Ethical Challenges in Predictive Policing:

Scrutinising the COMPAS Algorithm

Critical Studies Computational Ethics

Written by Nicole Stott

20-01-2025

Img: *https://petapixel.com/2017/10/06/two-photos-us-supreme-court-session/*

Table of Contents:

Ethical Challenges in Predictive Policing: Scrutinising the COMPAS Algorithm

As predictive algorithms increase their influence upon criminal justice systems, tools like the COMPAS algorithm participate directly in decisions related to bail, sentencing, and parole. The focus of COMPAS is the prediction of the probability of the defendant reoffending, and it provides data-driven insights for judges or other authorities. While supporters argue that these models enhance efficiency and objectivity in a long-time biased system, critics talk of innate biases, lack of transparency, and ethical issues surrounding autonomy and fairness. Given that predictive policing algorithms are trained using historical data which often reflects deep-rooted systemic inequalities, algorithms like COMPAS risk perpetuating these very inequalities they aim to diminish. This paper critically considers the COMPAS algorithm through the lenses of beneficence, justice, and respect, with useful improvements that could help in enhancing ethical integrity and reducing systemic biases in predictive policing: decision hygiene, feedback mechanisms, transparency, and data governance.

Historical Context and the Evolution of Predictive Policing

Predictive policing has its roots in early 20th-century criminology, where statistical models tried to identify risk factors for criminal behaviour. Statistical analysis in parole decisions began to grow in the 1970s, developing into modern tools such as COMPAS. The development of these tools was driven by the increase in availability of data and technological advancement, but their application remains controversial because past predictive models have demonstrated racial and socioeconomic biases, leading to disproportionate incarceration rates for minority groups (O'Neil, 2016).

Predictive policing has deep historical ties to racially biased law enforcement practices; the broken windows policing theory, which gained spotlight in the 1980s, emphasised aggressive policing of minor infractions in marginalised communities under the assumption that disorder ended up in serious crime. Many critics argue that modern predictive models, including COMPAS, replicate this approach by reinforcing pre-existing patterns of racial and economic disparities in criminal justice outcomes (Alexander, 2010). Furthermore, there is historical precedent for risk-assessment models disproportionately targeting marginalised communities due to the structural inequalities ingrained in society. These biases come not from the algorithm itself but from the data used to train such models, reflecting decades of systemic inequality.

Countries like the Netherlands and Canada have moved in parallel to more open and human-centred approaches to risk assessment in criminal justice, placing social work and mental health evaluations alongside predictive algorithms to make sure holistic decisions are made, their models indicate the possibility of more ethical frameworks in predictive policing. Besides, scholars like Michel Foucault have critically analysed the use of surveillance and data-driven policing, stating that predictive policing nurtures a long history of state power exercised through disciplinary mechanisms. From a Foucauldian perspective, COMPAS may evolve into an instrument of governance based on risk assessment, rather than individual justice, and could therefore be seen to erode personal freedoms and individuality.

Beneficence: *Maximising Social Good*

Beneficence as a core ethical principle requires that technologies be designed to maximise benefits for society while minimising harm. The potential of COMPAS lies in its ability to provide data-driven, objective insights into a traditionally subjective judicial system. Quantitative measures of recidivism risks assist judges in making decisions about how to allocate resources and prioritise cases. For example, high-risk offenders may be supervised more heavily, while low-risk offenders may be spared from unnecessary imprisonment. Ideally, this would ultimately reduce prison overpopulation, which has been a problem plaguing the U.S. justice system.

The effectiveness of COMPAS, however, rests on its use and development. Decision hygiene - disciplined and unbiased processes for making decisions - means algorithms remain a tool, not a determinant. If judges used COMPAS only as an added resource, then some of the risks of over-reliance and systemic harm may be mitigated. Furthermore, structured feedback from the judges could refine the algorithm's output over time, enabling it to evolve with society.

In this aspect, alternative predictive models may be considered to enhance the principle of beneficence within criminal justice systems. To illustrate this point, Canada has been conducting integrated risk assessment frameworks incorporating mental health screenings, rehabilitation programs, and direct support to the community to diminish the biases occurring in risk assessment. Predictive models that integrate human and social factors will have better standing in ethics, rather than a purely statistical prediction approach.

