SAFETY+JUSTICE CHALLENGE

THE PRESENT AND FUTURE OF AI IN PRE-TRIAL RISK ASSESSMENT INSTRUMENTS

BY ALEXANDRA CHOULDECHOVA AND KRISTIAN LUM

From music and romantic partner recommendation, to medical diagnosis and disease outbreak detection, to automated essay scoring, "Artificial intelligence" (AI) systems are being used to tackle prediction, classification, and detection tasks that impact nearly every sphere of our lives. Since the fundamental task of pre-trial risk assessment instruments is one of prediction, we anticipate that the success of AI technology in these other domains will inspire an increase in the availability of AI-based pre-trial risk assessment instruments in the coming years. **The purpose of this critical issue brief is primarily to equip practitioners considering adopting an AI-based pre-trial risk assessment tool to consider questions relevant to determining whether adopting such a system a will result in better predictions and ultimately move their jurisdiction towards fairer, more just and decarceral pre-trial decisionmaking that respects civil and human rights.**

Though AI may sound futuristic, in practice the term is used as a catch-all to describe any computational technologies that are capable of producing reasoned or "intelligent" outputs. Many currently available actuarial pre-trial risk assessment tools are themselves a form of AI. As such, the pertinent questions for deciding whether an "AI-based" pre-trial risk assessment tool is appropriate for a jurisdiction are largely the same as those one would consider before deciding to adopt any actuarial risk assessment tool. These include,

• What population was used to train the model? Is this different from my local population? Has the model been locally validated?
• Is the data measured accurately and without bias? What *exactly* is the model predicting?

• Does the process for obtaining the inputs for future assessments respect the rights and dignity of the accused?
• Is the model development process purely computational or does it require manual adjustment? If so, how and why were each of the adjustments made?
• Are there racial, ethnic, gender, or any other relevant disparities in the model's predictions?

More complex AI-based pre-trial risk assessment models will be subject to many of the same critiques and concerns that have been raised about current versions of pre-trial actuarial risk assessment tools. In most cases, we believe the issues with current tools will be magnified by the new data sources and more complex model-building approaches that are to come marketed as AI. Some additional questions to consider to assess to what extent an AI-based tool is worsening current issues are,

• Have inputs that themselves are the outputs of other AI systems been locally validated and found to be unbiased? Can those inputs be audited or contested?
• Are surveillance-based inputs derived from unevenly distributed surveillance systems?
• Does the tool broaden the definition of unacceptable pre-trial behaviors or widen the net of those eligible for pre-trial supervision or detention?
• Does the data contain any information that was obtained via legally or ethically questionable methods?
• Does collecting data to administer the assessment in the future require any morally objectionable or overly invasive procedures?
• Is the model understandable? Are the gains in predictive accuracy sufficient to offset the loss in interpretability?

Ultimately, questions that can be asked at the procurement stage are only part of the story and can only gauge the appropriateness of the model in isolation, not in the real world environment in which it will be used. Follow-up studies that assess how the model impacts pre-trial decision-making, including whether the deployment of the model resulted in reductions in the pre-trial population and in reduced racial disparities, are vital to understanding whether any model—AI-based or not—is having the intended effects.