

Data Sources that would improve our MVP

We can give the bias detection group information about the data we would ideally like to have to plug into an AI system that would detect bias, and give us the type of info we see in our prototype: top sheet, recommendations, links to data, reasons the system gave the recommendations. Plus the ability to annotate documents to explain its reasoning and show evidence. The bias detection team wants more detail on what factors go into charging people and choosing sentences. We can take a first pass but more user research would be best. User research could also help us understand more about the art and science of negotiating plea deals and sentences. How much punishment is a crime "worth"?

Key items are marked in bold.

Things we would ask in our questionnaire

- Have you experienced bias already?
- **The "priors" of the person accused: prior convictions (not AI)**
 - ideally, we could get some data on how fair the previous outcomes were!
 - could just be the questionnaire as a start
 - arrests
 - convictions
 - time in jail
 - violent vs. non violent offenses
 - question - what are they allowed to consider?
 - ideally, we could get some data on how fair the previous outcomes were!
- **Demographic information about the defendant (not AI)** - see the data in "data elements" <https://ibm.ent.box.com/notes/693387649567>
- **Personal history information about the defendant (not AI)**
 - from questionnaire

- Do they think they're innocent?

Data relevant to deciding on charging

- what the crimes are in the jurisdiction - what can you even be charged with?
- sentencing guidelines - what are the permitted sentences for each conviction?
- To classify crimes based on text input, we would need a database of text inputs for thousands of crimes in the past
- Historical data on what sentences people actually received for each charge
- Historical data on what sentences people actually served (vs. early release from jail)
- Historical data on conviction rates if you go to trial
 - Group by demographics
- Everything that's on our questionnaire is relevant data
- The fact pattern https://en.wikipedia.org/wiki/Fact_pattern
 - Find an example of a fact pattern for one crime
- Descriptions of the crime
 - Interviews
 - Evidence
 - Police reports
- What happens after people serve their sentence?
 - Likelihood to re-offend
 - What sentence could make them less likely to re-offend
 - Mental health treatment for example
 - Routing homeless people to social services
 - Prison
 - Probation
- How well does the crime fit a charge (how extreme is that charge compared to the reality of the crime)
- If using automation guidelines, what are the standards that they use to determine the outcomes?
- Case precedent - what other cases are similar to this one?
- Have similar crimes in the jurisdiction been charged the same way?

- Give me a group of similar crimes to compare to
- How bias is currently showing up in the charging outcomes
 - Especially for crimes similar to this one
 - Lesser charges for privileged groups?
- The "priors" of the person accused
- The impact to the accused of the charge - will it have an outsized/unfair impact on their life
 - Do they consider this at charging/plea bargaining time?
- How likely they are to avoid going to court
 - Incentives to prosecutors - how are they incentivized to avoid going to court?
 - The system can't handle every case going to trial - not enough court capacity
- Age/health of the defendant
- Prosecutor's history
 - Plea bargains
 - Trial convictions
 - Found not guilty
 - Not even charged
- strength of evidence?
 - Could AI help assess this?
- type of evidence?
 - physical evidence
 - testimony
- How likely is it that they're actually innocent?
 - At least from the initial data
- Factors that will help the accused decide whether to go to trial
 - Amount of time before the trial
 - Will they be in jail while waiting for the trial
 - Likely outcome of bail hearing
 - Cost in \$\$ of legal representation

- Testimony
- Did the person convicted have good legal representation?
 - Type of representation: self, public defender, private attorney
- What are the most common charges in this jurisdiction?
- Keywords or tone flags that might indicate bias
- Defense: How can we humanize the defendant? Get more sympathy?
 - Childhood: abuse, drugs, underlying problems
- DA: their focus is on the crime

Data relevant to deciding sentencing

- What the sentencing guidelines are for each crime in this jurisdiction
- Case precedent - what other cases are similar to this one?
- Other rules that come into play, like the "three strikes you're out" rule
 - mandatory minimums
- How bias is currently showing up in the sentencing outcomes
- What does the prosecutor consider to be the worth of the crime?
- Are the other types of sentences that would have a better impact/prevent the crime from occurring again?
- Was the defendant in jail during this time?
- have similar crimes in the jurisdiction been sentenced the same way?
- If charges are higher than they should be, is there a way to equalize it with the sentence?
- The impact to the accused of the sentence - will it have an outsized/unfair impact on their life
- what are the "facts of the case"
- For this type of crime, was it a more severe or less severe offense?
- Text/recorded description of the crime
- Text/recorded description of the impact to the victim
- Did the person convicted have good legal representation?
 - Type of representation: self, public defender, private attorney

Data on actual outcomes

- Start with data similar to what we see in the federal sentencing data (see: Data Elements <https://ibm.ent.box.com/notes/693387649567>)
- Text from dockets - what the judge wrote
 - transcripts of the court notes
- For federal cases, the text of the decisions
- Whether the decisions of lower courts were upheld or overturned
- Decisions from courts
 - the crime
 - demographic of the defendant
- Text/descriptions of the other points in the process - pre-trial motions and so on