Machine Learning Algorithms in the field of Criminal Justice

Poonam Kashyap
M21CS012

Abstract—**Artificial intelligence (AI) and automated decision-making (ADM) systems are increasingly used by European law enforcement and criminal justice authorities to profile people, predict their supposed future behaviour, and assess their alleged 'risk' of criminality or re-offending in the future.**

I. PROBLEM STATEMENT

European law enforcement and criminal justice agencies are increasingly using artificial intelligence (AI) and automated decision making (ADM) systems to profile persons, anticipate their alleged future behaviour, and estimate their alleged 'risk' of criminality or re-offending in the future. Constant monitoring, stop and search, fines, interrogation, arrest, custody, prosecution, sentence, and probation are just some of the policing and criminal justice outcomes that these forecasts, profiles, and risk assessments can affect, inform, or result in. They can also result in non-criminal judicial sanctions such as denial of welfare or other vital services, as well as the removal of children from their homes. These AI and ADM systems reproduce and reinforce discrimination based on race, socioeconomic status, and nationality, as well as engage in and infringe on fundamental rights such as the right to a fair trial and the presumption of innocence, the right to privacy and family life, and data privacy rights. The law enforcement and criminal justice data used to create, train, and operate AI and ADM systems reflects systemic, institutional, and societal biases that lead to over policing and disproportionate detention and imprisonment of Black people, Roma, and other minorities ethnic people across Europe. These biases are so deep and engrained that it's debatable if any system would not give similar results. Predictive analytics, risk assessment, and profiling Individuals are targeted and profiled as criminals by AI and ADM systems before they have carried out the alleged conduct for which they are being profiled, resulting in substantial criminal justice and civil outcomes and sanctions. In essence, these arrangements are designed to stifle the fundamental right to be assumed innocent. These systems are also frequently plagued by technological impediments that obstruct effective and meaningful oversight, transparency, and accountability. Concerns have also been raised about the absence of meaningful human input into these automated judgments, as well as their direct application and impact on children and adolescents.

II. CASE STUDY

Top 600 – Amsterdam Municipality, police social services (Netherlands) Used for: Risk modelling and profiling Created: 2012

In 2012, the Amsterdam Municipality started the 'Top600', an automated risk modelling and profiling system, in partnership with police and social services which attempts to profile the 'top 600' young people, over the age of 16, who are most at risk of committing 'High Impact Crime' in the future. Among those people profiled by the Top600, there are clear themes, both in terms of their ethnicity and their neighbourhoods, which suggests a clear element of discrimination on ethnic and socioeconomic grounds.

One mother of a child on the Top600 described that if people refuse to participate in the "care" side of the Top600 approach, "they will report you" and "threaten to remove your children from home".

III. POSSIBLE SOLUTIONS

Prohibition on predictive, profiling and risk assessment AI and ADM systems. Strict legal safeguards are needed against discrimination.

Implement required, independent bias testing during the design of algorithms and pre-deployment phases, as well as on an ongoing basis after deployment.

Data on criminal justice must be available in order to conduct this testing, and such data must be separated by race, ethnicity, and nationality. Most of the time the availability of official data becomes a barrier to test the model and therefore it becomes hard to determine the factors which cause bias and give partial output. Hence making data available to all will help to understand the machine learning tools better and make it more appropriate.

Before being utilised or even 'trialled' in real-world circumstances where they have genuine consequences on individuals or criminal justice results, AI or ADM systems should be thoroughly vetted. This procedure must also include affected organisations and people. If these tests aren't done, an AI or ADM system can't be demonstrated to be non-discriminatory, it should be prohibited from being used.

The lack of racial and ethnic data may also make it difficult to take preventative efforts to reduce racial bias. It's unlikely that engineers will be able to create systems that are devoid of racial prejudice if they don't have any data to compare

their work to. Data for monitoring and evaluation purposes will also need to have been collected well before the AI and ADM systems were implemented, in order to make a proper pre- and post-analysis comparison.

Ensuring transparency and accountability

In criminal justice, AI and ADM must be transparent, with system processes open source and not subject to legal safeguards like trade secrecy or intellectual property laws.

Their outputs must be understandable and scrutinised by their supervisors, decision-makers (such as suspects and accused persons), and the general public.

Individuals must be notified whenever an AI or ADM system has or may have influenced a criminal justice decision, whether it was helpful or not.

Through fully reasoned, case-specific, documented choices, human decision-makers must demonstrate how and in what way decisions were impacted, including what factors influenced a decision and if this involved AI or ADM system outputs.

Individuals wishing to protest or criticise AI and ADM judgments, or the systems themselves, must also have clear paths to challenge or redress.

REFERENCES

[1] Council of Europe, 'Unboxing Artificial Intelligence: 10 steps to protect Human Rights', May 2019. https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e6

[2] Automating Injustice: The use of Artificial Intelligence Automated Decision Making systems in criminal justice in Europe.