



This response is submitted by academic researchers with extensive experience of theory and practice of real-world ethical, legal, safeguarding and operational approaches to data analytics and AI in criminal justice. Our RAI Keystone Project: "Probable Futures", is a four-year research project reviewing Probabilistic AI systems across the criminal justice system to create a responsible and 'operational-ready' framework, working with multiple criminal justice partners.

Summary:

- 1. The impact of technological tools:** Technological tools cannot provide solutions to offender sentence and management challenges, if introduced into a situation where the technology cannot be deployed safely and responsibly. Indeed, technological tools may exacerbate or obfuscate existing challenges, making it harder to find effective solutions, while increasing risks to justice and community safety as well as the inherent risk to individual rehabilitation.
- 2. Clarity about the use of technological tools:** There is currently insufficient clarity about how tools may inform sentences, or which statutory purposes of sentencing are being achieved/prioritised with the deployment of technological tools. The use of tools, and the effectiveness of tools, requires further clarity.
- 3. AI in sentencing:** AI is not a cheap route to the quicker processing of cases. If justice is to be served, there are time-intensive measures, with resource implications, that would need to be implemented prior to the adoption of machine learning tools. There is a lack of high-quality data about the operation and impact of tools. Such tools also lack transparency, and the training models may 'bake in' judicial bias, rather than eliminate it. There may also be discrepancies in the values incentivising tool developers, when compared with the values of the criminal justice system.

Theme 3: Technology

Q: How can we use technology to be innovative in our sentencing options, including considering how we administer sentences and manage offenders in the community?

The impact of technological tools

It is vital that the practical impact of the deployment of technological tools is thoroughly interrogated. There have already been issues flagged with the use of predictive tools in probation, such as risk assessments. The risk scores produced by tools such as the Offender Group Reconviction Scale (OGRS) feature in Pre-Sentence Reports (PSRs) and are then relied upon by sentencers. The growing emphasis on speed of report delivery means that PSRs are now more commonly in a short format. In short format reports, risk scores can appear independently of potentially important context about the offender. Shorter format reports have been found to be of lower quality and less accurate, e.g. HM Inspectorate of Probation found that a focus on the timeliness of PSRs had reduced their quality (HMIP, 2020; HMIP, 2023a).

In a recent annual report (HMIP, 2023a) HM Inspectorate of Probation criticised the quality of risk assessments in PSRs, finding staff lacked sufficient time to properly review them due to staffing shortages, large caseloads and over-worked senior probation officers. These factors have also been identified as partially responsible for the perpetuation of race inequality in PSRs (HMIP, 2023b). In addition, there is evidence that reports do not contain adequate (or indeed accurate) information, which can hinder fair decisions for individuals and create negative outcomes, especially for vulnerable individuals (see, Paterson-Young *et al*, 2024 research on care-experienced individuals in CJS).

The implications of getting assessments of risk wrong are well understood. A recent Independent Serious Further Offence Review (HMIP, 2023c) following the Damien Bendall case, identified excessive workloads, inadequate supervision of junior probation officers, frequent staff turnover and lack of time to properly review cases as factors contributing to the incorrect calculation of risk. Notably, the same issues have been raised in other HMIP Serious Further Offence Reviews, e.g. *Jordan McSweeney* (HMIP, 2023d); *Joshua Jacques* (HMIP, 2024).

Prior to the deployment of technological solutions, it is vital that the realistic conditions in which these tools will be used are considered. The broader issues affecting probation ought to be addressed, and the risks of exacerbating existing problems weighed. Technology should assist and improve outcomes, and not be viewed as ‘sticking plaster’ solutions to wider issues, particularly when they introduce new risks, such as overworked staff being unable to properly oversee and safely/responsibly use the technology to best effect.

Clarity about the use of technological tools

There is lack of clarity about the use of predictive tools in sentencing, which need to be addressed prior to the introduction of further technological tools into the process. As outlined above, risk scores are included in PSRs and inform sentencing decisions, however it is unclear *how* they inform sentences or whether the way in which they inform sentences is consistent. It may be, for example, that a high risk score may make a sentencer more likely to impose a sentence for ‘punishment’ (s.57(2)(a) Sentencing Act 2020) or ‘public protection’ (s.57(2)(d) Sentencing Act 2020), but there is a lack of practice guidance around this – and a lack of evidence about the interaction between predictive risk tools and those using them, which could inform the production of such guidance. There is a need to understand the risk of bias in using predictive risk tools for individuals, especially individuals from marginalised and vulnerable groups.

It is also important to note that the statutory purposes of sentencing encompass competing sentencing objectives, based on either retributive or consequentialist philosophies of punishment, which are not easily reconcilable. Stating broadly that a technological tool should support the purposes of sentencing is therefore insufficiently precise and there should be greater clarity around *which* of the purposes particular technological developments are intended to support. Such guidance could provide the opportunity to encourage sentencers to focus on purposes which are more likely to result in a non-custodial sentence, e.g. ‘reform and rehabilitation’ (s.57(2)(c) Sentencing Act 2020) in appropriate cases.

