

Tamiyyah Shafiq

If algorithmic sentencing can be proved to be able to approximate a just sentence that is equal to or better than that delivered by a human sentencer, we would be morally obligated to employ this in our judicial system in interest of efficiency and more standardized sentences across the country by reducing human bias and error. To see if this statement is true, we can disprove the converse, that human sentencing is inherently superior to algorithmic sentencing.

Proponents of the converse claim that human sentencing is superior because the variety of factors that go into a case are too complex for an algorithm to process. This argument ignores the fact that every case has a multitude of factors that relate, but only a fraction of factors are relevant to the sentencing, and these are decided by the sentencing theory being used. In fact, human judgement is less suitable in weighing multiple factors, as humans often resort to “system 1 thinking” or intuitive judgement when faced with making a decision. Therefore, the claim that human sentencing is inherently superior is false.

There are three major concerns when it comes to algorithmic sentencing: privacy, transparency, and bias.

The first is the loss of privacy of the data used in algorithmic sentencing. This refers to two categories of data: that used in the developmental phase, and that used in the duration of the case. For the first, that would only be relevant to data if the model is trained using machine learning, which I would argue is unnecessary and even discouraged in training this sort of model, as it can introduce bias that has been present in history through example cases that may have been unfair. Rather, it would be better to train the model using theory of just sentencing, which would lead to the most fair sentencing of a case as it will comply solely to the most correct theory of just sentencing without being subject to human bias or error. For the second category of data, that used during the case, the algorithm has no more data than that of a human judge.

The second concern is transparency, saying that the logic behind the conclusion made by an algorithmic sentencer is opaque and therefore one cannot challenge or review the decision, or follow the logic behind it. This would be the case if private companies program the model and want to hide the code for business purposes, but there is no reason for this to be the norm. If a model is made fully available to the public, this would not be a concern. Furthermore, human-decision making is no more transparent than that of this model.

The third concern is that of bias, and this is again of two types: bias reproduced in training data, or bias that becomes present when looking at general trends that may differ across different relevant populations. The bias reproduced in training data can be avoided if machine learning is not used, as mentioned previously. As for the second bias, this is also present within human judges, and can perhaps be dealt with by having direct discrimination in the form of leniency towards a minority. Regardless, this bias is no less present in human judges, and algorithmic sentencing reduces the first bias which is present in humans.

Therefore, because algorithmic sentencing has been proved to be as just and even more efficient than human sentences, we should employ this practice in our judicial system.

Sources:

- <https://philpapers.org/archive/THOIEIEM-2.pdf>
- <https://www.tandfonline.com/doi/full/10.1080/0731129X.2023.2275967>
- <https://link.springer.com/article/10.1007/s43681-024-00442-5>