting Recidivism", *Behavioural Science and Law*, Volume 38, 2020, pp. 259-278 (on the use of machine learning algorithms to forecast recidivism in criminal justice settings). But see: Huq, Aziz Z., "Racial Equity in Algorithmic Criminal Justice", *Duke Law Journal*, Volume 68, 2019, pp. 1043, 1053, 1083-1102 (sheds a keen light on why constitutional law is fundamentally bereft of "creditable guidance" for granting racial equity in algorithmic risk assessment in criminal justice).

9 No surprise stems from the fact that prejudiced algorithms will feed *justice prédictive* - and that is something quite unsavoury and unpalatable in criminal justice indeed. See: Garapon, Antoine, "Les enjeux de la justice predictive", *Juris-Classeur périodique*, Volume 31, 2018, passim. See also : Larret-Chahine, Louis, "Le droit isométrique: un nouveau paradigme juridique né de la justice prédictive », *Archives de Philosophie du Droit*, Volume 60, 2018, pp. 287 and ff and passim.

humongous drawbacks[10]. Compellingly, artificial intelligence-embedded technologies – which forms part of a broader technological apparatus within which algorithm risk assessment operates[11] – often evince the devilishly harmful virus of algorithmic bias[12], which, I may venture to add, is wholly unbecoming in the purview of criminal procedure law.

To further compound our functional woes and mental throes, machine learning algorithms[13], if nothing else, tend to be beset by racial

---

[10] See: Chandler, Anupam, "The Racist Algorithm", *Michigan Law Review*, Volume 115, n.º 6, 2018, pp. 1023-1025 (arguing that the real-world facts on which algorithms used in criminal justice risk assessment are based are "deeply suffused with invidious discrimination"); Eaglin, Jessica M., "Constructing Recidivism Risk", *cit*., pp. 94-99 (argues that risk assessment has every chance to "compromise[e] equality"); Kim, Pauline T., "Data-Driven Discrimination at Work", *William and Mary Law Review*, Volume 58, 2017, pp. 857, 863-64 (shedding a keen light on the racial effects of algorithmic prediction in the remit of employment).

[11] Berk, Richard A., *Machine Learning Forecasts of Risk in Criminal Justice Settings*. New York: Springer, 2018 (broaches extensively algorithmic risk assessment in criminal justice settings).

[12] See: Baeza-Yates, R., "Bias on the web", *Communications of the ACM*, 61 (6), June 2018, 2018, pp. 54-61 (argues that "*algorithmic bias is when the bias is not present in the input data* and is added *purely by the algorithm*"). Italics added. See also: Hajian, Sara Hajian et al., "Algorithmic Bias: From Discrimination Discovery to Fairness-aware Data Mining", *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Volume 22, 2125, 2016; Griemmelmann, James/Westreich, Daniel, "Incomprehensible Discrimination ", *California Law Review Online*, Volume 7, 2018, pp. 164-171; Xiang, Alice, "Reconciling Legal and Technical Approaches to Algorithmic Bias", *Tennessee Law Review*, Volume 88, 2021, passim; See: Lemley, Mark A./Casey, Bryan, "Remedies for Robots", *The University of Chicago Law Review*, Volume 86, Issue 5, 2019, pp. 1311-1396 (1376) (alerting to the perils of algorithmic bias); Jackson, Maya C., "Artificial Intelligence and Algorithmic Bias: The Issues with Technology Reflecting History and Humans", *Journal of Business Technology Law*, Volume 16, 2021, pp. 299-316; West, Sarah Myers/Whittaker, Meredith/Crawford, Kate, *Discriminating Systems. Gender, Race and Power in AI*. New York: AI Now Institute, 2019, pp. 15 and ff and passim; Waldman, Ezra, "Privacy, Practice, and Performance", *California Law Review*, Volume 110, 2021, passim (discussing the central tenets of algorithmic bias).

[13] See: Murphy, Kevin P., *Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Learning series)*. Massachusetts: The MIT Press, 2012, pp. 1-1104 (states avowedly that "machine Learning attempts to combine deductive, inductive, and abductive reasoning to make person-like reasoning in AI Entities possible."). See also: Hutto-Schultz, Jess, "Dicitur Ex

skew[14], thereby exacting a toll on members of racialized communities[15]. Amongst whom, members of communities of colour that have been grappling with harrowing racial discrimination[16] for a long-winded span of time[17].

To forestall the gruelling effects of both racial discrimination and algorithmic unfairness[18], scholars have been neither coy nor bashful in voicing their deep-seated concerns about algorithmic injustice[19] facing members of communities of colour. An ensemble of strategies of resistance against racial discrimination have been accordingly devised: (I) the exclusion of input markers closely associated with race (colour-blindness[20]);

Machina: Artificial Intelligence and the Hearsay Rule", *George Mason Law Review*, Volume 27, Issue 2, 2020, pp. 684-718 (689). See: Chollet, François, *Deep Learning with Python*. Shelter Island, Nova Iorque: Manning, 2018, pp. 1-384 (passim). See also: Surden, Harry, "Machine Learning and Law", *Washington Law Review*, Volume 89, 2015, pp. 87 and ff. See also: Barocas, Solon/ Selbst, Andrew, "Big Data's Disparate Impact", *California Law Review*, Volume 104, pp. 2016, pp. 671-678.

14 Mayson, Sandra G., "Bias In, Bias Out", *Yale Law Journal*, Volume 128, 2019, pp. 2218-2300.

15 Rucker, J. M. and J. A. Rocheson, "Toward an Understanding of Structural Racism: Implications for Criminal Justice." *Science*, 374 (6565), 2021, pp. 286-90.

16 See: Feagin, J. R., *Systemic racism: a theory of oppression*. New York: Routledge, 2006. See also: Fiske, John, "Surveilling the City: Whiteness, the Black Man and Democratic Totalitarianism.", *Theory, Culture and Society*, 15 (2), 1998, pp. 67–88 (on the overhyped surveillance on black people as a byproduct of systemic racism).

17 Milligan, Joy, "Protecting Disfavored Minorities: Towards Institutional Realism", *UCLA Law Review*, Volume 63, 2016, pp. 918-921 (delves into the racial inequality in United States of America´s Department of Agriculture farm aid to both Whites and Blacks).

18 Berk, R. A., H. Heirdari, S. Jabbari, M. Kearns, and A. Roth, "Fairness in Criminal Justice Risk Assessments: The State of the Art." *Sociological Methods and Research*, 2018.

19 Ting-an Lin, "Democratizing AI" and the Concern of Algorithmic Injustice", *Philosophy and Technology*, Volume 37 (103), 2024, pp. 3-27; Okidegbe, Ngozi, "The Democratizing Potential of Algorithms?", *Connecticut Law Review*, Volume 54, 2021, passim; Katyal, Sonia, "Private Accountability in the Age of Artificial Intelligence", *UCLA Law Review*, Volume 66, 2019, pp. 54-99 (on the fundamental tenets of algorithmic injustice coupled with a host of measures to stymie it).

20 But see: Kleinberg, Jon et al., "Algorithmic Fairness", *AEA Papers & Proceedings*, Volume 108, 2018, pp. 22-23 (discussing, anchored in a batch of national

(II) adjustments[21] to computer-run algorithmic design to equalize forecasts across racial lines; and (III) rejection of algorithmic methods altogether[22]. This section´s main contention is that which these strategies are at best shallow, hollow, nomadic, and superficial and at worst run counter to algorithmic fairness[23] insofar the origin of racial inequity in algorithmic risk assessment lies neither in the input data[24], nor in a set of far-fetched algorithms[25], nor in algorithmic methodology itself[26].

---

data, that including race as an input variable to a machine-learning college-admissions algorithm both "improves predicted GPAs of admitted students" and can augment "the fraction of admitted students who are black").

[21] It is important to note that an algorithm designed for maximum accuracy will adjust itself to the majority data, and may be less accurate for members of the underrepresented group, thus breeding inductive bias. See: Shellenbarger, Sue, "A Crucial Step for Avoiding AI Disasters", *Wall Street Journal*, (Feb. 13, 2019, 9:57 AM ET), https://www .wsj.com/articles/a-crucial-step-for-avoiding-ai-disasters-11550069865?ns=prod/accounts -wsj [https://perma.cc/C28U-LAAE] (discussing this issue while emphasizing how diverse development teams pay more attention to unrepresentative and underrepresented data sets). Human programmers can circumvent and bypass this problem by weighting the minority-group data more heavily, by designing different algorithms for each racial group, or by striving to contain more data to equalize group representation in the data set. See also: Hamilton, Melissa, "The Biased Algorithm: Evidence of Disparate Impact on Hispanics", *American Criminal Law Review*, Volume 56, 2019, pp. 1553 and ff and passim (which clearly demonstrates that COMPAS was significantly less accurate for Hispanic than for white defendants by several measures and suggesting that smaller numbers of Hispanic defendants might be the cause). See also: Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[22] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[23] Hellman, Deborah, "Measuring Algorithmic Fairness", *Virginia Law Review*, 106 (4), 2020, pp. 811-66 (discusses the building blocks upon which stands algorithm fairness while emphasizing the roadblocks that stand in its way).

[24] Gillis, Talia B., "The Input Data Fallacy", *Minnesota Law Review*, Volume 106, 2022, pp. 1176-1263.

[25] Note, though, that there are far more unsettling issues with which criminal justice flounders. To begin with, the way inscrutable algorithms make their correlations and draw their inferences accordingly. This is coined as *black box*, that further enhances the shroud of mystery under which algorithms are cloaked. See: Bathaee, Yavar, "The Artificial Intelligence Black Box and the Failure of Intent and Causation", *Harvard Journal of Law and Technology*, n.º 37, 2018, p. 901.

[26] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

On the opposite pole, a deep dive into the remit of algorithmic decision-making[27] shows that the real problem lies elsewhere: the very nature of prediction itself. Compellingly, at its core, prediction[28] is backward-looking in that looks to the past in an earnest attempt to forecast the future[29]. Dispiritingly, though, we live in a racially skewed world against the backdrop of which predictions are slated to project racial inequalities of the past[30] onto the forthcoming future[31].

---

[27] See in Brazilian doctrine: Alves, Jones Figueirêdo/Pimentel, Alexandre Freire, "Breves notas sobre os preconceitos decisionais judiciais produzidos por redes neurais artificiais (Brief notes about the judicial decisional prejudices produced by artificial neural networks), *Lisbon Law Review*, Thematic Issue: Law and Technology, Year LXII, Numbers 1 and 2, 2022, pp. 555-577.

[28] Aggarwal, C. C., *Neural networks and deep learning*, Heidelberg, Springer, 2018, passim; Agrawal, A., Gans, J., & Goldfarb, A., "Prediction machines (updated and expanded): The simple economics of artificial intelligence", *Harvard Business Review Press*, 2022; Agrawal, A., Gans, J., & Goldfarb, A., "Power and prediction: The disruptive economics of artificial intelligence", *Harvard Business Review Press*, 2022; Bengio, Y., Hinton, G., Yao, A., Song, D., Abbeel, P., Harari, Y. N., Hadfield, G., Russell, S., Kahneman, D., & Mindermann, S., *Managing ai risks in an era of rapid progress*, arXiv preprint arXiv:2310.17688, 2023, passim (on neural networks and deep learning that operate jointly in the age of Artificial Intelligence-embedded technologies).

[29] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[30] Rudin, Cynthia/Wang, Caroline/Coker, Beau, "The Age of Secrecy and Unfairness in Recidivism Prediction", *Harvard Data Science Review*, Issue 2.1. Winter 2020, 2020, passim; Eckhouse, L., Lum, K., Conti-Cook, C./Ciccolini, J., "Layers of bias: A unified approach for understanding problems with risk assessment", *Criminal Justice and Behavior*, Volume 46, Issue 2, 2019, pp. 185-209; Stevenson, M. T./Slobogin, C., "Algorithmic risk assessments and the double-edged sword of youth", *Washington University Law Review*, Volume 96, 2018, passim.

[31] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300. See also: Ponce, Paula Pedigoni, "Direct and indirect discrimination applied to algorithmic systems: Reflections to Brazil", *Computer Law & Security Review*, Volume 48, 2023, 1057-1066.

Crucially, in the midst of the era of big data[32] and cybersurveillance[33] algorithms have pierced into all aspects of social life[34]. Burgeoningly, a plethora of energy-draining and time-consuming tasks are being currently carried out by algorithms[35], ranging from credit scoring[36], predictive policing[37], lending[38], mortgage redlining[39], screening resumes, evaluating

---

[32] Big data that is in the limelight these days. See: Sivarajah, Uthayasnkar et al, "Critical Analysis of Big Data Challenges and Analytical Methods", *Journal of Business Research*, Volume 70, 2017, pp. 263 and ff and passim; Jetten, Lieke /Sharon, Stephen, "Selected Issues concerning the Ethical Use of Big Data Health Analytics", *Washington & Lee Law Review Online*, Volume 72, 2016, passim; Wachter, S./Mittelstadt. B., "A right to Reasonable Inferences: Re-thinking Data Protection Law in the Age of Big Data and AI", *Columbia Business Law Review*, 2019, pp. 1 and ff and passim; See in French doctrine: Ferey, Samuel, « Analyse économique du droit, big data et justice prédictive », *Archives de Philosophie du Droit*, Volume 60, 2018, pp. 68-81; Završnik, A. "Algorithmic justice: Algorithms and big data in criminal justice settings", *European Journal of Criminology*, 2019, passim; Joh, Elizabeth E., "The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing." *Harvard Law and Policy Review*, 10 (1), 2016, pp. 15-42.

[33] Hu, Margaret, "From the National Surveillance State to the Cybersurveillance State", *The Annual Review of Law and Social Science*, Volume 13, 2017, p. 168 ("*as a result, the Cybersurveillance State, as a technological successor to the National Surveillance State, will execute bureaucratized biometric cybersurveillance*"). Italics added.

[34] Xi Chen, "Algorithmic proxy discrimination and its regulations", *Computer Law & Security Review*, *The International Journal of Technology Law and Practice*, Volume 54, September 2024, 2024, pp. 1-13.

[35] Ross, M./Taylor, J., "Managing AI Decision-Making Tools, Technology and Analytics", *Harvard Business Review*, 10 Nov. 2021, 2021.

[36] See: Hurley, Mikella/Adebayo, Julius, "Credit Scoring in the Era of Big Data", *Yale Journal of Law and Technology*, Volume 18, 2016, pp. 148-168.

[37] Brayne, S., "Big data surveillance: The case of policing", *American Sociological Review*, Volume 82, Issue 5, 2017, pp. 977-1008.

[38] Bruckner, Matthew Adam, "The Promise and Perils of Algorithmic Lenders' Use of Big Data", *Chicago-Kent Law Review*, Volume 93, 2018, pp. 25-29.

[39] In spite of its deep-seated racial bias, decision-makers continue to deploy redlining tactics through big data and may even justify such unlawful practices by claiming that the latter abide by cost-effectiveness. See: Humber, Nadiyah J., "A Home for Digital Equity: Algorithmic Redlining and Property Technology", *California Law Review*, Volume 111, Issue 5, 2023, pp. 1421-1484. See also: Hurley, Mikella/Adebayo, Julius, "Credit Scoring in the Era of Big Data", *Yale Journal of Law and Technology*, Volume 18, 2016, pp. 148, 151-156, 172. See also: O'Neill, Cathy, *Weapons of Math Destruction. How*

employee overall performance, establishing premiums, to underpinning the high-stakes[40] decisions[41] made by government officials[42], which, in turn, begets algorithmic accountability of which policymakers must not shy away from[43]. Albeit human decision-makers play a pivotal role in keeping individual[44] justice unscathed, Artificial Intelligence-embedded technologies are being increasingly deployed to make high-stake decisions[45],

---

Big Data Increases Inequality and Threatens Democracy. London: Penguin Random House, 2016, pp. 128 and ff and passim.

