

API Interface with Bias Detection System (Proposed)

We can give the bias detection group information about the outputs we want to 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 outputs we could use.

Key items are marked in bold.

Data relevant to public defenders' tool

Most likely-to-help items in **bold**.

- **How well does the crime fit a charge (AI - bias detection engine)**
 - how extreme is that charge compared to the reality of the crime
 - statistical component - 90% like this crime, 45% like this one
 - Have similar crimes in the jurisdiction been charged the same way?
- **Case precedent - what other cases are similar to this one? (AI - bias detection engine)**
 - outcomes
- **Give me a group of similar crimes to compare this to (AI - bias detection engine)**
- **Average sentencing data by conviction (not AI - federal sentencing data))**
- **Disparities in sentencing by race (not AI - federal sentencing data)**
- **How bias is currently showing up in the charging outcomes (AI - bias detection engine)**
 - **Especially for crimes similar to this one**
 - **Lesser charges for privileged groups?**
- **The "priors" of the person accused: prior convictions (not AI - from our questionnaire)**
 - ideally, we could get some data on how fair the previous outcomes were!
 - could just be the questionnaire as a start
- **Demographic information about the defendant (not AI - from our questionnaire)**

- **Personal history information about the defendant (not AI - from our questionnaire))**
 - from our questionnaire or other data that's already been entered into the system
- Factors that will help the accused decide whether to go to trial - hard data based on the charges (not AI)
 - Likely outcome of bail hearing - what percentage get bail, what percentage sit in jail and for how long
 - If you plead not guilty, what are your chances of being found guilty vs. innocent
 - If you're found guilty, what are the usual sentences?
- strength of evidence?
 - Could AI help assess this?
 - tricky
- How likely is it that they're actually innocent?
 - At least from the initial data
- Victim factors
- Keywords or tone flags that might indicate bias in the case documents (AI)
- **What happens after people serve their sentence? By sentence type. (not AI)**
 - Likelihood to re-offend or be rehabilitated
 - Help the public defender argue for a different option
 - What sentence could make them less likely to re-offend
 - Mental health treatment for example
 - Routing homeless people to social services
 - jail time
 - probation
 - home arrest
 - School-to-prison pipeline?

Data relevant to public dashboard

- **How bias is currently showing up in the sentencing outcomes: aggregate data (Some AI, some not)**
 - **Ideally, show the data per specific judge, prosecutor to help voters**
 - **Also, are they acting in accordance with their public statements?**

- **Less ideally, show the data per jurisdiction**
- **Could also show the data by type of crime**
- **Some kind of "bias score"?**
- **Evidence and sources**
- **Are there other types of sentences that would have a better impact/prevent the crime from occurring again? (Some AI, some not)**
 - **Help with policy**
- **Are mandatory sentencing guidelines contributing to biased outcomes?**
- **Did the person convicted have good legal representation?**
 - Type of representation: self, public defender, private attorney
- **Disparities in sentencing by type of representation (Some AI, some not)**
- **What happens after people serve their sentence? By sentence type. (Some AI, some not)**
 - Likelihood to re-offend or be rehabilitated
 - What sentence could make them less likely to re-offend
 - Mental health treatment for example
 - Routing homeless people to social services
 - jail time
 - probation
 - home arrest
 - School-to-prison pipeline?