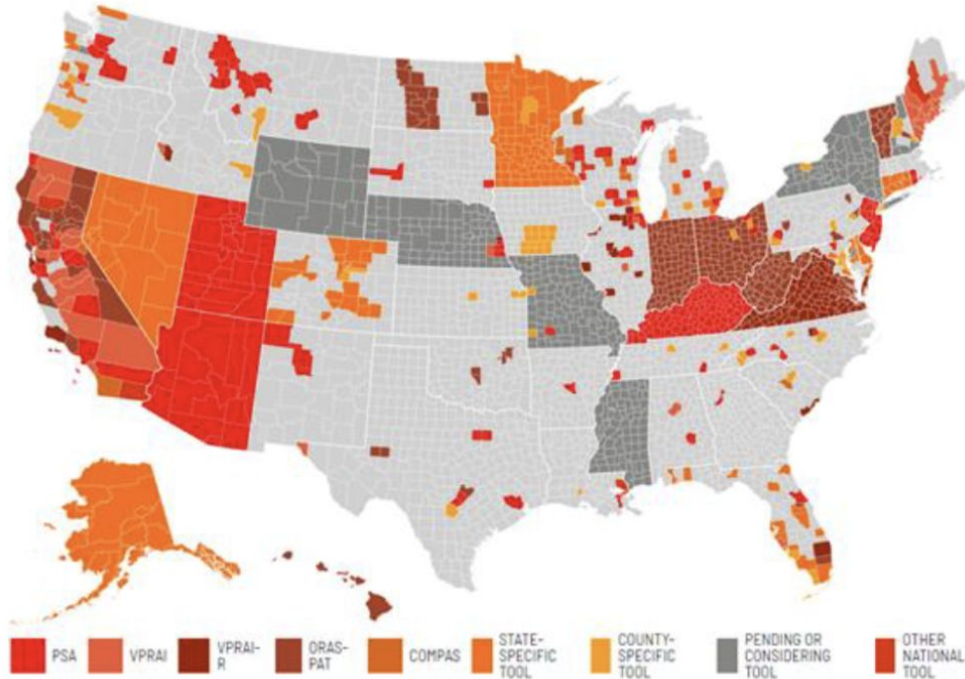


Deus Ex Machina: Algorithmic Sentencing?

Adrian Turkedjiev and Tamiyyah Shafiq

Overview

Figure 1. Adoption of Recidivism-Risk Assessment Tools across States¹



Assumption: the concept of just sentencing exists

if algorithmic sentencing can approximate a just sentence (equal to or better than humans), we have moral reason to use the sentencing method that best approximates a just sentence

Algorithmic sentencing: efficiency and standardization, reducing human bias and random errors

if algorithmic sentencing can approximate a just sentence (equal to or better than humans)

Human sentencing is inherently superior?

- The variety of factors is too complex for an algorithm
- Human judgement is inherently better

concerns

- 1) Privacy
 - a) Developmental
 - b) Use
- 2) Transparency
- 3) Bias
 - a) Reproducing bias in training data
 - b) Factors that vary across different relevant populations

therefore...we have moral reason to use the sentencing method that best approximates a just sentence

therefore...we have moral reason to employ algorithmic sentencing

Structural Algorithmic Flaws:

1. Algorithms are trained on data from past
 - a. This data is influenced by biases which we have only recently began to combat
 - i. African Americans “suffer from a lower probability of receiving alternative punishments, a higher chance of incarceration, and longer imprisonment time by **5.6%, 4.3%, and 30.7%**, respectively” (Tulane, 2023)
 - b. Continuing to act based on these algorithms makes their preconceived judgements a self-fulfilling prophecy. Consider the following:
 - i. *Members of group X have tended to not get parole within algorithm Y*
 - ii. *Thus, member of group X continue to not get parole under algorithm Y*
 - iii. *Thus, algorithm Y never receives data to make it believe group X ought get parole*
 - iv. *Thus, members of group X continuously do not get parola from algorithm Y*
2. Companies have incentives to use data from across different jurisdictions, meaning their algorithms are not trained in coherence to the laws of a particular jurisdiction

Implementation challenges

1. Judges are unduly biased towards consideration of the results of these algorithms, as they have an air of scientific legitimacy
 1. E.g., In *Zilly v. State of Wisconsin*
2. The algorithms themselves are often protected by trade secrets, making scrutiny of them nearly impossible
3. Private developers with whom the government works to create these models have an undue degree of control over what data is used in training
4. Prosecutors cannot simply shop around for a preferred algorithm to use, as they likely have skewed incentives in terms of the behavior they want from these algorithms

Legal concerns

1. Any judge or defendant must have full access to that which was used in deciding their case in order to ensure due process and allow for legitimate due process challenges. Algorithmic sentencing prevents this in three ways:
 1. Companies can avoid disclosing the algorithm via trade secret laws
 2. Even if a defendant or judge has access to the source code of an algorithm, they are unlikely to be able to interpret it well
 3. The 'black box problem' in AI means developers themselves are unlikely to fully understand why the algorithm is producing the results it is producing