40 See: Weizenbaum, Joseph, *Computer Power and Human Reason: From Judgment to Calculation*, 1st edition, New York, W. H. Freeman & Co, 1976, pp. 1-300 (227) (this founding father of Artificial Intelligence presciently noted that we should not entrust Artificial Intelligence-embedded technologies with "tasks that demand wisdom").

41 But see: Rudin, Cynthia, "Please stop explaining black box models for high stakes decisions", *Proceedings of NeurIPS 2018 Workshop on Critiquing and Correcting Trends in Machine Learning (NIPS 2018)*, 2018, passim.

42 Xi Chen, "Algorithmic proxy discrimination and its regulations", *cit*., pp. 1-13. But see: Williams, Rebecca, "Accountable Algorithms: Adopting the Public Law Toolbox Outside the Realm of Public Law", *Current Legal Problems,* Vol. 75, 2022, pp. 237–263; Williams, Rebecca, "Rethinking Administrative Law for Algorithmic Decision-Making", *Oxford Journal of Legal Studies*, Volume 42, 2021, pp. 468 e ff and passim (on the much-needed algorithmic accountability in the realm of administrative law).

43 See: Katyal, Sonia, "Private Accountability in the Age of Artificial Intelligence", *UCLA Law Review*, Volume 66, 2019, pp. 54-99 (on algorithmic accountability).

44 But see: M. Zalnieriute, L. B. Moses, G. Williams, "The Rule of Law and Automation of Government Decision-making, University of New South Wales Law Research Series", *Modern Law Review*, 82 (3), 2019, passim (on the sizzling hot challenges posed by an algorithmic-anchored decision-making at a government level).

45 See: Rudin, C., & Radin, J., "Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition", *Harvard Data Science Review*, 2019, Volume 1, Issue 2, passim; Tan, S., Caruana, R., Hooker, G., and Lou, Y., "Distill-and-compare: Auditing blackbox models using transparent model distillation". *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, Association for Computing Machinery*, 2018, pp. 303-310 (arguing that explainable models are too complex and fiddly to be understood by laypeople. As a result, interpretable models should be used instead).

which, in turn, creates a blazingly new cluster of challenges[46] that must be tackled heads-on[47].

Algorithmic decision-making can yield multifarious benefits to the overall advancement of modern societies[48]. Notably, it not only augments decision-making efficiency (*cost-benefit analysis*)[49] to adapt to the sheer amount of social changes catalysed by a seething social reality but, and foremost, appears to be more objective and neutral[50] than humans[51]

---

[46] See: Millar, Jason/Kerr, Ian, "Delegation, Relinquishment and Responsibility: The Prospect of Expert Robots", *Robot Law*, Calo, Ryan /Froomkin, A. Michael/Kerr, Ian, Cheltenham, Edward Elgar Publishing, 2016, pp. 102, 106-107 ("*Like the ancients, we will, quite rationally, come to rely upon them, knowing full well that we cannot necessarily explain the reasons for their decisions*"). Italics added. Converging: Casey, Anthony J./Niblett, Anthony, "Self-driving Laws", *University of Toronto Law Journal*, Volume 66, 2016, pp. 429-435 (contend that "as more information is generated, and the evolutionary algorithm updates and become a better forecaster, we imagine that judges will increasingly rely on the advice of the algorithm").

[47] Xi Chen, "Algorithmic proxy discrimination and its regulations", *cit.*, pp. 1-13.

[48] Xi Chen, "Algorithmic proxy discrimination and its regulations", *cit.*, pp. 1-13.

[49] Cost-benefit Analysis is in the limelight. Its development falls conspicuously outside the scope and aim of paper though. See: Sunstein, Cass R., "The Real World of Cost-Benefits Analysis: Thirty-Six Questions (and almost as many answers)", *Columbia Law Review*, 114 (1), 2014,177; Bayefsky, Rachel, "Dignity as a value in Agency Cost-Benefit Analysis", *Yale Law Journal*, 123 (2014): 1740; Sunstein, Cass R., Valuing Life: Humanizing the Regulatory State, 2014, 8 ff; Gordon, Jeffrey, "The Empty Call for Cost-Benefits Analysis for Financial Regulators", *Journal of Legal Studies*, 43, 2014, 351; Sunstein, Cass R., "The Limits of Quantification", *California Law Review*, 103, 2014, 1369; Posner, Eric A./Weyl E. Glen, "Cost-Benefit Analysis of Financial Regulations: Response to Criticisms", *Yale Law Journal*, 124, 2015, pp. 246-265 (parsing the relationship between agency cost-benefit analysis and human dignity).

[50] But see: Citron, Danielle Keats/Pasquale, Frank A., "The Scored Society: Due Process for Automated Predictions", *Washington Law Review*, Volume 89, 2014, pp. 1-34 (arguing that "[a]*dvocates applaud the removal of human beings and their flaws from the assessment process*," but arguing that bias remains "[*b*]*ecause human beings program predictive algorithms, their biases and values are embedded into the software's instructions*"). Italics added.

[51] But see: Jiang, H., O. Nachum, *Identifying and Correcting Label Bias in Machine Learning*, 15 Jan 2019, 2019, available online at https://arxiv.org/pdf/1901.04966.pdf (access: 14.12.2024); Chakraborty, J./Majumder, S./Menzies, T., "Bias in Machine Learning Software: Why? How? What to

owing to the fact that Artificial Intelligence-embedded technologies are not subject to any[52] prejudice whatsoever[53].

That holds true, but it is nonetheless truer that algorithms that are facially objective and seemingly neutral can lead to gruesome discrimination[54]. Shoddy data[55], botched, warped, and watered-down algorithms breed a slew of harmful consequences for both individuals and society. Worryingly, and perhaps not infrequently, even state-of-the art and top-notch algorithms suitably tailored to accomplish a given task

---

Do?, 9 Jul 2021, 2021, *arxiv.org*; Danks, D./A. J. London, "Algorithmic Bias in Autonomous Systems, Algorithmic Bias in Autonomous Systems", *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*, 2019, available online at https://www.ijcai.org/proceedings/2017/0654.pdf (access: 14.12.2024) (on the vast assortment of algorithmic biases).

[52] Xi Chen, "Algorithmic proxy discrimination and its regulations", *cit*., pp. 1-13.

[53] It is often averred, though, that bigoted algorithms are set to be brought to the fore as they are one foundational basis of *justice prédictive* so voguish in France. No surprise stems from the fact that prejudiced algorithms will feed *justice prédictive* - and that is something quite unsavoury and unpalatable in criminal justice indeed. See: Garapon, Antoine, "Les enjeux de la justice predictive", *Juris-Classeur périodique*, Volume 31, 2018, passim. See also : Larret-Chahine, Louis, "Le droit isométrique: un nouveau paradigme juridique né de la justice prédictive », *Archives de Philosophie du Droit*, Volume 60, 2018, pp. 287 and ff and passim.

[54] Xi Chen, "Algorithmic proxy discrimination and its regulations", *cit*., pp. 1-13. See also: Mehrabi, N./Morstatter, F. N. Saxena/Lerman, K./Galstyan, A., *A Survey on Bias and Fairness in Machine Learning*, 25 Jan 2022, available at https:// arxiv.org/pdf/1908.09635.pdf (access: 14.12.2024).

[55] See: *Tech, Bias and Housing Initiative, the Promise and Perils of Residential Proptech, Year 1, Research Summary Report, April 2023, TechEquity Collaborative*, p. 11, available at: https://techequity.us/wp-content/uploads/2023/04/TBHI-Y1-Research-Summary-Report.pdf (28.11.2024). ("*Once bad data enters the analysis of one factor, it gets baked into the subsequent process, multiplying the effect that one disadvantageous data point can have on someone's overall housing determination*"). Italics added.

and fed with clean data can mimic, reproduce[56] and exacerbate[57] previous discriminatory[58] patterns[59]. No befuddling surprise stems from the fact that Artificial Intelligence-embedded technologies have been flagged as sources of algorithmic bias[60], algorithmic discrimination[61] and algorithmic

---

[56] Benjamin, Ruha, *Race After Technology: Abolitionist Tools for the New Jim Code*. New York: Polity, 2019, pp. 1-172 (7) (discussing that Artificial Intelligence-embedded technologies are viewed as iron-clad tools of objectivity and devoid of any biases whatsoever). Ruha Benjamin contends that "the employment of new technologies that reflect *and reproduce existing inequities but that are promoted and perceived as more objective or progressive than the discriminatory systems of a previous era*"). Italics added.

[57] *Tech, Bias and Housing Initiative, the Promise and Perils of Residential Proptech, Year 1, Research Summary Report, April 2023, TechEquity Collaborative*, p. 5 available at: https://techequity.us/wp-content/uploads/2023/04/TBHI-Y1-Research-Summary-Report.pdf (28.11.2024). ("*Residential Proptech also has the potential to exacerbate the bias inherent in our housing system (...) Opaque algorithms incorporate biased data into their decision-making processes*"). Italics added.

[58] Roithmayr, Daria, *Reproducing Racism: How everyday choices lock in White Advantage*. New York: New York University Press, 2014, pp. 1-272 (contends "*that racial inequality reproduces itself automatically over time because early unfair advantage for whites has paved the way for continuing advantage. This book is designed to change the way we think about racial inequality* (...). Italics added. This book points out how White privilege is perpetuated throughout generations through residential segregation (*mortgage redlining*), abhorrent nepotism, and other bemoaning social practices that pester modern societies at large).

[59] Xi Chen, "Algorithmic proxy discrimination and its regulations", *cit*., pp. 1-13.

[60] Danks, D./London, A.J., "Algorithmic Bias in Autonomous Systems, Algorithmic Bias in Autonomous Systems", *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*, 2019; Ferrer, X./van Nuenen, T./Such, J. M./Cote, M./Criado, N., "Bias and Discrimination in AI: a cross-disciplinary perspective", King's College London, United Kingdom, 11 Aug 2020, 2020.

[61] See on algorithmic discrimination: Kleinberg, Jon, et al, "Discrimination in the Age of Algorithms", *Journal of Legal Analysis*, 2018, pp. 114-174; See: Kleinberg, Jon, et al, "Algorithms as Discrimination Detectors", *Proceeds of the National Academy of Science*, Volume 117, 2020, pp. 30096-30110; Griemmelmann, James/Westreich, Daniel, "Incomprehensible Discrimination ", *California Law Review Online*, Volume 7, 2018, pp. 164-171; Xiang, Alice, "Reconciling Legal and Technical Approaches to Algorithmic Bias", *Tennessee Law Review*, Volume 88, 2021, passim; Karnow, Curtis E. A., "The Opinion of Machines", *The Cambridge Handbook of the Law of Algorithms*, Bartfield, Woodrow (Editor). Cambridge, United Kingdom: Cambridge University Press, 2021, pp. 15 and ff and passim; Strandburg, Katherine J., "Rulemaking

injustice[62]. Take the example of COMPAS, a recidivism predictive tool used in some jurisdictions of the United States of America[63].

Whereas COMPAS does not use ethnicity-specific data[64], it nonetheless creates an algorithmic myth of colour-blindness[65] by bolstering the ungrounded hope that input exclusion («input data fallacy») can breed non-discriminatory algorithms[66]. Therefore, we are increasingly led to believe that colour-blind algorithms[67] amount to the panacea for a myriad of social diseases with which modern societies have been floundering for a long-haul[68]. Which could not be farther from the truth.

Strikingly, it bears mentioning that COMPAS does not go easy on coloured people. Tellingly, COMPAS is more likely to falsely flag members

---

and Inscrutable Decision Tools", *Columbia Law Review*, Volume 119, 2019, passim; Wachter, Sandra/Mittelstadt, Brent/Russell, Chris, "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR", *Harvard Journal of Law and Technology*, Volume 31, 2018, pp. 841 e ss and passim; Jackson, Maya C., "Artificial Intelligence and Algorithmic Bias: The Issues with Technology Reflecting History and Humans", *Journal of Business Technology Law*, Volume 16, 2021, pp. 299-316.

62 Ting-an Lin, "Democratizing AI" and the Concern of Algorithmic Injustice", *Philosophy and Technology*, Volume 37 (103), 2024, pp. 3-27. See also: Birhane, A., "Algorithmic injustice: A relational ethics approach", *Patterns*, 2 (2), 2021, pp. 1-9. Završnik, A. "Algorithmic justice: Algorithms and big data in criminal justice settings", *European Journal of Criminology*, 2019, passim.

63 An acronym that stands for *Correctional Offender Management Profiling for Alternative Sanctions* (COMPAS).

64 Xi Chen, "Algorithmic proxy discrimination and its regulations", *cit*., pp. 1-13.

65 Brewer, R. M./Heitzeg, N. A., "The racialization of crime and punishment: Criminal justice, Color-Blind racism, and the political economy of the prison industrial complex", *American Behavioral Scientist*, Volume 51, Issue 5, 2008, pp. 625-644.

66 Gillis, Talia B., "The Input Data Fallacy", *Minnesota Law Review*, Volume 106, 2022, pp. 1176-1263.

67 But see: Allen, James, "The Color of Algorithms: An Analysis and Proposed Research Agenda Deterring Algorithmic Redlining", *Fordham Urban Law Journal*, Volume 46, 2019, pp. 219-234 (229-234) (discussing how historical mortgage redlining greatly swayed modern algorithms generated by massive batches of data sets that consist of decades of information built on exclusion and racial discrimination in the purview of financing, marketing, and housing).

68 Sunstein, Cass R., "Governing by Algorithm? No Noise and (Potentially) Less Bias", *Duke Law Journal*, Volume 71, 2022, pp. 1175 and ff.

of racialized communities (i.e African-American and Latinos) as likely to engage in future criminal activities[69]. Thence, its false-positive rate is exceedingly higher for members of racialized communities than for White defendants[70], while its false-negative rate[71] is exceedingly greater[72] for the latter than for the former[73].

Compellingly, COMPAS holds members of historically disadvantaged groups (coloured people[74], immigrants[75], Latinos[76], women[77])

---

69 Xi Chen, "Algorithmic proxy discrimination and its regulations", *cit*., pp. 1-13.

70 Xi Chen, "Algorithmic proxy discrimination and its regulations", *cit*., pp. 1-13.

71 Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218 and ff and passim.

72 Flores, A.W./Lowenkamp, C.T./Bechtel, K., "False positives, false negatives, and false analyses: a rejoinder to "machine bias: there´s software used across the country to predict future criminal", *Federal Probation Journal*, Volume 80, 2016, pp. 38-46. Sun, R. (Ed.), *The Cambridge Handbook of Computational Cognitive Sciences*, Cambridge, Cambridge University Press, 2023.

