THE ALGORITHM IN THE DOCK: WHY INTERNATIONAL LAW MUST GOVERN AI-DRIVEN DECISIONS IN CRIMINAL JUSTICE

Zeeshan Akhtar Khan, BA LLB, Aligarh Muslim University Centre, Malappuram

ABSTRACT

The increasing use of artificial intelligence (AI) in criminal justice systems raises serious concerns about fairness, transparency and human rights. AI tools are being used more and more in sentencing, bail, parole, predictive policing, and facial recognition. But a handful of studies and incidents, including the wrongful arrest of Robert Williams, have shown that these systems can produce discriminatory and inaccurate results, particularly for racial minorities. Current international human rights law (including ICCPR) does not sufficiently regulate algorithmic decision-making or provide protections from AI-enabled injustice. The EU Artificial Intelligence Act provides a relevant regulatory framework, but its protection is regional and limited. This paper argues that international law must evolve to regulate AI in criminal justice through stronger transparency standards, human oversight, accountability mechanisms, and global legal safeguards to protect fair trial rights, equality, and individual liberty.

AI in Criminal Justice: A Dangerous Beginning

On the January of 2020, Robert Williams was detained at his home in Michigan in front of his wife and small kids. There are no witnesses who could identify him. There were no physical evidence linking him to the crime. The facial recognition algorithm output that the Detroit Police Department acted on had turned out to be wrong. Williams spent thirty hours in custody before investigators admitted the mistake. To the public knowledge, he is considered to be the first American arrested by reliance on a match generated by AI.

His case is not a deviation it is indicative of something. Criminal justice systems worldwide are using algorithmic tools to make or significantly shape decisions that affect whether a person goes to prison, remains in detention, or walks free, including sentencing, bail, parole and surveillance decisions. But the international laws on criminal procedure, fair trial rights and the prohibition of discrimination say almost nothing about the machines that are now embedded in those very systems. That silence is becoming nasty.

What the Algorithms Actually Do

The use of AI in criminal justice happen in various ways. Software that predicts reoffending is the most researched category of these programs. It assigns defendants a number, that indicates their risk, based on factors such as their previous offending, work history and home postcode. Judges and parole boards take these scores into consideration when deciding upon sentencing and release. One of the most widely used tools in the US is COMPAS (Correctional Offender Management Profiling for Alternative Sanctions). A ProPublica investigation found that COMPAS was almost twice as likely to wrongly classify Black defendants as high risk as they were for white defendants. The tool also low-balled the risk posed by white defendants. The organisation responsible for the software, Equivant, challenged the methodology but the main finding has been repeated and argued over in the academic literature ever since.

Facial recognition presents another serious problem but different, According to NIST, a majority of user facial recognition systems misidentified us in one experiment, and a lot worse: misidentified black and Asian faces at between a factor of 10 and 100 more than white faces. Despite having a documented inaccuracy rate of 94 per cent, police forces in the UK, the USA, India, and beyond continue to use the technology in investigatory contexts without any legislation about its use, accuracy thresholds, or availability of personal remedies.

In predictive policing, a third category which utilizes economic and crime data to predict the location of crime or the person likely to commit the crime. According to critics, including the ACLU, these systems embed and amplify the biases in historic policing data. Thus, they put communities under greater scrutiny for offences they never committed.

The International Law Gap

The primary issue from an international law point of view is that of legal silence and institutional fragility. They require a competent, independent, and impartial tribunal in ICCPR which guarantees the right to a fair trial violate one's right to a fair trial article 14. Article 26 guarantees everyone's right to not be discriminated against. Article 14(2) guarantees the presumption of innocence. Human Rights Committee's current interpretation of all these provisions is in no systematic way applied in relation to algorithmic decision-making tools.

The Human Rights Committee's General Comment No 32 on Article 14, which is the most authoritative interpretive guidance on fair trial rights, was adopted in 2007, before the current generation of AI tools existed in a deployable form. The proposal does not refer to automated decision-making, the right to inspect or contest algorithmic outputs, or what is meant by "impartial tribunal" when one input to the judicial decision is a black-box proprietary score. It is not only that there is a lack of academic commentary; it signifies that international human rights law has structurally failed to keep up with the technology it purports to govern.

The Council of Europe is the most advanced in this area the case law developed by the European Convention on Human Rights notably in Articles 5 (liberty), 6 (fair trial), and 8 (private life) has begun to deal with automated decision-making. In Loomis v Wisconsin (a Wisconsin Supreme Court decision rather than one out of Strasbourg), the court accepted COMPAS for sentencing but acknowledged its limitations, a decision that drew significant academic fire on due process grounds. The SyRI case at the European level of the Hague District Court in 2020 saw the court striking down a welfare-fraud prediction state system of the Dutch government under Article 8 ECHR. This establishes that algorithmic state systems must meet proportionality, transparency, and contestability requirements. Yet, these only apply for the nation and region and not globally.

The EU AI Act: A Partial Template

The most developed regulatory framework currently in existence, EU Artificial Intelligence

Act 2024. The legislation categorizes AI systems employed in criminal justice as "high-risk," like risk assessments and recidivism forecasts. It also largely bans the use of real-time biometric identification systems in public places by police and detection authorities. Mandatory conformity evaluation, transparency requirements, human oversight, and database registration at EU level for high-level risk systems. Affected individuals gain an explicit right to explanation under Article 86.

The Act is a remarkable milestone. However, it is by design a regional instrument the EU law would only apply to AI whose outputs are used in the EU. The risk assessment tool known as COMPAS is used in the United States. The National Automated Facial Recognition System (NAFRS) of India lies completely outside the jurisdictional reach of the EU. Through algorithmic systems, millions of defendants are processed in jurisdictions without equivalent domestic regulation and are completely unprotected.

What International Law Must Do

It is necessary to have three international interventions. To start with, the Human Rights Committee should adopt a new General Comment – or a dedicated General Comment – applying Articles 14 and 26 ICCPR to algorithmic decision-making in the criminal process. This should at a minimum prove that the right to a fair trial means: (i) defendants must be informed that an algorithmic tool has been used in their case; (ii) The defendant must have access to sufficient information about the tool's methodology to enable a meaningful challenge; and (iii) the right to human review of any algorithmic output that adversely impacts on liberty.

Second, the UN Special Rapporteur on the right to privacy and the Special Rapporteur on racism should produce a joint thematic report on the discriminatory impact of AI in criminal justice, building on the *existing work of the Special Rapporteur on contemporary forms of racism* to generate the evidentiary record that treaty body updating requires.

Third, the international community should negotiate a Protocol to the ICCPR — or a standalone treaty establishing binding minimum standards for the use of AI in criminal justice. This instrument should draw on the EU AI Act's risk-based framework but extend it universally, setting floors rather than ceilings: minimum transparency requirements, mandatory human oversight of liberty-affecting decisions, prohibition of AI systems with documented discriminatory impact, and individual remedies for AI-caused harm.

Conclusion

Robert Williams was eventually released and the charges against him were dropped. But he should not have been arrested in the first place. The algorithm that put him in handcuffs operated in a legal vacuum. No statute governed its use, no international standard constrained its deployment, and no right entitled him to challenge its output. That vacuum is not a technical accident. It is a failure of law to keep pace with power.

International human rights law was built precisely to protect individuals from the unchecked exercise of state power. As states increasingly exercise that power through algorithms, international law must adapt. The question is not whether the tools are new — every generation of rights lawyers has confronted novel instruments of state authority. The question is whether the international legal order will respond with the urgency that the scale and speed of algorithmic harm demand.