AI in sentencing

There are existing proposals for AI tools to be used in sentencing to predict the likely sentence in a given jurisdiction. The argument is that they could reduce arbitrariness in judicial decision-making, make

sentences more predictable, and increase the speed with which decisions are made (Bagaric *et al*, 2022). For the reasons outlined below, there are significant issues with these proposals and we would not advise the development or adoption of such tools.

Proposed AI tools for sentencing would be trained on previous judicial decisions to calculate the modal sentence for particular offences, based on a set number of factors. The stated aim is to increase consistency and proportionality in sentencing (Chiao, 2019). A proof-of-concept AI (machine learning) model was developed in New Zealand (but not deployed in the criminal justice system). This tool was trained on 302 previous assault cases and predicted likely sentence lengths with a 'Mean Absolute Error' of about one year (Rodger, Lensen and Betkier, 2023). This can serve as an example of the potential for discrepancies between the values of tool developers and the values of the criminal justice system: the academics developing the tool found this 1-year error 'encouraging', whereas in legal practice this would be considered to represent significant injustice in terms of an 'unjust' sentence (too harsh or too lenient).

Even where a (more) accurate prediction of the modal sentence can be achieved, there are significant concerns about lack of transparency and bias in the use of such tools (Tiarks, 2021). These tools are proposed as a way of minimising judicial bias, but if there is existing judicial bias, the data used to train these tools (previous judicial decisions) are flawed and therefore such bias would be 'baked in', and potentially exacerbated, by the tool. Transparency is difficult to achieve with such tools, as it can be unclear how a particular output has been reached, making it hard to challenge (Villasenor and Foggo, 2020).

Aside from concerns about bias and transparency, it is doubtful whether these tools can – as is suggested – improve consistency and proportionality in sentencing. High quality data – and a large quantity of data – are needed to train effective tools. As things currently stand, the quantity and quality of data available from sentencing hearings is insufficient. Most sentences follow a guilty plea, rather than a trial, so usually no evidence is heard, just a case summary. Proceedings are not recorded in the magistrates courts, where over 90% of criminal cases are concluded. As outlined above, there has been a move towards less information being available in PSRs and judges are also encouraged to avoid lengthy reasons for sentence (Chin-Charles and Cullen [2019] EWCA Crim 1140). Expensive and time-intensive changes would need to be made to the way that sentencing hearings are carried out, if AI tools trained on existing decisions were to be developed, e.g. greater use of full PSRs, longer more detailed sentencing hearings and recording of proceedings in the magistrates courts.

References

Bagaric, M., Svilar, J., Bull, M., Hunter, D. and Stobbs, N. 'The Solution to the Pervasive Bias and Discrimination in the Criminal Justice System: Transparent and Fair Artificial Intelligence' (2022) 59 Am Crim L Rev 95.

Chiao, V. 'Predicting Proportionality: The Case for Algorithmic Sentencing' (2018) 37(3) Criminal Justice Ethics 238.

Chin-Charles and Cullen [2019] EWCA Crim 1140.

HM Inspectorate of Probation (2020) *The quality of pre-sentence information and advice provided to courts*, Research & Analysis Bulletin 2020/04. Manchester: HM Inspectorate of Probation.

HM Inspectorate of Probation (2021) *Race equality in probation: the experiences of black, Asian and minority ethnic probation service users and staff*. Manchester: HM Inspectorate of Probation.

HM Inspectorate of Probation (2023a) *HM Inspectorate of Probation Annual Report 2022/2023*. Manchester: HM Inspectorate of Probation.

HM Inspectorate of Probation (2023b) *Race equality in probation follow-up: A work in progress*. Manchester: HM Inspectorate of Probation.

HM Inspectorate of Probation (2023c) *HM Inspectorate of Probation Independent serious further offence review of Damien Bendall*. Manchester: HM Inspectorate of Probation.

HM Inspectorate of Probation (2023d) *HM Inspectorate of Probation Independent serious further offence review of Jordan McSweeney*. Manchester: HM Inspectorate of Probation.

HM Inspectorate of Probation (2024a) *The quality of pre-sentence information and advice provided to courts – 2022 to 2023 inspections*. Manchester: HM Inspectorate of Probation.

Paterson-Young, C., Corrales, R., Warren, I., McNamara, P. 'Impact of trauma and placement in residential care on the criminalisation of children' (2024) *Children and Youth Services Review*, 166.

Rodger, H., Lensen, A. and Betkier, M. 'Explainable artificial intelligence for assault sentence prediction in New Zealand' (2023) *Journal of the Royal Society of New Zealand*, 53.

Sentencing Act 2020.

Tiarks, E. 'The impact of algorithms on legitimacy in sentencing' (2021) 2 *Journal of Law, Technology & Trust*.

Villasenor, J. and Foggo, V. 'Artificial Intelligence, Due Process and Criminal Sentencing' (2020) *Mich St L Rev* 295.