73 Xi Chen, "Algorithmic proxy discrimination and its regulations", *cit*., pp. 1-13.

74 See: Chandler, Anupam, "The Racist Algorithm", *Michigan Law Review*, Volume 115, n. º 6, 2018, p. 1026 (on racist algorithms, noting that "the possibilities for discriminatory manipulation are legion"). But see: Altenburger, Kristen M./Ho, Daniel E., "When Algorithms Import Private Bias into Public Enforcement: The Promise and Limitations of Statistical Debiasing Solutions", *Journal of Institutional & Theoretical Economics*, Volume 175, 2019, pp. 97-98 (calls for a "debias predictive algorithms" to assuage racist and/or biased, doctored and bigoted algorithms nevertheless).

75 Maxwell, J./Tomlinson, J., *Experiments in Automating Immigration Systems*. Bristol: Bristol University Press, 2022, passim; Borgesius, FJ Zuiderveen, "Strengthening legal protection against discrimination by algorithms and artificial intelligence", *International Journal of Human Rights*, Volume 24, Issue 10, 2020, pp. 1573 and ff and passim. Hu, Margaret, "From the National Surveillance State to the Cybersurveillance State", *The Annual Review of Law and Social Science*, Volume 13, 2017, p. 168 ("Because of the routinized and administrative nature of the government-led *big data program* or *data surveillance (dataveillance) program, contemporary cybersurveillance is likely to be viewed as justified under crime, immigration control, and counterterrorism policy rationales*"). Italics added.

76 See: Hamilton, Melissa, "The Biased Algorithm: Evidence of Disparate Impact on Hispanics", *American Criminal Law Review*, Volume 56, 2019, pp. 1553 and ff and passim.

77 Hamilton, Melissa, "The sexist algorithm", *Behavioral Sciences and the Law*, Volume 37, Issue 2, Special Issue: the Use of Statistics in Criminal Cases, March-April 2019, 2019, pp. 1 and ff ("*multiple measures of algorithmic equity*

and whites to radically different standards of algorithmic fairness. Put differently, COMPAS treats the former less fairly than the latter[78] who are scored by it[79]. Bafflingly, and not in the least disturbingly, this racial discrimination does not spring from any malicious[80] intent[81] whatsoever. Rather, it flows from the fact that the algorithm, ensnared on its own inscrutable inner workings[82], leverages on a cluster of apparently harmless proxies to disproportionately target[83] members of racialized communities.

---

and predictive accuracy are provided to support the conclusion that this algorithm is sexist."). Italics added.

78 Note, though, that algorithmic fairness lends itself to several incompatible standards. Whilst COMPAS falls short of meeting the baseline standards of "equal opportunity" and "equalized odds", it refrains itself from discriminating African-Americans from the prism of "predictive parity" nonetheless. A full-fledged analysis of the extant state of this scorching topic can be found in: Hübner, Dietmar, "Two Kinds of Discrimination in AI-Based penal Decision-Making", *SIGKDD Explorations*, 2021.

79 Xi Chen, "Algorithmic proxy discrimination and its regulations", *cit*., pp. 1-13.

80 It is noteworthy that some forms of racial discrimination may not be fully intentional though. More likely than not, it can be related with the perils of big data colour-blindness that nonetheless relies on proxies to race and gender to make algorithmic decisions. Another problem is the one related with "creditworthiness by association" that causes more problems than it solves. Hurley, Mikella/Adebayo, Julius, "Credit Scoring in the Era of Big Data", *Yale Journal of Law and Technology*, Volume 18, 2016, p. 149. See also: Campisi, Natalie, From Inherent Racial Bias to Incorrect Data The Problems With Current Credit Scoring Models, Forbes (Feb. 26, 2021), https://www.forbes.com/advisor/creditcards/from-inherent-racial-bias-to-incorrect-data-the-problems-with-current-credit-scoring-models/ [https://perma.cc/EWA4-CS8E] (highlighting rampant racial bias in credit scoring systems); Oyama, Rebecca, "Do Not (Re)Enter: the Rise of Criminal Background Tenant Screening as a Violation of the Fair Housing Act", *Michigan Journal of Race and Law*, Volume 15, 2009, pp. 181, 184-93 (discussing a broad range of historical/criminal accounts as a way to screen undesirable tenants, such as African-Ameriacans and Latinos)

81 Xi Chen, "Algorithmic proxy discrimination and its regulations", *cit*., pp. 1-13.

82 See: Scherer, Matthew U., "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies", *Harvard Journal of Law and Technology*, Volume 29, 2016, pp. 353-369 (frames the crippling problem of opacity as "the possibility that the inner workings of an AI system may be kept secret and may not be susceptible to reverse engineering").

83 Xi Chen, "Algorithmic proxy discrimination and its regulations", *cit*., pp. 1-13.

In an era propelled by more intelligent algorithms and gargantuan/behemoth data generation[84], it takes scarce imagination to foresee the ubiquity and ineluctability of the algorithm discrimination caused by such proxies[85].

Bearing this very firmly in mind, this paper adamantly avers that machines should not be entrusted with the task of making high-stake decisions that may heavily bear on citizens' fundamental rights (i.e. the right to a due process, a central tenet of criminal procedure law) or inordinate harms may otherwise occur (i.e. the loss of functional reputation of the system of administration of justice). To cater for that, this paper leverages on a *law-in-context methodology*[86] – that seeks to deftly flesh out the theoretical underpinnings of the concepts brought forth throughout this long-winded paper – while shedding a keen light on the roadblocks that stand in the way of a sought-after «technological due process»[87] in the remit of criminal procedure law.

Building upon this, this paper unfolds as follows. Part II leverages on a piece of empirical research to premise the main contention that there

---

[84] See: Wonyoung, So, "Which Information Matters? Measuring Landlord Assessment of Tenant Screening Reports", *Housing Policy Debate*, Volume 33, n. º 6, 2023, pp. 1484–1510 (1501-1502) ("*Algorithmic scoring using criminal records is problematic because training data—a data set that is used to train a machine learning model—is inaccurate and the outcome variable ("desirable tenants") is unclear*"). Italics added.

[85] Xi Chen, "Algorithmic proxy discrimination and its regulations", *cit.*, pp. 1-13.

[86] About law-in-context methodology: Santos, Hugo Luz dos Santos, *Towards a Four-Tiered Model of Mediation*. New York: Springer Nature, 2023, 1-216; Santos, Hugo Luz dos, *Multidisciplinary Dynamics of Mediation*. New York: Springer Nature, 2025, Volume I, pp. 1-686; Santos, Hugo Luz dos, *Multidisciplinary Dynamics of Mediation*. New York: Springer Nature, 2025, Volume II, pp. 1-926; Santos, Hugo Luz dos, *Controllable Artificial Intelligence and the Future of Law*. New York: Springer Nature, 2025, pp. 1-1313; Santos, Hugo Luz dos/Leong, Cheng Hang, "Culture Matters": Expedited Arbitration and Arb-Med in Macau", Hong Kong Law Journal, Volume 54:3, 2024. See also: Santos, Hugo Luz dos, *Inteligência Artificial e Processo Penal*. Braga: NovaCausa Edições Jurídicas, 2022.

[87] Citron, Danielle K., "Technological Due Process", *Washington University Law Review*, Volume 85, 2008, pp. 1249-1253 ("computer programs seamlessly combine rulemaking and individual adjudications without the critical procedural protections owed either of them").

is an «impossibility of race neutrality» in the remit of algorithmic risk assessment in criminal justice settings. On the heels of such a sweeping assertion, Part III adamantly avers that the building blocks of statistics that underpin algorithmic fairness (e.g. predictive parity) are impossible to be attained in algorithmic risk assessment undertaken in the remit of criminal procedure law settings. Drawing on these findings, Part IV contends that machines should not be entrusted with the task of making high stake decisions (e.g. those related with citizens 'fundamental rights, like freedom) in criminal procedure law settings. As an upshot, Part V cobbles together the findings arising out of this paper.

2. «The quest for predictive parity» (or lack thereof) in algorithmic risk assessment: the problem of equality trade-offs in Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)

As adumbrated earlier in this essay, there is no such thing as a «one size fits all», let alone a catch-all concept, amenable to deftly measure racial equality in algorithmic risk assessment[88] in criminal procedure law settings. Quite the opposite, there is a feast of potentially applicable measures in this regard[89]. But here´s the twist: it is utterly

---

[88] See: Dieterich, William et al., "COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity", *Northpointe Inc*. 1, 2-3, 8-13 (July 8, 2016), http://go.volarisgroup.com/rs/430 -MBX-989/images/ProPublica_Commentary_Final_070616.pdf [https://perma.cc/K4GM -RBQY] (stating that a predictive instrument is bigoted/flawed only if a certain score, or a given classification, means a starkly likelihood of the forecasted outcome for members of one racial group than members of the other racial group). See also: Flores: Anthony W. et al., "False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks," *Federal Probation*, Volume 80, 2016, pp. 38-40 (positing that "well-established and accepted standards exist to test for bias in risk assessment"); Skeem, Jennifer L./Lowenkamp, Christopher T., "Risk, Race, and Recidivism: Predictive Bias and Disparate Impact", *cit*., p. 685 (asserting that if "a given score [has] the same meaning regardless of group membership," the instrument is "unbiased"). See: Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[89] See: Selbst, Andrew D., "Disparate Impact in Big Data Policing", *Georgia Law Review*, Volume 52, 2017, p. 109 (arguing that law enforcement agencies

impossible to attain racial equality according to every single measure concomitantly[90]. The *ProPublica* debate vividly portrays this definitional quandary. *ProPublica* posited that the algorithmic tool *Correctional Offender Management Profiling for Alternative Sanctions* (COMPAS)[91] was «biased against coloured people».[92] In a nutshell, COMPAS was a warped algorithm[93] that ran counter to algorithmic fairness[94].

should carry out "algorithmic impact statements" with a view to measure the potential discriminatory impact of computer-run predictive policing technologies).

[90] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[91] Dobbie, Will, Jacob Goldin, and Crystal S. Yang, "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges," *American Economic Review*, Volume 108, 2018, 201–240. See: Baughman, Shima Baradaran, "Costs of Pretrial Detention", *Boston University Law Review*, Volume 97, 2017, p. 1 (calculating the whooping costs of detention); Heaton, Paul et al., "The Downstream Consequences of Misdemeanor Pretrial Detention", *Stanford Law Review*, Volume 69, 2017, 711, 759-69 (concluding, amongst other things, that pretrial detention ballooned the likelihood that a defendant would accrue new criminal charges within 18 months of a bail court hearing).

[92] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[93] From the prism of social sciences, warped algorithms accomplish nothing but the impairment of procedural justice. See: Thibaut, John/ Walker, Laurens, *Procedural Justice: A Psychological Analysis*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975, pp. 1-150 (procedural justice is tightly interlocked to raising the levels of fairness in legal settings. Absent of which procedural fairness is to mar. Lawmakers must not lose sight of such assertion). See also: Simmons, Ric, "Big Data and Procedural Justice: Legitimizing Algorithms in the Criminal Justice System", *Ohio State Journal of Criminal Law*, Volume 15, 2018, pp. 576 and ff and passim.

[94] Selbst, Andrew D., "Disparate Impact in Big Data Policing", *cit*., at 123 (arguing that "[t]he words 'discrimination,' 'fairness,' and 'bias' evoke a family of related concepts").

Leveraging on data from a jurisdiction where COMPAS was used to assess the likelihood that a pretrial[95] defendant would be rearrested[96] if he stayed at liberty[97], the *ProPublica* researchers compared COMPAS's risk classifications with defendants' actual outcomes - whether each defendant was rearrested[98] or not[99] - over the following two years. Outraged and fazed, *Northpointe*, the corporation that owns COMPAS, rejoindered that

---

[95] See: Stevenson, Megan, "Assessing Risk Assessment in Action", *Minnesota Law Review*, Volume 103, 2018, p. 303 (discussing the burgeoning use of pretrial risk assessment as a mandatory component of bail decisions in Kentucky). It is important to note that other tools have other target variables, yet the analysis in this offering applies to several other target variables as well. In the pretrial purview, for example, risk-assessment tools also forecast "failure to appear," outlined in terms of data points that vary across jurisdictions. See also: Mayson, Sandra G., "Dangerous Defendants", *Yale Law Journal*, Volume 127, 2018, pp. 509-13. There is also a spate of risk-assessment instruments that seek to forecast violence but in fact predict any hint or allegation of violence, whether it reflects in arrest or not (let alone conviction). See, e.g., Yang, Min et al., "The Efficacy of Violence Prediction: A Meta-Analytic Comparison of Nine Risk Assessment Tools", *Psychology Bulletin*, 2010, pp. 740-742 (arguing that "[t]he range of possible criterion variables for violence is wide, and "includes self-reports to third-party reports . . . , informal social service or police contact, formal contact or police charges, formal adjudication and court convictions, and incarceration"). See: Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[96] McIntyre, Frank/Baradaran, Shima, "Race, Prediction, and Pretrial Detention", *Journal Empirical Legal Studies*, Volume 10, 2013, pp. 742-759 (mentions black defendants' exceedingly high likelihood of being arrested on drug-related charges as a potential cause of the race gap).

[97] Stevenson, Megan/Mayson, Sandra G., "Pretrial Detention and Bail", Luna, Erik (Editor), *Reporting Criminal Justice: A Report of the Academy for Justice, Bridging the Gap between Scholarship and Reform*, 21, 2017, pp. 34-35 (poring over a batch of studies hinting that actuarial risk assessment can improve accuracy of pretrial risk judgments). See: Stevenson, Megan/Mayson, Sandra G., "Pretrial Detention and Bail", Luna, Erik, (Editor), *Reforming Criminal Justice: A Report of the Academy for Justice, Bridging the Gap between Scholarship and Reform*, pp. 21, pp. 45-47 (assessing evidence on efficacy of electronic monitoring and pinpointing key costs of electronic monitoring). See: Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[98] Milgram, Anne, et al., "Pretrial Risk Assessment: Improving Public Safety and Fairness in Pretrial Decision Making", *Federal Sentencing Report*, Volume 27, 2015.

[99] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

ProPublica's own dataset evinced that COMPAS was reportedly race[100] neutral![101] Oddly, both *ProPublica* and *Northpointe* had a point. Note, though, they were just focusing on different metrics of racial equality[102] in algorithmic decision-making[103].

Tellingly, *Northpointe* claimed that race neutrality was mirrored on the fact that both coloured and white defendants labelled[104] as high risk by COMPAS were indeed rearrested at equal rates[105]. Relatedly, a high-risk classification mirrored the very same likelihood of rearrest for a coloured defendants for a[106] white one (roughly 60% on the any-arrest-risk scale and 20% on the violent-arrest-risk scale, over a two-year period).[107]

---

[100] But see: Skeem, Jennifer L./Lowenkamp, Christopher T., "Risk, Race, and Recidivism: Predictive Bias and Disparate Impact", *Criminology*, Volume 54, 2024, pp. 680, 683-84, 704-06 (argues that criminal history correlates with race in their respective data set).

[101] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[102] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[103] Kleinberg, Jon, et al., "Human Decisions and Machine Predictions", *National Bureau of Economic Research*, Working Paper No. 23180, 2017, 2017, https://www.nber.org /papers/w23180.pdf [https://perma.cc/3WHJ-TWLJ].