Respect: *Safeguarding Individual Rights and Autonomy*

The ethical use of predictive tools, such as COMPAS, would call for respect for individual rights and dignity. On the other hand, exclusive ownership of the COMPAS

algorithm is highly problematic to transparency and accountability. Defendants cannot often have access to the data underlying their risk scores, which further diminishes their autonomy and self-advocacy in court. This has brought about a lack of transparency that has even been challenged legally, with critics arguing it undermines the fundamental right to a fair trial.

Transparency is the most crucial element in ethical algorithm design. For instance, unless there is clarity on how the risk scores have been calculated, defendants cannot dispute flawed or biased assessments. Introducing decision hygiene, such as asking judges to record how they are using the COMPAS scores in making decisions, may help maintain accountability and build trust. Additionally, allowing access to the defendants regarding their respective risk assessments and data would give them a voice against inaccuracies. Recent debates over the legal questions have pressed for a standard of explainability in the AI models, especially when being used to make/support such vital decisions. Some scholars put forward arguments that the European Union's General Data Protection Regulation should be extended in its scope towards all predictive policing models across the world in order to effect the "right to explanation" from AI-driven decisions. It would therefore strengthen defendants' rights by allowing them to contest the decisions made by algorithms.

Legal Precedents and AI Regulation in Predictive Policing

The use of AI-driven predictive tools like COMPAS raise serious legal concerns regarding due process, accountability, and individual rights. In the European Union, the General Data Protection Regulation has established the right to explanation, requiring that individuals subject to automated decision-making process have the right to understand how those decisions were made. If applied to predictive policing, this regulation would require COMPAS and similar algorithms to provide explicit reasoning for their assessments, preventing unclear decision-making from determining legal outcomes.

In the U.S., legal challenges to AI-based sentencing have emerged under the Due Process Clause of the Fourteenth Amendment. The Eric Loomis case in Wisconsin highlighted the lack of transparency in COMPAS's risk assessment. Loomis argued that inability to contest the methodology behind the algorithm denied him his right to a fair trial. The court maintained the use of COMPAS, but the case underlined the growing need for more transparency and algorithmic accountability in criminal justice. Other legal experts have argued that predictive policing technologies could fall under the ambit of the Equal Protection Clause, which protects against AI-driven sentencing that would disproportionately affect minority communities. Courts may have to create

judicial standards regarding the admissibility of AI evidence, like the Daubert standard that dictates the admissibility of scientific evidence during trials.

Global Perspectives: How Different Nations Approach AI in Criminal Justice

Predictive policing throws up ethical questions of varying proportions across different legal and cultural regimes. For instance, Norway has mostly rejected the risk-based sentencing model and relies more on rehabilitation and reintegration programs as methods of preventing recidivism. While China has enthusiastically adopted AI-driven surveillance and predictive analytics in law enforcement, there are many worries about state overreach and violations of civil liberties. Germany has reacted with a severe backlash against AI by placing strict regulations on the use of predictive algorithms in sentencing to ensure transparency and due process. These approaches strongly indicate that predictive policing must be meticulously tailored to meet societal values. If AI tools are to be implemented in an ethical manner, they must be tailor-made to regional legal traditions and human rights standards, rather than one-size-fits-all.

Justice: Ensuring Fairness and Equity

Justice essentially calls for fairness in the design, implementation, and outcomes of algorithms such as COMPAS. Critics argue that COMPAS perpetuates racial disparities in assigning higher risk scores to Black defendants. This partially emanates from the historical data the algorithm was trained on, reflecting systemic over-policing in marginalised communities. These concerns are, in part, about the limits of correlation versus causation: socioeconomic factors such as poverty are highly correlated with recidivism, but these correlations are indicative of general social injustices rather than overt algorithmic bias. The mitigation of such biases involves setting data in a wider context of social justice. Reweighting training data to reduce the impact of historical biases may help achieve more fairness, but a more efficient – at least in the long term- solution to enhance justice would involve algorithmic fairness audits, where predictive models are independently tested and refined on a regular basis to ensure they do not perpetuate systemic discrimination. Regular audits will force developers of AI to be neutral in the outcome of risk assessment.

