The use of AI in sentencing and the management of offenders

A workshop on 27 February 2023 jointly hosted by The Alan Turing Institute, Northumbria University's Centre for Evidence and Criminal Justice Studies, and the Sentencing Academy explored the role that artificial intelligence plays—and could play—in the sentencing and management of offenders.

Executive Summary

- The implementation of AI into sentencing is complex due to non-universal definitions, imperfect or non-uniform data, an overload of information of varying value, and the risk of mistakes reducing the legitimacy of sentencing on which the judicial system relies.
- AI can provide some benefits, such as mitigating against possible judicial bias or improving the speed of access to justice in the case of nations with sentencing guidelines. However, these benefits are often offset by sacrificing the right to speak in one's own defence at sentencing hearings, and risks such as unexplainable decision-making with inadequate reasoning given.
- AI may overlook the fact that the exact same offence committed against two different victims will have different impacts, and this is something that currently requires a human's lived experience to analyse the harm caused by an offence.
- AI is built, trained, and validated on existing data which can risk creating a false sense of objectivity, and subsequently over-trust in outputs.
- AI may have a place in a research or advisory capacity when it comes to sentencing and help to detect and/or reduce bias or discrimination in sentencing.
- AI may be particularly useful to the judiciary in the tracking and analysis of their sentencing habits, helping them to self-assess their own potential hidden biases and correct them with training.
- AI has a role to play within sentencing, but that role must be chosen carefully and only where it can make a measurable and safe impact. It should not be used for its own sake, or for the appearance of modernisation.

Presentation Summaries

The workshop was split into two panels: *Technical Challenges for the Use of AI in Sentencing* and *Exploration of the Issues for Practice arising from the Use of AI in Sentencing and the Management of Offenders*.

Section One: Technical Challenges for the Use of AI in Sentencing

Dr Elizabeth Tiarks - Northumbria University

Core Message:

Implementing AI into sentencing is particularly complex due to the absence of a single understanding of definitions; imperfect, variable amounts of information available; an overabundance of information of varying value; risks of incorrect decisions; and the importance of high accuracy to ensure legitimacy. Sentencing also has significant impacts on the family of those sentenced and the wider community, and therefore is vital to get right.

Content:

The issues to be solved by AI in sentencing include concerns about bias and the arbitrary nature of sentencing, and the ability of AI to reduce or mitigate against bias. The focus very much remains on consistency and proportionality.

There are important questions around which factors make cases different from one another and how AI interprets these, as well as what difference contextual factors should make. There are aggravating or mitigating factors to various extents, with contextual interplay between factors.

Dr Tiarks discussed the five statutory purposes of sentencing: punishment, reduction of crime, reform of offenders, the protection of the public, and reparation. There is no singular purpose, and the aims are somewhat contradictory. AI may well struggle with such vague and conflicting goals.

There is no single understanding of consistency to use as a benchmark, while data are too variable and contextual to reliably count on. Many sentences are highly impacted by barristers and magistrates. Mitigating factors are also taken into account in some cases and not others, depending on the level of representation.

Dr Miri Zilka – University of Cambridge

Core Message:

AI is built on existing data which can risk creating a false sense of objectivity. This is due to technical factors in how artificial intelligence and machine learning technologies work.

Content:

Dr Zilka discussed how criminal justice tools based on AI can often be non-objective due to a range of technical factors. This excessive trust in tools can cause problems when deployed in a criminal justice setting.

Section Two: Exploration of the Issues for Practice arising from the Use of AI in Sentencing and the Management of Offenders

Professor Julian Roberts - Sentencing Academy and University of Oxford

Core Message:

AI should have a limited role at sentencing. Though some scholars have advocated replacing human judges by an algorithmically determined sentence, it should not be ignored that AI enhances certain rule of law requirements but undermines others.

Content:

AI could have a positive influence on impartial treatment, by constraining a potentially biased human decision-maker. A judge could compare his or her proposed sentence to one generated by an algorithm which has processed all the information available to a court. AI would be better able to devise a sentence which conforms to established sentencing principles such as proportionality, restraint and equity. At the same time, AI will be blind to extra-legal factors which may have influenced a human judge, for example the defendant's race, ethnicity, immigration status, and employment record.

Impartiality can also be improved: sentencing patterns or sentencing judgments can be scrutinised by AI to identify sources of bias or specific courts or judges which routinely depart from the established sentence range or tariff for a specific offence.

In jurisdictions which operate formal sentencing guidelines (such as England and Wales, South Korea and many US states), AI can identify elements of these guidelines which trigger inequitable outcomes. For example, AI is much better able (than human researchers) to identify indirect sources of discrimination at sentencing.

However, AI-determined sentencing can never replace the sentencing hearing. This is a key element of common law sentencing. The hearing gives both the victim and the offender an opportunity to be heard. The crime victim may wish to explain to both the court and the offender the impact of the crime. From the offender's perspective, defendants are likely to perceive the sentencing process as more legitimate if they have an opportunity to address the court before a sentence is imposed.