[104] But see: Corbett-Davies, Sam/Goel, Sharad, "The Measure and Mismeasure of Fairness: a critical Review of Fair Machine Learning", *Cornell University Library*, 2018, p. 18 (broaches a type of *measurement bias* coined "feature bias," which is neatly a bias in the predictors x. Note, though, that there is a second type of measurement bias deemed "label bias," which, in turn, amounts to a bias in y. The authors argue that that *label bias* is the worst bias there is).

[105] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[106] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[107] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

This particular metric of racial equality[108] is frequently coined as predictive parity[109]. The fact that led *ProPublica* to believe that COMPAS was fraught with racial bias was something uniquely ingenious: a coloured defendant who would not be rearrested within a given span of time was far

---

[108] But see: Kim, Pauline T., "Data-Driven Discrimination at Work", *William and Mary Law Review*, Volume 58, 2017, p. 918 ("If the goal is to reduce biased outcomes, then a simple prohibition on using data about race or sex could be either wholly ineffective or actually counterproductive due to the existence of class proxies and the risk of omitted variable bias."); Lipton, Zachary et al., "Does Mitigating ML's Impact Disparity Require Treatment Disparity?", NIPS Proceedings 1, 2018, https://arxiv.org/pdf/1711.07076.pdf [https://perma.cc/X8VF-EY53] (arguing, premised on statistical examples, that a prohibition on race or sex data is rather counterproductive indeed); Corbett-Davies, Sam/Goel, Sharad, "The Measure and Mismeasure of Fairness: a critical Review of Fair Machine Learning", *Cornell University Library*, 2018, passim (discussing the "[l]imitations of anti-classification" as a fairness metric). See also: Kim, Pauline T., "Data-Driven Discrimination at Work", *cit*., p. 867 ("[I] f the goal is to discourage classification bias, then the law should not forbid the inclusion of race, sex, or other sensitive information as variables, but seek to preserve these variables, and perhaps even include them in some complex models."); Kroll, Joshua A., et al., "Accountable Algorithms", *University of Pennsylvania Law Review*, Volume 165, 2017, pp. 633, 693-95. See also: Kim, Pauline T., "Data-Driven Discrimination at Work", *cit*., p. 866 ("*classification bias* occurs when employers rely on classification schemes, such as data algorithms, to sort or score workers in ways that worsen inequality or disadvantage along the lines of race, sex, or other protected characteristics."). Italics added. These two uses of the word "bias" correspond to the notions of irrational versus rational (or statistical) discrimination. Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[109] Recall, the literature on algorithmic fairness has labelled measures of equality as either "group fairness" or "individual fairness" metrics. This phosphorescent dichotomy, though, can be very misleading indeed. Almost every tentative measure of "group fairness" can be framed using the term "individual" (for example, predictive parity entails that, for any individual, a given risk score conveys the same average risk irrespective of race). Relatedly, any "individual-fairness" metric can be framed using the term "group" (for example, a single-threshold rule requires that the group of people who poses any degree of risk should receive the same risk score). The difference is that which "individual-fairness" metrics relates to how the algorithm reaches its output in each individual case, whilst "group-fairness" metrics relates to the distribution of outputs and/or their accuracy across particular groups. See: Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

more likely to be labelled[110] as high risk (44.9%) than a white defendant who would not be rearrested (23.5%).[111]

Statistically speaking, the false-positive rate was much higher for the coloured defendants than the white defendants.[112] In the meantime, a white defendant who would be rearrested was far more likely to be tagged as low risk (47.7%) than a coloured defendant who would be rearrested (28.0%).[113] Notably, the false-negative rate was far higher for white defendants than for coloured defendants[114].

[110] Corbett-Davies and Goel coined this sweltering issue as "label bias" and diagnose it as "perhaps the most serious obstacle facing fair machine learning." But see: Corbett-Davies, Sam/Goel, Sharad, "The Measure and Mismeasure of Fairness: a critical Review of Fair Machine Learning", *Cornell University Library*, 2018, p. 18 (broaches a type of *measurement bias* coined "feature bias," which is neatly a bias in the predictors x. Note, though, that there is a second type of measurement bias deemed "label bias," which, in turn, amounts to a bias in y. The authors argue that that label bias is the worst bias there is). See also: Barocas, Solon/Selbst, Andrew D., "Big Data´s Disparate Impact", *California Law Review*, Volume 104, 2016, pp. 677-78 (arguing that predictive data is predictive of a given/specific target variable with training data which is both correctly "labelled" and "collected."); Kleinberg, Jon/Ludwig, Jens/Mullainathan, Sendhil/Sunstein Cass R., "Discrimination in the Age of Algorithms", *Journal of Legal Analysis*, Volume 10, 1, 2018, pp. 28-29, 33-34 (draw a distinction between "group differences in the raw data" and biases for the "choice of predictors.").

[111] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[112] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[113] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[114] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

ProPublica viewed these racial disparities in COMPAS's error rates as an inexcusable, and deep-seated, injustice[115] and racial discrimination[116]. The racial disparity in error rates was not, though, a byproduct of the devilishly harmful virus of unfairness besetting the COMPAS algorithm itself. Rather, it was a mathematical outcome of the different rates of arrest between the black and white defendants in the underlying dataset.

---

[115] See: Stevenson, Megan/Mayson, Sandra G., "Pretrial Detention and Bail", Luna, Erik, (Editor), *Reforming Criminal Justice: A Report of the Academy for Justice, Bridging the Gap between Scholarship and Reform*, pp. 21, 34-39 (assessing the risk of pretrial risk assessments while highlighting a slew of deep-rooted concerns regarding accuracy, racial-equality, and procedure); See: Chandler, Anupam, "The Racist Algorithm", *Michigan Law Review*, Volume 115, n.º 6, 2018, pp. 1023-1025 (arguing that the real-world facts on which algorithms used in criminal justice risk assessment are based are "deeply suffused with invidious discrimination"); Eaglin, Jessica M., "Constructing Recidivism Risk", *cit*., pp. 94-99 (argues that risk assessment has every chance to "compromise[e] equality"); Mayson, Sandra G., "Dangerous Defendants", *cit*., pp. 494-96; Selbst, Andrew D., "Disparate Impact in Big Data Policing", *cit*., *passim*; Kim, Pauline T., "Data-Driven Discrimination at Work", *William and Mary Law Review*, Volume 58, 2017, pp. 857, 863-64 (shedding a keen light on the racial effects of algorithmic prediction in the remit of employment). Converging: Zywicki, Todd J./Adamson, Joseph D., "The Law and Economics of Subprime Lending", *University of Colorado Law Review*, 2009, pp. 1-9 (contending that Black workers with roughly the same abilities and educational background earn far less than comparable white workers or have considerably fewer employment opportunities); King, Allan G./Markovich, Marko J., "Big Data" and the Risk of Employment Discrimination, *Oklahoma Law Review*, Volume 68, 2016, pp. 555-563 (on employment discrimination). Dobbie, Will, Jacob Goldin, and Crystal S. Yang, "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges," *American Economic Review*, Volume 108, 2018, 201–240.

[116] On racial discrimination: Hellman, Deborah, "What Makes Genetic Discrimination Exceptional?", *American Journal of Law and Medicine*, Volume 29, 2003, pp. 77, 83-86; Morrow, Jeffrey S., "Insuring Fairness: The Popular Creation of Genetic Antidiscrimination", *Georgia Law Journal*, Volume 98, 2009, pp. 215, 230-32; Prince, Anya E.R., "Insurance Risk Classification in an Era of Genomics: Is a Rational Discrimination Policy Rational?", *Nebraska Law Review*, Volume 96, 2018, pp. 624, 630-34, 641-42 (shedding light on "fair" and "unfair" discrimination). Note, though, that the word "discrimination" can also be used in a technical legal sense. In this sense it would mean that such differential treatment or disparate impact would incur liability pursuant to the fullest extent of antidiscrimination law. See: Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

Since the rate of arrest was much higher among the coloured defendants, they, roughly, had higher arrest-risk classifications[117].

Whenever the average risk is much higher for one group than for another, a higher proportion of the former group will be forecasted to be rearrested. In addition to that, a higher proportion of that group will be mistakenly predicted to be rearrested as well[118].

This holds a wealth, and treasure trove, of truth regardless of how carefully programmed/devised the machine learning algorithm is, provided that the machine learning algorithm is equally seeking to have equal predictive accuracy[119] for each racial group[120].

The pinpointed machine learning algorithm engages in predictions[121] that are equal across the two racial lines in one dimension: a positive prognosis is equally correct for each racial line. For both the black and the white groups, 50% of those forecast for rearrest are eventually rearrested. Whenever the machine learning algorithm is used prospectively,

---

[117] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[118] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[119] It is important to note, though, that an algorithm designed for maximum accuracy will adjust itself to the majority data, and may be less accurate for members of the underrepresented group, thus breeding *inductive bias*. See: Shellenbarger, Sue, "A Crucial Step for Avoiding AI Disasters", *Wall Street Journal*, (Feb. 13, 2019, 9:57 AM ET), https://www.wsj.com/articles/a-crucial-step-for-avoiding-ai-disasters-11550069865?ns=prod/accounts -wsj [https://perma.cc/C28U-LAAE] (discussing this issue while emphasizing how diverse development teams pay more attention to unrepresentative and underrepresented data sets). Human programmers can circumvent and bypass this problem by weighting the minority-group data more heavily, by designing different algorithms for each racial group, or by striving to contain more data to equalize group representation in the data set. See: Barua, Sukarna et al., "MWMOTE—Majority Weighted Minority Oversampling Technique for Imbalanced Data Set Learning", *IEEE Transactions on Knowledge & Data Engineering*, Volume 26, 2014, pp. 405, 405-06. See also: Hamilton, Melissa, "The Biased Algorithm: Evidence of Disparate Impact on Hispanics", *American Criminal Law Review*, Volume 56, 2019, pp. 1553 and ff and passim (which clearly demonstrates that COMPAS was significantly less accurate for Hispanic than for white defendants by several measures and suggesting that smaller numbers of Hispanic defendants might be the cause). See: Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[120] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[121] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

a positive vaticination for any individual will equal a 50% likelihood of rearrest irrespective of whether the person is white or black[122].

For purposes here, the machine learning algorithm indeed achieves predictive parity, the equality metric that *Northpointe* accounted for and remarked upon. In many ways, though, the algorithm breeds blatantly unequal results. Expound on the rate of false predictions amongst those who will not be rearrested – the oft-called false-positive rate. Out of the eight white figures who will not be rearrested, two are erroneously forecast for rearrest. Out of the nine black figures who will not be rearrested, only one is erroneously forecast for arrest[123]. This racial disparity amounts to both *disparate treatment*[124] and *disparate*[125]

---

[122] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[123] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[124] Howard, Justin D Howard, "Refashioning the Disparate Treatment and Disparate Impact Doctrines in Theory and in Practice", *Howard Law Journal*, Volume 41, 1998, passim; Seiner, Joseph, "Disentangling Disparate Impact and Disparate Treatment: Adapting the Canadian Approach", *Yale Law & Policy Review*, Volume 25, 2006; Stephanopoulos, Nicholas O. "Disparate Impact, Unified Law", *Yale Law Journal*, Volume 128, 2019, pp. 1566-1595. Note, though, that there are two main tools to be accounted for when it comes to entertaining discrimination claims: the Equal Protection Clause of the Federal Constitution (and analogous state constitutional regulations) and federal and state statutes that forbid discrimination on various grounds, including, but not limited to, race. A discrimination claim (lawsuit) pursuant to the Equal Protection Clause must bring forth facts to prove *disparate treatment*. Absent of which, the claim will not succeed; a glimpse – or a mere hint - of *disparate impact* alone will not suffice in this regard. Antidiscrimination statutes also allow disparate treatment claims, and some likewise allow disparate impact claims. See: Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300. As Richard Primus expounds, whilst there are significant technical differences in the constitutional and statutory disparate treatment frames of references, substantive analysis of a disparate treatment claim pursuant to either is essentially the same. See: Richard Primus, "The Future of Disparate Impact", *Michigan Law Review*, Volume 108, 2010, pp. 1341, 1354-56. See: Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300. See also: Bornstein, Stepanhie, "Antidiscriminatory Algorithms", *Alabama Law Review*, Volume 70, 2018, p. 535 (positing that algorithmic discrimination at large could fall under the all-embracing umbrella of anti-stereotyping category within Title VII's disparate treatment).

[125] But see: Barocas, Solon/Selbst, Andrew D., "Big Data´s Disparate Impact", *California Law Review*, Volume 104, 2016, p. 699 (discussing how discriminatory data mining is analogous to unintentional disparate impact analysis).

impact[126] that befalls on members of racialized communities[127].

In spite of the widely accepted conceptual differences between intent-based and effect-based theories of disparate impact, both approaches often dovetail seamlessly. As far as the fair lending goes, the definition of disparate impact cases is somewhat shallow, nomadic, and erratic, and dispiritingly vague. Note, though, that the loan officer´s discretion often leads to exceedingly higher rates of rejections for members of racial minorities. See: Ayres, Ian/Klein, Gary/West, Jeffrey, "The Rise and (Potential) Fall of Disparate Impact Lending Litigation", *Evidence and Innovation in Housing Law and Policy*, Fennell, Lee Anne/Keys, Benajmin J. (Editors), Cambridge, Cambridge University Press, 2017, pp. 231-254. However, Schwemm and Taren avowedly contend that these cases might somewhat be deemed hybrid impact/intention cases. See: Schwemm, Robert G./Taren, Jeffrey L., "Discretionary Pricing, Mortgage Discrimination, and the Fair Housing Act", *The Harvard Civil Rights-Civil Liberties Law Review*, 2010, passim. The conduct being granularly parsed is discretion ascribed to brokers – which is considered by many stakeholders a harmless neutral practice -, may nonetheless lead to brokers to intentionally discriminate against members of racial minorities. See also: Selmi, Michael, "Indirect Discrimination and the Anti-Discrimination Mandate", Oxford, Oxford University Press, *Philosophical Foundations of Discrimination Law (Philosophical Foundations of Law)*, Hellman, Deborah/Moreau, Sophia (Editors), 2013, pp. 1-304 (pp. 257 and ff and passim).

[126] To be clear-cut, none of these output measures give rise to disparate impact liability under the applicable law. As adumbrated, only the first step in a legal disparate impact analysis is related with outputs; the ultimate question is whether and to what extent the challenged disparate impact is warranted or fully justified, a question that is undoubtably just as much about the decision-making process as a disparate treatment analysis. See: Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300. See also: Zatz, Noah D., "Disparate Impact and the Unity of Equality Law", *Boston University Law Review*, Volume 97, 2017, pp. 1357-1362 (contending that disparate impact and disparate treatment liability are fundamentally "separated superficially by the presence or absence of discriminatory intent but united fundamentally in addressing a common injury: status causation").

[127] See: Sunstein, Cass R., "Algorithms, Correcting Biases", *Social Research*, Volume 86, 2019, pp. 449-506 (this famed author tagged disparate impact as "disturbing in itself, in the sense that a practice that produces such an impact helps entrench something like a caste system"). In spite of the widely accepted conceptual differences between intent-based and effect-based theories of disparate impact, both approaches often dovetail seamlessly though. See: Ayres, Ian/Klein, Gary/West, Jeffrey, "The Rise and (Potential) Fall of Disparate Impact Lending Litigation", *Evidence and Innovation in Housing Law and Policy*, Fennell, Lee Anne/Keys, Benajmin J. (Editors). Cambridge: Cambridge University Press, 2017, pp. 231-254.

