

Machine-Learnt Bias? Algorithmic Decision Making and Access to Criminal Justice

Overall Winner, Justis International Law & Technology Writing Competition 2020, by Malwina Anna Wojcik of the University of Bologna

The pressure on the criminal justice system in England and Wales is mounting. Recent figures reveal that despite a rise in recorded crime, the number of defendants in court proceedings has been the lowest in 50 years.¹ This indicates a crisis of access to criminal justice. Predictive policing and risk assessment programmes based on algorithmic decision making (ADM) offer a prospect of increasing efficiency of law enforcement, eliminating delays and cutting the costs. These technologies are already used in the UK for crime-mapping² and facilitating decisions regarding prosecution of arrested individuals.³ In the US their deployment is much wider, covering also sentencing and parole applications.⁴

While the lack of undue delay is an important component of access to justice, so are equality and impartiality in decision making. Can we trust algorithms to be not only efficient, but also effective in combating discrimination in access to justice? One of the greatest promises of ADM is its presumed ability to eliminate subconscious bias that inevitably underlies all human decision making. An algorithm is believed to be capable of providing fairer and more accurate outcomes. Kleinberg's study showed that a machine learning algorithm trained on a dataset consisting of bail decisions from the New York City taken between 2008 and 2013 was able to outperform judges in crime prediction.⁵

Whilst a prospect of speedy and accurate decision making in standardised cases sounds very appealing to the underfunded and overloaded criminal justice system, there are numerous pieces of evidence suggesting that algorithms might not be as unbiased as we would expect them to be. This is because they can only be as fair as their creators and the data sets they are presented with. This problem is well illustrated by the research undertaken by Joy Buolamwini, a computer scientist who discovered racial bias embedded in the leading facial recognition technologies. Because of the

lack of racial and gender diversity in their data sets, these algorithms incorrectly recognised women with darker skin tones as men.⁶ The conclusion is that underrepresentation of minorities in the datasets results in lack of accuracy and generates a deeply biased output.

However, even the most inclusive datasets do not guarantee a fair result when the training data itself indicates bias or historical discrimination. Chiao argues that prioritising certain predictive factors in criminal risk assessment might lead to entrenching racial disparities within the algorithm.⁷ There is evidence that individual risk assessment programs, such as COMPAS, show racial bias and inaccuracy by incorrectly associating higher risk of reoffending with black people.⁸ It is so because seemingly neutral predictive factors, such as neighbourhood or previous arrest rate, in fact reveal racial bias based on unfair policing practices.

Similar concerns have been raised by a human rights watchdog, Liberty, which warns against the use of historical arrest data in crime-mapping, arguing that it does not reflect the actual crime rate and is likely to reinforce over-policing of marginalised communities.⁹ Reliance on such biased technologies could entail a violation of ECHR which requires a reasonable suspicion for an arrest to be lawful.¹⁰ Feeding the predictive algorithms with data containing historical bias will not lead to an increase in equality and fairness. On contrary, it will hinder access to justice by disadvantaging minorities and decreasing their trust in the criminal justice system.

On the other hand, it might be that the bias present in ADM will be easier to detect and combat than human bias. Algorithms can already be trained to be fairness-aware through incorporating anti-discriminatory constraints during data processing or removing the sources of bias prior to processing.¹¹ In practice however, there is a major disagreement on how the notion of fairness in a mathematical model should be construed. It is virtually

impossible for an algorithm to satisfy all definitions of fairness and maximise accuracy at the same time.¹² Nevertheless, many suggest that algorithms are generally capable of reaching a better trade-off between fairness and accuracy than humans.¹³ Thus, ADM could become a valuable tool for understanding the barriers to access to justice and removing them.¹⁴