The Accuracy and Bias of COMPAS: Examining Quantitative Evidence

Quantitative studies on COMPAS show significant racial disparities in the risk assessment it calculates. A 2016 investigation by ProPublica reported that COMPAS incorrectly classified Black defendants as high-risk at nearly twice the rate of White defendants, while also misclassifying White defendants as low risk more frequently than Black defendants (Angwin et al., 2016). The results are indicative of a systemic bias from the training data, representative of historically biased law enforcement practices. Further, Dressel and Farid (2018) compared the predictions of COMPAS to humans who had never received any training in criminal risk assessment. Their study found that the humans and COMPAS performed just about equally well, which, considering its complexity, does not really translate into greater predictive accuracy. And this begs the question as to whether the utility of predictive policing tools brings something over and above human judgment or merely automates the same with pre-existing biases. In turn, some developers of AI have suggested re-weighting data sets or incorporating fairness constraints in order to limit algorithmic bias. Critics argue that data cannot overcome systemic inequities, and only broad reforms to criminal justice will treat disparities at their source rather than relying on tweaks to the algorithm.

Ethical Counterarguments: The Case for Predictive Policing

While many critiques have focused on ethical flaws in COMPAS, supporters of predictive policing believe such tools can provide real value if implemented with proper safeguards. Law enforcement agencies say risk assessment algorithms ensure resources are used efficiently, with high-risk individuals getting necessary intervention and low-risk individuals spared from unwarranted incarceration. Other legal experts suggest AI, if used properly and given the correct data, could minimise human biases rather than perpetuate them. It has been proved that human judges are prone to cognitive biases, such as the racial empathy gap and disparity in sentencing based on characteristics of defendants. If calibrated properly, AI can offer a uniform data-driven method of sentencing and reduce personal prejudices that all too often drive judicial decisions. Another argument that can be made for predictive policing is crime prevention. The ability to identify high-risk individuals allows intervention programs to take place before crimes are committed, therefore steering criminal justice toward rehabilitation instead of punishment. Programs that combine AI risk assessments with social services have seen some promising results in recidivism reduction, especially when combined with vocational training and mental health initiatives. But if these advantages are to flourish, auditing, transparency measures, and legal oversight will have to be established. AI in policing should augment and not replace human judgment, with predictive tools advisory but not deterministic.

Extending the Moral Landscape: Virtue Ethics and Predictive Policing

Much of the ethical discussion surrounding the use of COMPAS has been framed in terms of the principles of beneficence, justice, and respect; however, another salient perspective is that of virtue ethics, which places moral character above rule-based decision-making. The virtue ethics approach to predictive policing, question whether reliance on algorithms promotes moral virtues such as fairness, wisdom and compassion in the judicial system. If predictive tools motivate judicial officers to depend on statistical probabilities rather than individualised assessments, they erode the moral duty to treat defendants as unique persons, rather than mere data points. Ethics training for legal professionals is one possible solution, where AI remains supplementary to, not a replacement for, human moral judgment in sentencing decisions.

Conclusion

The COMPAS algorithm epitomises in equal measures the promise and pitfalls of predictive policing: It offers efficiency and objectivity, yet also runs the risk of fermenting systemic biases and eroding individual rights. Ethical computing principles - beneficence, respect, and justice - need to support all future development and deployment of predictive policing tools if they are to serve society without further fermenting inequalities, this said, we are wrong to call it an algorithmic/machine bias as the bias does not come from the algorithm nor from the machine.

In this aspect, a judicial system should incorporate mechanisms for decision hygiene, transparency, and continuous feedback loops when using COMPAS. Moreover, policy interventions and independent oversight bodies are needed to audit predictive algorithms for fairness and accountability.

As AI continues to evolve, ethical considerations in its role in criminal justice will be increasingly important. The future of predictive policing lies in balancing technological advancements with the protection of human rights in such a way as to enhance, not compromise, justice, dignity, and fairness in legal decision-making.

References –

Alexander, M. (2010) *The new Jim Crow: Mass incarceration in the age of colour-blindness*. New York: The New Press.

Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (2016) 'Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks', *ProPublica*, 23 May. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (Accessed: 3 December 2024).

Dressel, J. and Farid, H. (2018) 'The accuracy, fairness, and limits of predicting recidivism', *Science Advances*, 4(1), pp. 1-5. Available at: https://www.science.org/doi/10.1126/sciadv.aao5580 (Accessed: 07 December 2024).

European Parliament and Council (2016) 'General Data Protection Regulation (GDPR)', *Official Journal of the European Union*, L119, pp. 1-88. Available at: https://gdpr-info.eu (Accessed: 19 December 2024).

Foucault, M. (1977) *Discipline and punish: The birth of the prison*. Translated by A. Sheridan. London: Penguin.

Loomis v. Wisconsin, 137 S. Ct. 2290 (2017).

O'Neil, C. (2016) *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Crown Publishing.

Wisconsin Supreme Court (2016) *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016)