The false-positive rate is exceedingly higher for the white group (25%) than for the black one (11%). This is the form of egregious inequality that *ProPublica* identified in the COMPAS data. Similarly to *ProPublica´s* body of empirical evidence, this machine learning algorithm produces unequal results also in another dimension: twice as many white figures as black ones are forecast for rearrest. Compellingly, the machine learning algorithm exacts a toll on the group with the higher base rate: the black group. In the taxonomy of which data scientists are fond of, the machine learning algorithmic risk assessment is far from achieving statistical parity[128].

It is possible to tweak the algorithm to equalize the false-positive rates for the two groups[129]. Note, though, that this tweaking is not without sizable costs[130]. Whenever modification is made in the machine learning

---

[128] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300. See: Corbett-Davies, Sam/Pierson, Emma/Feller, Avi/Goel, Sharad Goel/Huq, Aziz, "Algorithmic Decision Making and the Cost of Fairness", *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Ass'n for Computing Machinery ed., 2017), 2017, passim (coining statistical parity as a "popular" concept of fairness in the risk-assessment and algorithmic-fairness literature); Feldman, Michael et al., "Certifying and Removing Disparate Impact (July 16, 2015), 2015 https://arxiv.org/pdf/1412.3756.pdf [https://perma.cc/NNQ4-NHUH] (an early work in the algorithmic-fairness literature that adopts a statistical-parity metric). Note, though, that, theoretically, one should not shy away, or veer away, from bifurcating/splitting the "prediction" issue from the "decision" issue. See: Corbett-Davies, Sam/Pierson, Emma/Feller, Avi/Goel, Sharad Goel/Huq, Aziz, "Algorithmic Decision Making and the Cost of Fairness", *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Ass'n for Computing Machinery ed., 2017), 2017, passim.

[129] In practical terms, "people with otherwise identical risk prognoses" will include a cohort of people who have exactly the same observable risk traits, with exception to race. But it might also include two people who each have radically different traits, but who nevertheless pose equivalent statistical risk in light of the best method of estimation. See: Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[130] Compellingly, Sam Corbett-Davies and peers, parsing the same Broward County data that ProPublica did, concluded that attaining parity in false-positive rates while still optimizing for public safety (and without detaining any other defendants) would beget a 7% uptick in violent crime. See: Corbett-Davies, Sam/Pierson, Emma/Feller, Avi/Goel, Sharad Goel/Huq, Aziz, "Algorithmic Decision Making and the Cost of Fairness", *Proceedings of the 23rd*

algorithm, there is a chain reaction that ripples across the stagnant waters of algorithmic risk assessment. For instance, predicting arrest for a greater proportion of the black group is quite cumbersome. Tellingly, after tweaking in the algorithm to ensure greater racial equity, Professor Sandra G. Mayson noted that 25% of the non-rearrestees – for both black group and the white group - were mistakenly forecast[131] for rearrest[132]. As an upshot, the false-positive rate was 25% for each group[133].

The global number of people predict for rearrest is likewise much closer across racial lines. Note, though, the ripple effect on the

---

ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Ass'n for Computing Machinery ed., 2017), 2017, p. 802. Moreover, 17% of those detained would be low-risk people for whom detention was either unwarranted or uncalled for. See: Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

131 Berk and his peers have also provided a scintillating portray of these equality/accuracy trade-offs, drawing on real arraignment data and a machine-learning algorithm that predicts new arrests for a domestic-violence offense within a period of twenty-one months. The base rate amongst black defendants was 11%, whereas the base rate amongst white defendants was 6%. Relatedly, Berk and his peers found that, if the machine learning algorithm is devised to attain predictive parity regarding to a forecast of no-rearrest, the "false negative and false positive rates vary dramatically by race." Specifically, the false negative rate is 49% for black defendants and 93% for white defendants; the false positive rate is 2% for white defendants and 24% for black defendants. See: Berk, Richard et al., "Fairness in Criminal Justice Risk Assessments: The State of the Art", *Sociological Methods and Research*, 2018, pp. 12-18 (July 2, 2018), https://journals.sagepub.com/doi/pdf/10.1177 /0049124118782533 [https://perma.cc/LB82-47SB]. p. 32. See: Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

132 See: Berk, Richard, "Accuracy and Fairness for Juvenile Justice Risks Assessments", *Journal of Empirical Legal Studies*, Volume 16, 2019, p. 175 (comes to the terms with the fact that tweaking/altering data to attain statistical parity yields extremely high false-negative rates); Zachary Lipton et al., "Does Mitigating ML's Impact Disparity Require Treatment Disparity?", NIPS PROC. 1 (2018), https://arxiv.org/pdf/1711.07076.pdf [https://perma.cc/X8VF-EY53] (sets forth compelling evidence that enforced blindness to protected traits exacts a toll on the algorithms´ overall accuracy); Petersilia, Joan/Turner, Susan, "Guideline-Based Justice: Prediction and Racial Minorities", *Crime and Justice*, 1987, pp. 151-174 (argues that omitting markers correlated with race from a recidivism prediction algorithm exacts a toll on the overall accuracy of the model).

133 Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

accuracy of the rearrest forecasts themselves. For the white group, a forecast of rearrest remains 50% likely to be true. On the flip side, only about 30% likely to be true for the black group. Whenever the machine learning algorithm is used prospectively, a rearrest prediction will produce something meaningfully different depending on whether the figure is white or black[134]. This is the kernel of disparate treatment[135] that deserves to be overtly decried[136] in criminal justice settings.

One may argue that, to recoup predictive parity[137], suffice to alter the white group for whom rearrest is forecast. Again, that will inevitably create a shockwave of pure disparity, which is something you should not hold your breath for in criminal justice settings. As of now, amongst those who are indeed rearrested, the machine learning algorithm accurately predicts rearrest for 100% of the black arrestees, but is nonetheless oblivious of 50% of the white arrestees. Put otherwise, the machine learning algorithm now misses 50% of the white arrestees. There is now a humongous disparity in false-negative rates[138].

As this example vividly portrays, if the base rate of the forecasted outcome varies greatly across racial lines, it is utterly untenable, unsound, illogical, fallacious, and ultimately impossible to attain: (i) predictive parity;

---

[134] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[135] See, e.g., *Ricci v. DeStefano*, 557 U.S. 557, 577 (2009) (framing disparate treatment as another form of "intentional discrimination"); see also *Washington v. Davis*, 426 U.S. 229, 239- 41 (1976) (arguing that differential treatment of people of different races infringes the Equal Protection Clause only, and if, motivated by clear and proven "discriminatory racial purpose"). See: Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[136] See: Howard, Justin D., "Refashioning the Disparate Treatment and Disparate Impact Doctrines in Theory and in Practice", *Howard Law Journal*, Volume 41, 1998, passim; Seiner, Joseph, "Disentangling Disparate Impact and Disparate Treatment: Adapting the Canadian Approach", *Yale Law & Policy Review*, Volume 25, 2006; Stephanopoulos, Nicholas O. "Disparate Impact, Unified Law", *Yale Law Journal*, Volume 128, 2019, pp. 1566-1595.

[137] Berk, Richard et al., "Fairness in Criminal Justice Risk Assessments: The State of the Art", *cit*., p. 14 (describing "conditional use accuracy equality"); Dieterich, William et al., "COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity", *cit*., passim (describing "predictive parity"); Hardt, Moritz et al., "Equality of Opportunity in Supervised Learning", *cit*., passim.

[138] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

(ii) parity in false-positive rates[139]; (iii) parity in false-negative rates at the same time (unless prediction is quick-wittedly flawless, which is a mirific in this regard) and iv) colour-blindness[140]. Crucially, pundits have brought to the fore irrebuttable mathematical evidence of this axiom. Put it coarsely, whenever base rates vary greatly, polymaths should shift their gaze downstream – i.e. prioritize one of these metrics over the other. We can´t have both. Race neutrality is therefore unattainable[141] in criminal procedural law settings.

---

[139] When Richard Berk trained the algorithm only to maximize for overall accuracy, it predicted rearrest for 17% of the white subgroup and 33% of the black subgroup (a 16 percentage-point difference). Strikingly, the false-positive rates were 16% for the white subgroup whereas for the black groups was 28% (a 12 percentage-point difference). See: Berk, Richard, "Accuracy and Fairness for Juvenile Justice Risks Assessments", *cit*., p. 180. Notably, when Richard Berk changed the algorithm to equalize the *cost ratios*, it predicted rearrest for 10% of the white subgroup and 29% of the black subgroup (a 19 percentage-point difference). Relatedly, the false-positive rates were 8% for the white subset and 22% for the black subset (a 14 percentage-point difference). See: Berk, Richard, "Accuracy and Fairness for Juvenile Justice Risks Assessments", *cit*., p. 185. See: Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[140] For this reason, Sam Corbett-Davies and Sharad Goel refer to colorblindness as "anti-classification." See: Corbett-Davies, Sam/Gaebler, Johann D./Nilforoshan, Hamed/Shrof, Ravi/Goel, Sharad, "The Measure and Mismeasure of Fairness", *Journal of Machine Learning Research*, Volume 24, 2023, pp. 1-117 (p. 36) ("*Further, heavier policing in communities of color might lead to Black and Hispanic defendants being arrested, and later convicted, more often than White defendants who commit the same offense (Lum and Isaac, 2016). Poor outcome data might thus cause one to systematically underestimate the risk posed by White defendants. The second, related, issue is that our target of interest is a counterfactual outcome; it corresponds to what would have happened had a defendant been released*."). Italics added. See also: Balkin, Jack M./Siegel, Reva B., "The American Civil Rights Tradition: Anticlassification or Antisubordination?", *University of Miami Law Review*, Volume 58, 2003, pp. 9, 10-11 (explaining the distinction between the anticlassification and the antisubordination approaches to equality law).

[141] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

3. Why do predictive tools run counter to algorithmic fairness in the remit of criminal justice settings?

3.1. Why does algorithmic risk assessment (always) exact a toll on members of racialized communities? Algorithmic prediction lends itself to a reflecting mirror that projects the biased past into the future

As emphasized by esteemed scholars[142], the rationale behind the impossibility to achieve algorithmic equality by every metric when base rates vary greatly is straightforward: algorithmic prediction lends itself to a reflecting mirror that projects the biased past into the future. A key takeaway jumps out as ripe: bereft of a meaningful intervention to smooth out the ill-gotten prediction, history is doomed to repeat itself, which will, again, inordinately befall communities of colour. Just like in the past. So much for the sought-after, and much-acclaimed, algorithmic fairness[143] in risk assessment.

---

[142] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[143] See on the dire need to attain algorithmic fairness: Zajko, M, "Conservative AI and social inequality: conceptualizing alternatives to bias through social theory", *Artificial Intelligence and Society*, Volume 36, 2021, pp. 1047-1056; Dwork, Cynthia/Hardt, Moritz/Pitassi, Tonian/Reingold/Zemel, Richard, "Fairness Through Awareness, Proceedings of the 3rd Innovations", *Theoretical Computer Science Conference*, 2012, https://dl.acm.org/doi/10.1145/2090236.2090255 [https://perma .cc/T2U3-TNQ6] (discussing an individual fairness approach to algorithmic fairness); Meharabi, Ninareh/Morstatter, Fred/Saxena, Nripsuta/Lerman, Kristina/Galstyan, Aram, "A Survey on Bias and Fairness in Machine Learning", *ACM Computing Survey*, Jul. 2022; Fuster, Andreas/Goldsmith-Pinkham, Paul/Tarun Ramadorai/Walther, Angsar, "Predictably Unequal? The Effects of Machine Learning on Credit Markets", *Journal of Finance*, Volume 76, 2022, passim (providing a description of the variation in interest rates across racial groups carried out by machine learning algorithms); Zarsky, Tal, "Transparency in Data Mining: From Theory to Practice", *Discrimination and Privacy in the Information Society: Data Mining and Profiling in Large Databases*, Bart Custers et al. eds., Belin, Springer, 2013; Kroll, Joshua A. et al., "Accountable Algorithms", *University of Pennsylvania Law Review*, Volume 165, 2017, passim; Green, B./Hu L., "The Myth in the Methodology: Towards a Recontextualization of Fairness in Machine Learning", *Machine Learning Debates Work 35th Internation Conference of Machine Learning*, 2018, passim; West, SM./Whittaker, M./Crawford, K., "Discriminating systems gender, race, and power in AI", AI Now Institute, 2019; Xu, K./Nosek, B./Greenwald, AG., "Data from the race implicit

This assertion is firmly rooted in the fact that algorithmic predictions seek to identify patterns in past data and convert them into projections about future events. Strikingly, if there is a blatant racial disparity in the dataset[144], there will be inevitably racial disparity in the algorithmic prediction tool too[145]. As hinted above, it is possible to replace one form of inequality with another, but nonetheless impossible to wipe it out/scrub it out altogether with a snap of a finger. Let alone concomitantly. This finding about prediction is not unique to actuarial algorithmic methods[146]. Actuarial prediction mirrors a vividly clear glimpse of noticeable, palpable, tangible, detectable quantified data, whilst subjective prediction reflects a squalid/gloomy/fuggy/misty/smoggy image of anecdotal data[147]. Note, though, that both subjective and

---

association test on the project implicit demo website", *Journal of Open Psychology* Data 2:e3, 2014 (on the central tenets of algorithmic fairness).

144 Selbst, Andrew D., "Disparate Impact in Big Data Policing", *cit.*, at 123 (arguing that "[t]he words 'discrimination,' 'fairness,' and 'bias' evoke a family of related concepts"). But see: Barocas, Solon/Selbst, Andrew D., "Big Data's Disparate Impact", *California Law Review*, Volume 104, 2016, pp. 671-682 (discussing how data is often flawed, rigged, incomplete, imperfect, and therefore algorithms inherit the prejudice of the original decision makers). But see: Citron, Danielle Keats/Pasquale, Frank, "The Scored Society: Due Process for Automated Predictions", *Washington Law Review*, Volume 89, 2014, pp. 1, 4-5 (highlighting how human beings programming automated systems is directly conducive to, and leads, to fundamentally inaccurate results since the source code, predictive algorithms and datasets may encompass human biases which, in turn, have a disparate impact on historically disadvantaged groups); Crawford, Kate/Schultz, Jason, "Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms", *Boston College Law Review*, Volume 55, 2014, pp. 93, 99-101 (pinpointing pathways through which predictive analytic tools perpetuate discriminatory practices). But see: Bornstein, Stepanhie, "Antidiscriminatory Algorithms", *Alabama Law Review*, Volume 70, 2018, pp. 519, 524-528 (contending, however, that the tagged "facially neutral" algorithms producing discriminatory outcomes should be quashed).

145 Mayson, Sandra G., "Bias In, Bias Out", *cit.*, pp. 2218-2300.

146 Mayson, Sandra G., "Bias In, Bias Out", *cit.*, pp. 2218-2300.

147 Starr, Sonja B., "Evidence-Based Sentencing and the Scientific Rationalization of Discrimination." *Stanford Law Review*, 2014, 66:803-72; Kusner, MJ., Loftus, JR., "The long road to fairer algorithms", *Nature*, Volume 578, 2020, pp. 34-36.

algorithmic prediction alike look to the past as a portend for the future inasmuch past inequalities ineluctably ripple forward[148].