But how can we ensure that algorithms live up to their full potential in combating inequalities? The best way to start is by increasing transparency and accountability of ADM technologies, which are often not properly tested or audited before their deployment in the justice system. The fact that private entities creating the technology can have a share in value-laden decisions concerning criminal justice is deeply problematic. It risks a fundamental erosion of access to justice by favouring what Buolamwini calls the ‘coded gaze’, a reflection of preferences and prejudices of those who have the power to shape technology.¹⁵ Public oversight over algorithms is crucial for ensuring non-discrimination and compliance with human rights. Worryingly, the Law Society has expressed serious concerns regarding the lack of openness and transparency about the use of ADM in the criminal

justice system. Some of its recommendations include development of the Statutory Procurement Code for ADM systems in criminal justice,¹⁶ creation of the National Register of Algorithmic Systems¹⁷ and extension of the public sector equality duty to algorithms’ design process.¹⁸ Increasing transparency and intelligibility is crucial for protecting the due process guarantees, especially the right to be given reasons for a decision and the right to challenge it. Therefore, it is equally important to introduce individual explanation facilities aimed at helping individuals understand how each decision was reached and how can it be appealed.¹⁹

A troubling conclusion is that algorithms seem to expose a fundamental lack of substantial equality in the criminal justice system. Whether ADM will be a force for good in removing this crucial obstacle for access to justice, depends on the presence of public scrutiny and auditing mechanisms. Carefully monitored algorithms can make a major improvement by detecting social inequalities and providing a better balance between accuracy and fairness. On the other hand, opaque systems based on “pale male” data sets and entrenched bias are likely to reinforce inequality and hinder access to justice.

Footnotes

- ¹ ‘Numbers dealt with in justice system at 50-year low’ (*The Times*, 15 November 2019), <<https://www.thetimes.co.uk/article/numbers-dealt-with-in-justice-system-at-50-year-low-k5hnb50kd>> accessed 15 November 2019.
- ² For example: PredPol used by Kent Police between 2013 and 2018 or MapInfo used by West Midlands Police.
- ³ For example: Harm Assessment Risk Tool (HART). See: The Law Society Commission on the Use of Algorithms in the Justice System, *Algorithms in the Criminal Justice System* (June 2019) para 7.3.1.
- ⁴ For example: Correctional Offender Management Profiling for Alternative Sanctions (COMPAS). See: ‘Practitioner’s Guide to COMPAS Core’ (Equivant 2019) <<http://www.equivant.com/wp-content/uploads/Practitioners-Guide-to-COMPAS-Core-040419.pdf>> accessed 15 November 2019.
- ⁵ J Kleinberg et al, ‘Human Decisions and Machine Predictions. (2018) *Quarterly Journal of Economics* 133, 237.
- ⁶ J Buolamwini, ‘Compassion through Computation: Fighting Algorithmic Bias’ (speech at the 2019 World Economic Forum in Davos) <https://www.youtube.com/watch?v=_sgji-Bladk> accessed 15 November 2019.
- ⁷ V Chiao, ‘Fairness, Accountability and Transparency: Notes on Algorithmic Decision-making in Criminal Justice’ (2019) 15 *International Journal of Law in Context* 126, 127.
- ⁸ J Angwin et al, 2016, ‘Machine Bias’ (*ProPublica*, 23 May 2016) <<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>> accessed 15 November 2019.
- ⁹ H Couchman, ‘Policing by Machine. Predictive Policing and the Threat to our Rights’ (*Liberty*, January 2019) 15.
- ¹⁰ Art. 5(1)(c) ECHR.
- ¹¹ The European Parliament Research Service, *Understanding algorithmic decision-making: Opportunities and challenges* (March 2019) 46.
- ¹² *ibid* 55.
- ¹³ Chiao (n 7) 129.
- ¹⁴ The Law Society Commission (n 3) para 8.4.
- ¹⁵ R Feloni, ‘An MIT Researcher who Analyzed Facial Recognition Software Found Eliminating Bias in AI is a Matter of Priorities’ (*Business Insider*, 23 January 2019) <<https://www.businessinsider.sg/biases-ethics-facial-recognition-ai-mit-joy-buolamwini-2019-1/>> accessed 15 November 2019.
- ¹⁶ The Law Society Commission (n 3) para 8.2, sub-recommendation 4.3.
- ¹⁷ *ibid* para 8.3, sub-recommendation 1.7.
- ¹⁸ *ibid* para 8.4, sub-recommendation 3.1.
- ¹⁹ *ibid* para 8.3, sub-recommendation 4.4.