A *viva voce* hearing also offers legal counsel the opportunity to respond to each party's submission on sentence, an essential element of an adversarial proceeding. In short, AI can best contribute to sentencing in a research or advisory capacity, uncovering sources of disparity or discrimination. Sentencing, however, should remain in the hands of a human adjudicator.

Robin Allen KC - Cloisters Legal Practice

Core Message:

AI in sentencing may have some uses as a support tool for the judiciary but should not be used for the actual sentencing exercise itself. Though there are concerns around handing over control of sentencing to non-human machines without the cultural context and life experience of a human decision-maker, AI's use as a support tool could help to mitigate against elements such as bias which are prone to creeping into human decision-making.

Content:

The judiciary can sentence for a variety of reasons, with some seeing the influence of social outcomes an important part of the concept of justice. This contextual information on—and understanding of— society makes it very difficult for non-human decision-making to get things right.

However, AI may have a place within sentencing as a judicial support tool. It could help to detect and mitigate against judicial bias by informing the judiciary of their own sentencing habits. It could also provide a second opinion for a court by predicting what that particular judicial member has imposed as a sentence for similar offences in the past.

Alexander (Sacha) Babuta - The Alan Turing Institute

Core Message:

Sacha Babuta (Director of the Centre for Emerging Technology and Security) presented the findings from his doctoral research which comprised a process evaluation of a machine learning risk assessment project undertaken in a large UK police force. The project under evaluation was a pilot project to develop a data-driven risk scoring system for forecasting risk of future offending. This included using machine learning and other statistical modelling to assign a 'harm score' to each supervisee, as well as a predictive score forecasting each individual's risk of escalating to more serious offending.

The evaluation found that the project had not delivered the intended benefits that were envisaged at the time the system was developed. The beta-testing phase of the project was inherently limited due to the lack of pre-defined evaluation metrics by which the success (or otherwise) of the project would be assessed. Notably, the research highlighted a significant divergence in perspectives between senior officers overseeing the project on the one hand, and more junior officers expected to use the system operationally.

Content:

Senior officers were far more complimentary and positive regarding the system than operational users. The research also identified fundamental deficiencies in the user interface of the system and data integration challenges, meaning that regardless of the effectiveness of the statistical modelling used, the software has provided limited operational benefit for users. Finally, a lack of sufficient training was highlighted as another important reason for the pilot project not achieving its intended outcomes.

Sacha then summarised the findings from recent Centre for Emerging Technology and Security research on the use of human-machine teaming in intelligence analysis. The project engaged directly with operational intelligence analysts working within UK national security to understand how users interact with the output from machine learning systems, and how machine-generated insights can be effectively integrated into existing decision-making processes.

The research found that the way an analyst treats the output from a machine learning system is highly context-specific, and will depend on the urgency of the decision, the priority of the operation, and the perceived impact of subsequent decisions on resources and outcomes. In many cases, a technical explanation of the model's behaviour was not particularly helpful for an analyst to calibrate the appropriate level of trust in the system. Instead, analysts require

context-specific, user-specific and interactive technical information about the model that is presented in accessible, standardised language. The complexity of these explanations should be determined by the complexity of the problem on which the model is deployed. The project made specific recommendations for the effective integration of machine learning systems into intelligence analysis processes.

Summary

Incorporating AI into sentencing stands to bring some measurable benefits to the justice system in England and Wales, such as the reduction of bias. However, AI also poses some risks to sentencing on both a philosophical/theoretical and a practical level.

A live hearing is a fundamental part of the adversarial common law system of justice. AI-based sentencing does not include a hearing and the absence of this key procedural element may undermine perceptions of fairness and confidence in the sentencing process.

There was also concern that the varying and potentially conflicting purposes of sentencing, such as reform of offenders, reparation, and punishment, would be difficult for AI to handle contextually due to the contextual interplay between factors. This was a recurring theme: that the difficulty of implementing the automation of sentencing and the risks involved in doing so would outweigh the benefits.

However, there was recognition that AI could help to ensure that sentences better adhere to core sentencing principles uniformly, so that sentencing is carried out consistently across the whole of England and Wales.

In terms of the practical administration of sentencing, there was concern among some scholars that the implementation of AI within sentencing was particularly complex due to a multitude of administrative and technical factors, including non-universal agreement of definitions, imperfect and non-uniform data collection and storage, and the significant costs of errors on both individuals and in the continued legitimacy of the justice system.

However, there was a shared belief that AI could help with many of the more administrative aspects of sentencing, such as ensuring fairness by the judiciary, helping the judiciary to make decisions, and providing a clearer picture of the sentencing landscape to decision-makers in government. This was due to the fact that the judiciary need not necessarily rely on the AI, but merely factor its advice into their decision-making process.

To summarise, AI has a role to play at sentencing, but only where it can make a measurable and safe contribution. It should not be used for its own sake, or for the appearance of modernisation.