Strikingly, the gist of the problem lies not in algorithmic methodology. Notably, any form of prediction that heavily relies on data about the biased past[149] will inevitably create racial disparity[150] if the past data portrays the event that one aims to forecast[151] - the *target variable*[152] - transpiring with unequal rates across

---

[148] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[149] See: Fergus, Rachel, "Biased Technology: The Automated Discrimination of Facial Recognition", *ACLU Minnesota*, February 29, 2024, available at: https://www.aclu-mn.org/en/news/biased-technology-automated-discrimination-facial-recognition (access: 20.09.2024) ("*Especially when that technology – and law enforcement members using it - have been shown over and over to be biased toward marginalized groups*"). Italics added.

[150] Morley J, Kinsey L, Elhalal A et al, "Operationalising AI ethics: barriers, enablers, and next steps", *Artificial Intelligence and Society*, 2021, passim.

[151] See: Hamilton, Melissa, "Back to the Future: The Influence of Criminal History on Risk Assessments", *Berkeley Journal of Criminal Law*, Volume 20, 2015, pp. 75, 78 (exuding concerns with the recurrent use of criminal history in risk assessment owing to "the potential that criminal history is an unfortunate proxy for race and social disadvantage"); Barry-Jester, Anna Maria et al., "The New Science of Sentencing: Should Prison Sentencing Be Based on Crimes That Haven't Been Committed Yet?", *Marshall Project* (Aug. 4, 2015, 7:15 AM), https://www .themarshallproject.org/2015/08/04/the-new-science-of-sentencing [https://perma.cc /J4UW-BDKP] (arguing that risk-assessment outcomes gives rises to disparate racial impact that befall Blacks).

[152] It is important to note that other tools have other target variables, yet the analysis in this offering applies to several other target variables as well. In the pretrial purview, for example, risk-assessment tools also forecast "failure to appear," outlined in terms of data points that vary across jurisdictions. See: Mayson, Sandra G., "Dangerous Defendants", *Yale Law Journal*, Volume 127, 2018, pp. 509-13. There is also a spate of risk-assessment instruments that seek to forecast violence but in fact predict any hint or allegation of violence, whether it reflects in arrest or not (let alone conviction). See, e.g., Yang, Min et al., "The Efficacy of Violence Prediction: A Meta-Analytic Comparison of Nine Risk Assessment Tools", *Psychology Bulletin*, 2010, pp. 740-742 (arguing that "[t]he range of possible criterion variables for violence is wide, and "includes self-reports to third-party reports . . . , informal social service or police contact, formal contact or police charges, formal adjudication and court convictions, and incarceration"). See: Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

racial groups[153]. Coherently, if a machine learning algorithm's predictions are accurate at equal rates across racial lines[154], as were the COMPAS

[153] See: McIntyre, Frank/Baradaran, Shima, "Race, Prediction, and Pretrial Detention", *Journal Empirical Legal Studies*, Volume 10, 2013, pp. 742-759 (mentions black defendants' exceedingly high likelihood of being arrested on drug-related charges as a potential cause of the race gap); Skeem, Jennifer L./Lowenkamp, Christopher T., "Risk, Race, and Recidivism: Predictive Bias and Disparate Impact", *Criminology*, Volume 54, 2024, pp. 680, 683-84, 704-06 (argues that criminal history correlates with race in their respective data set). See also: Pager, Devah, *Marked: Race, Crime, and Finding Work in an Era of Mass Incarceration*. Chicago: University of Chicago Press, 2007; Fornili, K. S., "Racialized mass incarceration and the war on drugs: A critical race theory appraisal", *Journal of Addictions Nursing*, Volume 29, Issue 1, 2018, 65-72; Jacoby, S. F., Dong, B., Beard, J. H., Wiebe, D. J., & Morrison, C. N., "The enduring impact of historical and structural racism on urban violence in Philadelphia", *Social Science & Medicine*, 1982/2018, 199, 2018, pp. 87-95; Travis, Jeremy, Bruce Western, and Steve Redburn (Eds.), *Growth of Incarceration in the United States: Exploring Causes and Consequences*, Washington, DC: National Academy of Science, 2014. See: Bar-Gill/Oren/Warren, Elizabeth, "Making Credit Safer", *University of Pennsylvania Law Review*, Volume 157, 2008, pp. 1-66 (argue that Black defendants are far more likely to be incarcerated, arrested, surveilled and harassed than White folks, which leads to the conclusion that the use of credit scores and income presents another way in which credit decisions leverage on structural disadvantage).

[154] Dieterich, William et al., "COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity", *Northpointe Inc*. 1, 2-3, 8-13 (July 8, 2016), http://go.volarisgroup.com/rs/430 -MBX-989/images/ProPublica_Commentary_Final_070616.pdf [https://perma.cc/K4GM -RBQY] (stating that a predictive instrument is bigoted/flawed only if a certain score, or a given classification, means a starkly likelihood of the forecasted outcome for members of one racial group than members of the other group). See also: Flores: Anthony W. et al., "False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks," *Federal Probation*, Volume 80, 2016, pp. 38-40 (positing that "well-established and accepted standards exist to test for bias in risk assessment"); Skeem, Jennifer L./Lowenkamp, Christopher T., "Risk, Race, and Recidivism: Predictive Bias and Disparate Impact", *cit*., p. 685 (asserting that if "a given score [has] the same meaning regardless of group membership," the instrument is "unbiased").

forecasts in Broward County, any inequality in prediction inevitably mirrors inequality[155] in the underlying dataset[156].

Relatedly, to grasp and remedy disparity in prediction, it is therefore necessary to thoroughly understand how and when racial inequality emerges/sprouts in the data that we shift our gaze to as a reflection of past crime[157]. In addition to that, the deployment of race-aware algorithms[158] is also a remedy to be kept front of mind if one endeavours to weed out algorithmic unfairness from the field of algorithmic risk assessment.

Bearing this backdrop very firmly in mind, «playing with the data»[159] entails keeping a «human in the loop».[160] Which is nothing but

---

[155] Stevenson, Megan/Mayson, Sandra G., "Pretrial Detention and Bail", Luna, Erik, (Editor), *Reforming Criminal Justice: A Report of the Academy for Justice, Bridging the Gap between Scholarship and Reform*, pp. 21, 34-39 (assessing the risk of pretrial risk assessments while highlighting a slew of deep-rooted concerns regarding accuracy, racial-equality, and procedure); See: Chandler, Anupam, "The Racist Algorithm", *Michigan Law Review*, Volume 115, n.º 6, 2018, pp. 1023-1025 (arguing that the real-world facts on which algorithms used in criminal justice risk assessment are based are "deeply suffused with invidious discrimination"); Eaglin, Jessica M., "Constructing Recidivism Risk", *cit*., pp. 94-99 (argues that risk assessment has every chance to "compromise[e] equality"); Mayson, Sandra G., "Dangerous Defendants", *cit*., pp. 494-96.

[156] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[157] Mayson, Sandra G., "Bias In, Bias Out", *cit*., pp. 2218-2300.

[158] On the dire need to achieve algorithmic fairness through the creation of both race-aware algorithms and meaningful trade-offs to attain fairer risk scores: Kim, Pauline, "Race-Aware Algorithms: Fairness, Non-discrimination and Affirmative Action" *California Law Review*, 2022; Kleinberg, Jon/S. Mullainathan/M. Raghavan, "Inherent Trade-offs in the Fair Determination of Risk Scores", *Proceedings of the 8th Conference on Innovations in Theoretical Computer Science (ITCS)*, 2017; Kroll, Joshua A./J. Huey/S. Barocas/E. W. Felten/J. R. Reidenberg/D. G. Robinson/and H. Yu, "Accountable Algorithms", *University of Pennsylvania Law Review*, 165 (3), 2017, pp. 633-705.

[159] Lehr, David/Ohm, Paul, "Playing with the Data: What Legal Scholars Should Learn About Machine Learning", *University of California, Davis*, Volume 51, 2017, pp. 653-718.

[160] Arguing that humans should remain "in the loop" at the output-end of algorithmic decision-making: Saxena, Devansh/Badillo-Urquiola, Karla/Wisniewski, Pamela J./Guha, Shion, "A Human-Centered Review of the Algorithms Used Within the U.S. Child Welfare System", *2020 Proceedings of Conference on Human Factors in Computing Systems*, Apr. 2020, 2020, passim.

formulaic: deep-rooted concerns have been accordingly voiced regarding the perils of trusting too much on machines. The next section will add plausibility, and give credence, to this main contention.

4. Robot-judges – and algorithmic risk assessment - must not enter the courtroom or inordinate harms might otherwise occur: defendants must be both granted the right to a human decision in the realm of criminal procedural law and the procedural chance to challenge the outputs generated by algorithmic decision-making

Against this background, algorithmic risk assessment in the realm of criminal procedure law – i.e. COMPAS - is not without insurmountable drawbacks, algorithmic fairness wise. That is exactly why COMPAS should not be deployed to decide who should go to jail and who should walk free in criminal procedure law settings. For that reason alone, robot-judges should not be entrusted with the task of making high-stake decisions in

---

See: Salloch, S., & Eriksen, A., "What are humans doing in the Loop? Co-reasoning and practical Judgment when using machine learning-driven decision aids", *The American Journal of Bioethics*, 2024, pp. 1-12; Binns, Reuben, "Human Judgment in Algorithmic Loops: Individual Justice and Automated Decision-Making", *Regulation & Governance*, Volume 16, 2022, passim; Simon & Schuster, Green, B., "Algorithmic realism: expanding the boundaries of algorithmic thought", *Proceedings of the 2020 Conference on Fairness Accountability and Transparency*, 2020, passim; Green, B./Hu L., "The Myth in the Methodology: Towards a Recontextualization of Fairness in Machine Learning", *Machine Learning Debates Work 35th Internation Conference of Machine Learning*, 2018, passim; Green, B., Chen Y, "Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments", *FAT* 2019, Proceedings of 2019 Conference of Fairness, Accountability, Transparency*, 2019, pp. 90-99 (on the central tenets of algorithmic fairness). See also: Alimardani, Armin/Jane, Emma A., "Human Hallucinations: GenAI and the Mirage of Personalised Learning for All", 2024, p. 13 ("*The output generated by GenAI models includes an inherent randomness and the same prompt may result in various outputs. Therefore, it is likely that the outputs of GenAI models are never entirely accurate and reliable. In other words, we contend that existing GenAI systems represent a technology that requires a 'human in/on the loop' to verify outputs before they reach the end user. When incorporating GenAI into an educational chatbot, students are the end users, making it challenging to identify and rectify these errors promptly.*"). Italics added.

the remit of criminal justice. Regardless, it is often ventured that, in the forthcoming future, robot judges will write better judicial decisions[161] than human judges[162]. But here is the kicker: human judges[163], unlike robot-judges, are fully equipped to make sound value judgments[164]. No surprise stems from the fact that value judgments inhere and permeate judicial decision-making[165]. But here is the twist: Artificial Intelligence, if ever enters the courtroom[166], must rely on human value judgments

---

[161] Kozlov, Yuri/Shutova, Maria/Bajwa, Taaha, *Automated Judge is Not a Task For LegalTech But For DeepTech*, 24th February of 2025, 2025, p. 1 ("*Unlike predictors, JudgeAI replicates legal reasoning with 96% accuracy in real cases*, offering transparent, *logic-driven decisions through algorithmic evidence analysis and behavior modelling*"). Italics added.

[162] Volokh, Eugene, "Chief Justice Robots", *Duke Law Journal*, Volume 68, 2019, pp. 1135 and ff and passim ("*this Essay argues, the same technology can be used to create AI judges, judges that we should accept as no less reliable (and more cost-effective) than human judges. If the software can create persuasive opinions, capable of regularly winning opinion-writing competitions against human judges—and if it can be adequately protected against hacking and similar attacks—we should in principle accept it as a judge, even if the opinions do not stem from human judgment*"). Italics added.

[163] See: Zou, Mimi/Leffley, Ellen, "Generative Artificial Intelligence and Article 6 of the European Convention on Human Rights: The Right to a Human Judge?", Mimi Zou, Martin Ebers, Cristina Poncibò and Ryan Calo (Editors), *The Cambridge Handbook of Generative AI and the* Law, Cambridge, Cambridge University Press 2025, passim (championing for a "right to a human judge").

[164] See: Davis, Joshua P., "Of Robolawyers and robojudges", *Hastings Law Journal*, Volume 73, Issue 5, 2022, pp. 1176-1201 (1198). See also: Davis, Joshua P., *Unnatural Law: AI, Consciousness, Ethics, and Legal Theory*. Cambridge: Cambridge University Press, 2023.

[165] Davis, Joshua P., "Of Robolawyers and robojudges", *cit*., pp. 1176-1201 (1198).

[166] See: Grimm, Paul V./Grossman, Maura R./Gless, Sabine/Hildebrant, Mireille, "Artificial Justice: The Quandary of AI in the Courtroom", *Judicature International*, September 2022, Bolch Judicial Institute at Duke Law, Duke University School of Law, 2022, p. 1 ("*what happens when machine learned and AI-generated data enter the courtroom*? Should that evidence be considered reliable?"). Italics added. See on the multifarious issues arising out of evidence produced by Artificial Intelligence-powered devices in criminal trials: Gless, Sabine/Lederer, Fredric/Weigend, Thomas, "AI-Based Evidence in Criminal Trials? AI-Based Evidence in Criminal Trials?", *Tulsa Law Review*, Volume 59, Issue 1, 2024, p. 35 ("*Criminal courts therefore face the question of whether to admit various kinds of device data offered as evidence*"). Italics added.

embedded in past judicial decisions[167] to generate outputs[168]. And these are sobering news.

Compellingly, and not in the least worryingly, rigged past data[169] may contain deep-rooted human biases that we may not wish to re-entrench into computer-generated algorithms[170] that will decide our future[171]. But much more lies beyond that[172].

Whilst Artificial Intelligence excels in some time-consuming legal tasks[173] and performs uncannily well in a sheer number of recreational

---

[167] Davis, Joshua P., "Of Robolawyers and robojudges", *cit*., pp. 1176-1201 (1198).

[168] But see: Schwarcz, Daniel/Sam Manning/Patrick Barry/David R. Cleveland/JJ Prescott/Beverly Rich, "AI-Powered Lawyering: AI Reasoning Models, Retrieval Augmented Generation, and the Future of Legal practice", 2025, pp. 1-2 ("*We find that both AI tools significantly enhanced legal work quality, a marked contrast with previous research examining older large language models like GPT-4*"). Italics added.

[169] See: Barocas, Solon/Selbst, Andrew D., "Big Data's Disparate Impact", *California Law Review*, Volume 104, 2016, pp. 671-682 (discussing how data is often flawed, rigged, incomplete, imperfect, and therefore algorithms inherit the prejudice of the original decision makers). But see: Citron, Danielle Keats/Pasquale, Frank, "The Scored Society: Due Process for Automated Predictions", *Washington Law Review*, Volume 89, 2014, pp. 1, 4-5 (highlighting how human beings programming automated systems is directly conducive to, and leads, to fundamentally inaccurate results since the source code, predictive algorithms and datasets may encompass human biases which, in turn, have a disparate impact on historically disadvantaged groups).

[170] Balkin, Jack M., "The Three Laws of Robotics in the Age of Big Data", *Ohio State Law Journal*, Volume 78, Issue 5, 2017, pp. 1217-1241.

[171] *Tech, Bias and Housing Initiative, the Promise and Perils of Residential Proptech, Year 1, Research Summary Report, April 2023, TechEquity Collaborative*, p. 11, available at: https://techequity.us/wp-content/uploads/2023/04/TBHI-Y1-Research-Summary-Report.pdf (28.11.2024). ("*Once bad data enters the analysis of one factor, it gets baked into the subsequent process, multiplying the effect that one disadvantageous data point can have on someone's overall housing determination*"). Italics added.

[172] For instance, Artificial Intelligence, with the right settings, can classify relevant case law. See: Sargeant Holli/Izzidien, Ahmed/Steffek, Felix, "Topic classification of case law using a large language model and a new taxonomy for UK law: AI insights into summary judgment", *Artificial Intelligence and Law*, 2025, p. 1.

[173] Nielsen, Aileen/Stavroula Skylaki/Milda Norkute/Alexander Stremitzer, "Building a better lawyer: Experimental evidence that artificial intelligence can increase legal work efficiency", *Journal of Empirical Legal Studies*, Volume

chores[174], it nonetheless falls noticeably short of capturing the nuances underpinning each lawsuit[175] that find its way to the criminal courtroom. Compellingly, there is no such thing as two identical lawsuits in judicial settings (except made, perhaps, to class actions).

Unable to grasp this ground truth, and bereft of an entrenched pattern to build upon, robot-judges, robot-prosecutors[176] or robot-lawyers[177] will either provide a downright bogus opinion[178] or a plausible, yet inaccurate, one. Which brings me to the second line of reasoning against the deployment of robot-judges to decide lawsuits in criminal courts.

21, 2024, pp. 979-1022 (979) ("*Our results show that AI support can dramatically increase the efficiency of legal task completion*, but finding the optimal form of AI assistance is a fine-tuning exercise".). Italics added.

[174] Fellin, Teppo/Hollweg, Mathias, Theory Is All You Need: AI, Human Cognition, and Decision Making, *cit*., pp. 1-53 (3) ("*Artificial intelligence (AI) now matches or outperforms humans in any number of games, standardized tests, and cognitive tasks that involve high-level thinking and strategic reasoning*"). Italics added.

[175] Scholars have not been slow to pinpoint the pressing challenges arising out of the deployment of Artificial Intelligence in the courtroom. See: Gless, Sabine, "AI in the Courtroom: A Comparative Analysis of Machine Learning in Criminal Trials", *Georgetown Journal of International Law*, Volume 51, 2020, p. 250 ("*Where machine evidence is proffered as evidence in a criminal trial, it must be adequately contextualized and tested for reliability. Such evidence—just like human testimony—is not infallible*"). Italics added.

[176] Anderson, Stephen E., "Should Robots Prosecute and Defend?", *Oklahoma Law Review*, Volume 72, Number 1, 2019, pp. 1-20 (1) ("*Even when we achieve the 'holy grail' of artificial intelligence—machine intelligence that is at least as smart as a human being in every area of thought—there may be classes of decisions for which it is intrinsically important to retain a human in the loop*"). Italics added.

[177] Anderson, Stephen E., "Should Robots Prosecute and Defend?", *cit*., pp. 1-20 (1) ("*Thus, while many details need to be worked out, we might within decades have a criminal justice system consisting of robo-defense lawyers and robo-prosecutors*. And even if we never do, their consideration *provides another lens through which to consider these roles and, ultimately, our criminal justice system*"). Italics added.

[178] See: Wachter, Sandra, "Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond", *Yale Journal of Law & Technology*, Volume 26, Issue 3, 2024, pp. 671-672 ("*Generative AI (GAI) creates hallucinations and inaccurate or harmful information, which can lead to misinformation, disinformation, and the erosion of scientific knowledge*"). Italics added.

It is pivotally important to aver that human behaviour is shaped by underlying values, whereas machines 'predictions are shaped by patterns with an entrenched presence in the dataset. Which prompts the vexing question: why must not machines be entrusted with the task of making high stake decisions in criminal procedure law settings? Machines, unlike human beings, have no human values[179] (e.g. sympathy, empathy, judicial independence, forbearance, lenience). Instead, they operate through deep-seated patterns and entrenched optimal predictive rules that run counter to the central tenets underpinning criminal procedure law settings.

For that reason alone, machines should not be entrusted with the task of making high-stake decisions (e.g. those pertaining to fundamental rights, such as freedom) in criminal procedure law settings. Defendants must therefore be afforded the right to a human decision-making in criminal justice settings. Furthermore, defendants must also be awarded the right to challenge the outputs generated by artificial intelligence-embedded technologies in the compass of algorithmic risk assessment.

To the extent that Rule of Law is tightly interlocked to procedural[180] concerns[181] (Procedural Rule of Law[182]), defendants must be thereupon

---

[179] See: Sourdin, Tania, *Judges, Technology and Artificial Intelligence: The Artificial Judge*, Cheltenham: Edward Elgar, 2021, p. 257 ("*it is inevitable that a values-based approach must be incorporated into such a framework ... this should reflect the values underpinning the judicial system within a particular jurisdiction*"). Italics added.

[180] Tashima, A. Wallace, "The War on Terror and the Rule of Law", *Asian American Law Journal*, 15 2008, pp. 245-263 (echoing the same procedural approach to the Rule of Law while asserting that independent and impartial legal officials are needed to uphold the Rule of Law).

[181] Waldron, Jeremy, "The Rule of Law and the Importance of Procedure", *New York Public Law and Legal Theory*, 2010, p. 4 (noting that "the Rule of Law is violated when due attention is not paid to these procedural matters or when the institutions that are supposed to embody these procedure are undermined or interfered with". Further down this author would admit that there are some substantive dimensions of Rule of Law, such as respect for right for property and presumption of liberty and so forth).

[182] See: Santos, Hugo Luz dos, *Controllable Artificial Intelligence and the Future of Law*, *cit*., passim (Chapter V, Part V).

allotted the procedural chance («technological due process»[183]) to challenge the legal decisions[184] arisen from biased algorithmic risk assessment.

Concluding remarks

1. The rationale behind the impossibility to attain algorithmic equality by every metric when base rates vary greatly is straightforward: algorithmic prediction lends itself to a reflecting mirror that projects the biased past into the future. A key takeaway jumps out as ripe: bereft of a meaningful intervention to falloff the ill-gotten prediction, history is doomed to repeat itself, which will, again, inordinately befall communities of colour. Just like in the past.

2. As a result, algorithmic fairness in algorithmic risk assessment is set to dispiritingly dwindle. This line of reasoning sits easily with the mounting scepticism amongst pundits about the convenience (or lack thereof) of the wide-ranging stance of debarring protected markers, such as race and gender.

3. This assertion is firmly rooted in the fact that algorithmic predictions seek to identify patterns in past data and convert them into projections about future events. Strikingly, if there is a blatant, and rampant, racial disparity in the dataset there will be inevitably racial disparity in the algorithmic risk assessment too.

4. Strikingly, the legal quagmire of the problem lies not in the oft-trotted out algorithmic methodology. Compellingly, any form of prediction that heavily relies on data about the biased past will inevitably create racial disparity if the past data portrays the event that one aims to

---

[183] Citron, Danielle K., "Technological Due Process", *Washington University Law Review*, Volume 85, 2008, pp. 1249-1253 ("computer programs seamlessly combine rulemaking and individual adjudications without the critical procedural protections owed either of them").

[184] Taekema, Sanne, "The Procedural Rule of Law: Examining Waldron´s Argument on Dignity and Agency", *Jahrbuch für Recht und Ethik*, Band 21, (2013): 133 ff (noting that "Waldron´s understanding of the Rule of Law as procedural which demands the opportunity for individuals to contest legal decisions in the formalized processes of courts and tribunals").

forecast - the *target variable* - transpiring with unequal rates across racial groups in criminal justice settings.

5. Coherently, if a machine learning algorithm's predictions are accurate at equal rates across racial lines, as were the COMPAS forecasts in Broward County, any inequality in prediction inevitably mirrors inequality in the underlying dataset[185], which, in turn, will inordinately befall on members of racialized communities while wobbling the foundations upon which stands algorithmic fairness in criminal procedure law settings.

6. A key takeaway accordingly jumps out as ripe in this regard: AI-embedded technologies, if nothing else, are prone to reify, further compound and exacerbate deep-seated social and ethnic concerns with which modern societies have been scuffling for a long-haul, such as misogyny, racism, xenophobe, and political hatred which, taken together, are rife for misfire, social outcry, uproar, and a plethora of grievances alike. *Algorithmic Dictatorship*[186] is a living testament of such gruesome reality.

7. *Algorithmic Fairness* is therefore sorely needed – paired with tools like race-aware machine learning algorithms - to assuage the darkest patterns of Artificial Intelligence-embedded technologies arising out of the dingiest corners of the digital reality.

8. Another major finding arising out of this paper is that which Artificial Intelligence-embedded technologies fall noticeably short of capturing the nuances underpinning each lawsuit that find its way to a criminal courtroom.

9. Unable to grasp this ground truth, and bereft of an entrenched pattern to build upon, robot-judges will either provide a downright bogus opinion or a plausible, yet inaccurate, one. Which brings me to the second line of reasoning against the deployment of robot-judges to decide lawsuits in criminal courts.

10. It is pivotally important to aver that human behaviour is shaped by underlying values, whereas machines 'predictions are borne out by deep-seated patterns with an entrenched presence in the dataset. Which

---

[185] Mayson, Sandra G., "Bias In, Bias Out", *cit.*, pp. 2218-2300.

[186] See: Santos, Hugo Luz dos, *Controllable Artificial Intelligence and the Future of Law*, *cit.*, passim (Chapter IV, Part IV).

prompts the vexing question: why must not machines be entrusted with the task of making high stake decisions in criminal procedure law settings?

11. Machines, unlike human beings, have no human values (e.g. sympathy, empathy, judicial independence, forbearance, lenience). Instead, they operate through deep-rooted patterns and embedded optimal predictive rules. Criminal procedure law settings should not - must not - be largely bereft of human values that render criminal justice human or avoidable miscarriages of justice are otherwise slated to occur.

12. For that reason alone, machines should not be entrusted with the task of making high-stake decisions (e.g. those pertaining to fundamental rights such as freedom) in criminal procedure law settings. Defendants must therefore be granted the right to a human decision-making in criminal justice settings.

13. To be fully compliant with the central tenets of algorithmic fairness underpinning both criminal procedure law (in general) and algorithmic risk assessment (in particular), a procedural chance must be given to the defendants as to challenge the biased outputs arising out of algorithmic risk assessment or avoidable miscarriages of justice (e.g. wrongful arrests, wrongful convictions, biased outputs conducive to the defendant´s constant harassment by law enforcement agencies) may otherwise occur. This is the kernel of the much-touted Procedural Rule of Law of which relevant stakeholders must not be unbeknownst of.

References

Adadi, Amina/Berrada, Mohammed, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI), Institute of Electrical and Electronics Engineers, Volume 6, 2018, passim.

Agrawal, Ajay/Gans, Joshua/Goldfarb, Avi, Prediction Machines: The Simple Economics of Artificial Intelligence, 2nd Edition. Massachusetts: Harvard Business Review Press, 2022.

Alimardani, Maryam/Hiraki, Kazuo, "Passive Brain-Computer Interfaces for Enhanced Human-Robot Interaction", Frontiers of Robotics and Artificial Intelligence, Volume 7, 2020, pp. 1 and ff and passim.

Alves, Jones Figueirêdo/Pimentel, Alexandre Freire, "Breves notas sobre os preconceitos decisionais judiciais produzidos por redes neurais artificiais (Brief notes about the judicial decisional prejudices produced by artificial neural networks), *Lisbon Law Review*, Thematic Issue: Law and Technology, Year LXII, Numbers 1 and 2, 2022, pp. 555-577.

Ashley, Kevin D., Artificial Intelligence and Legal Analytics. New Tools for Law Practice in the Digital Age. Cambridge: Cambridge University Press, 2018.

Avbelj, Matej, "The Rule of Law, Comprehensive Doctrines, Overlapping Consensus, and the Future of Europe", Ratio Juris, Volume 36, Issue 3, September, 2023, pp. 242–258.

Avery, Mallory/Leibbrandt, Andreas/Vecci, Joseph, "Does Artificial Intelligence Help or Hurt Gender Diversity? Evidence from Two Field Experiments on Recruitment on Tech, Monash University, Monash Business School, 2023, pp. 1-70.

Baller, Stéphane/Deffains, Bruno, "Intelligence artificielle et devenir de la professions d´avocat ; l´avenir est présent », Revue pratique de la prospective et d'innovation, Volume 1, 2018, pp. 14 and ff.

Bar-Gill, Oren/Sunstein, Cass R./Talgam-Cohen, Inbal, "Algorithmic Harm in Consumer Markets", Journal of Legal Analysis, Volume 15, n. º 1, 2023, pp. 1-47.

Bathae, Yavar, "The Artificial Intelligence Black Box and the Failure of Intent and Causation", Harvard Journal of Law and Technology, n.º 37, 2018, p. 901.

Beriain, Iñigo de/Estrada/ Pérez, M. Josune, "La inteligencia artificial en el proceso penal español: un análisis de su admisibilidad sobre la base de los derechos fundamentales implicados", Revista de derecho UNED (RDUNED), Volume 25, 2019, pp. 531-561.

Bibas, Stefanos, "Foreword. Prosecutors' Changing Roles at the Hub of Criminal Justice", passim, The Oxford Handbook of Prosecutors and Prosecution, Wright, Ronald F./ Levine, Kay L./Gold, Russell M. (Editors), Oxford: Oxford University Press, 2021, pp. 1-654.

Blattner, Laura/Nelson, Scott/Spiess, Jann, Unpacking the Black Box: Regulating Algorithmic Decisions. Working Paper. Redwood City, California: Stanford University Press, 2021.

Boden, Margaret A., Artificial Intelligence: A Very Short Introduction. Oxford: Oxford University Press, 2018.

Borgesius, FJ Zuiderveen, "Strengthening legal protection against discrimination by algorithms and artificial intelligence", International Journal of Human Rights, Volume 24, Issue 10, 2020, pp. 1573 and ff and passim.

Bornet, Pascal/Barkin, Ian/Wirtz, Jochen, Intelligent Automation: Learn how to harness Artificial Intelligence to boost business and make out world more human, 2020.

Bornstein, Stephane, "Antidiscriminatory Algorithms", Alabama Law Review, Volume 70, 2018, pp. 519-570.

Brennan-Marquez/Henderson, Stephen E., "Artificial Intelligence and Role-Reversible Judgment", Journal of Criminal Law and Criminology, Volume 109, 2019, pp. 137-163.

Brigant, Jean-Marie, "Les risques accentués d´une justice pénale prédictive », Archives de Philosophie du Droit, Volume 60, 2018, pp. 238 and ff and passim.

Brownsword, Roger, Law, Technology and Society: Re-Imagining the Regulatory Environment. London: Routledge, 2020.

Burk, Dan L., "Algorithmic Fair Use", University of Chicago Law Review, Volume 86, 2019, pp. 283-288.

Calo, Ryan, "Artificial Intelligence Policy: Primer and Roadmap", U.C. Davis Review, Volume 51, 2018, pp. 309-404.

Calo, Ryan, "Robotics and the Lessons of Cyberlaw", California Law Review, Volume 103, 2015, pp. 512 ff.

Cataleta, Maria Stefania, "Artificial Intelligence vs Human Intelligence", in: Martin, L. Miraut/Zalucki, M. (editors), Artificial intelligence and Human Rights, Dickinson eBook. Krakow: AFM Krakow University, 2021, pp. 117-127.

Cataleta, Maria Stefania/Cataleta, Anna, "Artificial Intelligence and Human Rights: An Unequal Struggle", CIFILE Journal of International Law, Volume 1, N. º 2, 2020.

Chabert, J.L. et al, A History of Algorithms: From the Pebble to the Microchip. New York/Heidelberg: Springer, 2013.

Chandler, Anupam, "The Racist Algorithm", Michigan Law Review, Volume 115, n. º 6, 2018, p. 1026.

Chasemi et al, Medhi, "The application of Machine Learning to a General Risk-Need Assessment Instrument in the Prediction of Criminal Recidivism", Criminal Justice and Behavior, Volume 48, 2020, pp. 518-538.

Citron, Danielle Keats, "Technological Due Process", Washington University Law Review, Volume 85, 2008, pp. 1256-1257.

Corbett-Davies, Sam/Goel, Sharad, "The Measure and Mismeasure of Fairness: a critical Review of Fair Machine Learning", Cornell University Library, 2018, pp. 3 and ff and passim.

Crootof, Rebecca, "Cyborg Justice" and the Risk of Technological-Legal Lock-In", Columbia Law Review, Volume 119, 2019, pp. 233 and ff.

Davis, Joshua P. "Law Without Mind: AI, Ethics, and Jurisprudence", California Western Law Review, Volume 55, 2018, pp. 181 e ss and passim.

Davis, Joshua P., "Of Robolawyers and robojudges", Hastings Law Journal, Volume 73, Issue 5, 2022, pp. 1176-1201.

Davis, Joshua P., Unnatural Law: AI, Consciousness, Ethics, and Legal Theory. Cambridge: Cambridge University Press, 2023.

Dazeley, Richard/Vamplew, Peter/Foale, Cameron/ Young, Charlotte/ Aryal, Sunil/ Cruz, Francisco, "Levels of explainable artificial intelligence for human-aligned conversational explanations", Artificial Intelligence, Volume 299, 2021, passim.

Deffains, Bruno, "Le monde du droit face à la transformation numérique", Revue Française d´Études Constitutionnelles et Politiques, Volume 170, 2019, pp. 49 and ff and passim.

Dino, Dylan, "The Rule of Law and the Rule of Empire: A. V. Dicey in Imperial Context", The Modern Law Review, Volume 81, Issue 5, 2018, pp. 739-764 (739-741).

Dixon, H. B., "The Evolution of a High Technology Courtroom", Future Trends in State Courts. Virginia: National Center for State Courts, 2011.

Dockrill, Peter, "Brain Implant Translates Paralyzed Man´s Thoughts into Text With 94 % Accuracy", SCI Alert, 2021, passim.

Dong et al, Qi,, "Imbalanced deep learning by minority class incremental rectification, IEEE Transactions on Pattern Analysis & Machine Intelligence, Volume 41, 2019, p. 1367.

Donoghue, Jane, "The Rise of Digital Justice: Courtroom Technology, Public Participation and Access to Justice", The Modern Law Review, Volume 80, Issue 6, 2018, p. 995.

Edwards, Lilian/Veale, Michael, "Slave to the Algorithm? Why a "Right to an Explanation" is Probably not the Remedy You are Looking for", Duke Law and Technology Review, Volume 16, 2018, pp. 18-67.

Engelhart, Marc, Sanktionierung von Unternehmen und Compliance – Eine rechtsvergleichende Analyse des Straf-und Ordnungswidrigkeitenrechts in Deutschland und den USA, 2. Auflage, Schriftenreihe des Max-Planck-Instituts für ausländisches und internationales Strafrecht, Reihe S: Strafrechtliche Forschungsberichte (MPIS), Band 121, Berlin: Duncker & Humblot, 2012, pp. 284-290.

Eubanks, Virginia, Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor. New York: St. Martin's Press, 2018.

Ferey, Samuel, « Analyse économique du droit, big data et justice prédictive », Archives de Philosophie du Droit, Volume 60, 2018, pp. 68-81.

Griemmelmann, James/Westreich, Daniel, "Incomprehensible Discrimination ", California Law Review Online, Volume 7, 2018, pp. 164-171.

Grimm, Paul W., "Practical Considerations for the Admissibility of Artificial Intelligence Evidence", Maryland Bar Journal, Volume 2, Issue 3, 2021, pp. 39 and ff and passim.

Hamilton, Melissa, "The Biased Algorithm: Evidence of Disparate Impact on Hispanics", American Criminal Law Review, Volume 56, 2019, pp. 1553 and ff and passim.

Hanke, Philip, "Computers with Law Degrees? The Role of Artificial Intelligence in Transnational Dispute Resolution, and Its Implications of the Legal Profession", Transnational Dispute Management, Volume 14, Issue 2, 2018, passim.

Huq, Aziz Z., "Racial Equity in Algorithmic Criminal Justice", Duke Law Journal, Volume 68, 2019, pp. 1043-1072.

Ifeoma, Ajunwa, "The Paradox of Automation as Anti-Bias Intervention", Cardozo Law Review, Volume 41, 2020, pp. 1672, 1726-1727.

Irti, Natalino, Un Diritto Incalcolabile. Torino: G. Giappichelli Editore, 2016, pp. 20 and ff and passim.

Jackson, Maya C., "Artificial Intelligence and Algorithmic Bias: The Issues with Technology Reflecting History and Humans", Journal of Business Technology Law, Volume 16, 2021, pp. 299-316.

Katyal, Sonia, "Private Accountability in the Age of Artificial Intelligence", UCLA Law Review, Volume 66, 2019, pp. 54-99.

Kleinberg, Jon, et al, "Discrimination in the Age of Algorithms", Journal of Legal Analysis, 2018, pp. 114-174.

Kleinberg, Jon, et al, "Algorithms as Discrimination Detectors", Proceeds of the National Academy of Science, Volume 117, 2020, pp. 30096-30110.

KÖCK, Elisabeth, "Nemo-tenetur-Grundsatz für Verbände", Festschrift für Manfred Burgstaller zum 65. Geburtstag, Graft, Christian/Medigovic, Ursula, (Hrsg.). Wien/Graz: Neuer Wissenschaftlicher Verlag, 2004, pp. 267-279.

Mayson, Sandra G., "Bias In, Bias Out", Yale Law Journal, Volume 128, 2019, pp. 2218 and ff and passim.

O'Neill, C., Weapons of Math Destruction. London: Penguin Random House, 2016.

Okidegbe, Ngozi, "The Democratizing Potential of Algorithms?", Connecticut Law Review, Volume 54, 2021, passim.

Osaba, Osonde/Welser, William IV, An Intelligence in Our Image – The Risk of Bias and Errors in Artificial Intelligence, Santa Monica, California, Rand, 2018.

Pagallo, Ugo/Quattrocolo, Serena, "The Impact of AI on Criminal Law and its twofold procedures", Barfield, W./Pagallo, Ugo (Editors), Research Handbook on the Law of Artificial Intelligence, 2018. Cheltenham: Edward Elgar, 2018, pp. 385-409.

Pasquale, Frank, "A Rule of Persons, Not Machines: The Limits of Legal Automation", George Washington Law Review, Volume 87, 2019, pp. 1 and ff and passim.

Pasquale, Frank, The Black Box Society: The Secret Algorithms that Control Money and Information. Cambridge, Massachusetts: Harvard University Press, 2015.

Queck, Nadine, Die Geltung des Nemo-tenetur-Grundsatzes Die Geltung des nemo-tenetur-Grundsatzes zugunsten von Unternehmen. Berlin: Duncker and Humblot, 2005.

Samek, Wojciech/Müller, Klaus-Robert, "Explainable AI: interpreting, explaining and visualizing deep learning", Towards Explainable Artificial Intelligence, Samek, Wojciech/Montavon, Grégoire/ Vedaldi, Andrea/Hansen, Lars Kai/ Müller, Klaus-Robert (Editors). Heidelberg/Berlin/New York: Springer, 2019, pp. 5-22.

Santos, Hugo Luz dos, Inteligência Artificial e Processo Penal. Braga: NovaCausa Edições Jurídicas, 2022.

Santos, Hugo Luz dos Santos, Towards a Four-Tiered Model of Mediation. New York: Springer Nature, 2023.

Santos, Hugo Luz dos/Leong, Cheng Hang, "Culture Matters": Expedited Arbitration and Arb-Med in Macau", Hong Kong Law Journal, Volume 54:3, 2024.

Santos, Hugo Luz dos, Multidisciplinary Dynamics of Mediation. New York: Springer Nature, 2025a, Volume I.

Santos, Hugo Luz dos, Multidisciplinary Dynamics of Mediation. New York: Springer Nature, 2025b, Volume II.

Santos, Hugo Luz dos, Controllable Artificial Intelligence and the Future of Law (Artificial Intelligence and the Rule of Law Series). New York: Springer Nature, 2025c.

Sourdin, Tania, Judges, Technology and Artificial Intelligence: The Artificial Judge. Cheltenham: Edward Elgar, 2021.

Surden, Harry, "Machine Learning and Law", Washington Law Review, Volume 89, 2015, pp. 87 and ff.

Susskind, Richard, Tomorrow's Lawyers: An Introduction to your Future, 2d edition. Oxford: Oxford University Press, 2018.

Susskind, Richard, "The Future of Courts", Remote Courts, Volume 6, n.º 5, July/August 2020, 2020, pp. 1 and ff and passim.

Wachter, S/Mittelstadt. B., "A right to Reasonable Inferences: Re-thinking Data Protection Law in the Age of Big Data and AI", Columbia Business Law Review, 2019, pp. 1 and ff and passim.

Watcher, Sandra/Mittelstadt, Brent/Russell, Chris, "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR", Harvard Journal of Law and Technology, Volume 31, 2018, pp. 841 e ss and passim.

Waytowhich, N./Lawhern, VJ/Garcia, JO, et al., "Compact Convolutional neural networks for classification of asynchronous steady-state visual evoked potentials", Journal of Neural Engineering, Volume 15, Issue 6, 2018, passim.

West, Sarah Myers/Whittaker, Meredith/Crawford, Kate, Discriminating Systems. Gender, Race and Power in AI. New York: AI Now Institute, 2019.

Wexler, Rebecca, "Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System", Stanford Law Review, Volume 70, 2018, pp. 1343 and ff.

Williams, Rebecca, "Accountable Algorithms: Adopting the Public Law Toolbox Outside the Realm of Public Law", Current Legal Problems, Vol. 75, 2022, pp. 237–263.

Williams, Rebecca, "Rethinking Administrative Law for Algorithmic Decision-Making", Oxford Journal of Legal Studies, Volume 42, 2021, pp. 468 e ff and passim.

Willis, Lauren, "Deception by Design", Harvard Journal Law and Technology, Volume 34, 2020, pp. 115-190.

Wisser, Leah, "Pandora´s Algorithmic Black Box: The Challenges of Using Algorithmic Risk Assessments in Sentencing", American Criminal Law Review, Volume 56, 2019, pp. 1811-1832.

Workman, W., "Advancements in technology: New opportunities to investigate factors contributing to differential technology and information use", International Journal of Management and Decision Making, Volume 39, 2007, pp. 317 ff.

Wu, Tim, "Will Artificial Intelligence Eat the Law? The Rise of Hybrid Social-Ordering Systems", Columbia Law Review, Volume 119, 2019, pp. 2001 and ff and passim.

Volokh, Eugene, "Chief Justice Robots", Duke Law Journal, Volume 68, 2019, pp. 1135 and ff and passim.

Xiang, Alice, "Reconciling Legal and Technical Approaches to Algorithmic Bias", Tennessee Law Review, Volume 88, 2021, passim.

Yang, C./Hang, X./Wang, Y., et al, "A dynamic window recognition algorithm for SSVEP-based brain-computer interfaces using spatial temporal equalizer", International Journal of Neural Systems, Volume 28, Issue 10, 2018, passim.

Yang/Kai Hao, "Selling Consumer Data for Profit: Optimal Market-Segmentation Design and its Consequences", American Economic Review, Volume 112, Issue 4, April 2022, 2022, pp. 1364-1393.

Yu, Peter K, "Can Algorithms Promote Fair Use?", Fiu Law Review, Volume 14, 2020, pp. 328 and ff and passim.

Yu, Peter K., "The Algorithmic Divide and Equality in the Age of Artificial Intelligence", Florida Law Review, Volume 72, 2020, pp. 355-360.

Zhang, X./Sheng, QZ et al., "Converting your thoughts to texts: enabling brain typing via deep feature learning of eeg signals", IEEE International Conference on Pervasive Computing and Communications. Athens: IEEE, 2018, passim.

Zhu, Mirilla, "Jury, Using Artificial Intelligence to Predict Recidivism Rates", Yale Scientific, 2020, passim.

Kozlov, Yuri/Shutova, Maria/Bajwa, Taaha, Automated Judge is Not a Task For LegalTech But For DeepTech, 24th February of 2025, 2025, p. 1.

Zou, Mimi/Leffley, Ellen, "Generative Artificial Intelligence and Article 6 of the European Convention on Human Rights: The Right to a Human Judge?", Mimi Zou, Martin Ebers, Cristina Poncibò and Ryan Calo (Editors), The Cambridge Handbook of Generative AI and the Law, Cambridge, Cambridge University Press 2025, passim.

Zuboff, Zhoshana, The Age of Surveillance Capitalism, The Fight for a Human Future at the New Frontier of Power. New Yor City: PublicAffairs Books, 2019.

Authorship information

Hugo Luz dos Santos. PhD in Law and University Professor at City University of Macau, China, Fellow of the Royal Society of Arts of the United Kingdom "in recognition of his outstanding contributions to the field of justice, rule of law and policy worldwide". hugo.miguel.luz@gmail.com

Additional information and author's declarations (*scientific integrity*)

Conflict of interest declaration: the author confirms that there are no conflicts of interest in conducting this research and writing this article.

Declaration of authorship: all and only researchers who comply with the authorship requirements of this article are listed as authors; all coauthors are fully responsible for this work in its entirety.

Declaration of originality: the author assures that the text here published has not been previously published in any other resource and that future republication will only take place with the express indication of the reference of this original publication; he also attests that there is no third party plagiarism or self-plagiarism.

Data Availability Statement: In compliance with open science policies, all data generated or analyzed during this study are included in this published article.

Editorial process dates (https://revista.ibraspp.com.br/RBDPP/about)

- Submission: 05/08/2025
- Desk review and plagiarism check: 10/09/2025
- Review 1: 14/11/2025
- Review 2: 15/11/2025
- Preliminary editorial decision: 02/12/2025
- Correction round return: 02/12/2025
- Final editorial decision: 16/01/2026

Editorial team

- Editor-in-chief: 1 (VGV)
- Assistant-editor: 1 (DDE)
- Reviewers: 2

How to cite (ABNT Brazil):

SANTOS, Hugo Luz dos. Biased Algorithmic Risk Assessment in criminal justice settings: How is COMPAS fraying the fabric of the right to a human decision-making in criminal procedure law. *Revista Brasileira de Direito Processual Penal*, vol. 12, n. 1, e1287, jan./abr. 2026. https://doi.org/10.22197/rbdpp.v12i1.1287