



# Value Cards

## Model Cards

8 Models

1 Model Overview

## Persona Cards

Judge

Defendants

Fairness Advocate

Community Member

## Checklist Card

Understanding Societal Values in AI

Identifying Stakeholders

Analyzing Impacts

### Model Number: 1

#### Accuracy

Over all: 68.4%

White American: 71.3%

African American: 65.5%

#### Disparity

Disparity in Accuracy: 5.8%

Disparity in FPR: 6.1%

Disparity in FNR: 14.9%

#### False Positive Rate

Over all: 8.0%

White American: 5.2%

African American: 11.3%

#### False Negative Rate

Over all: 74.6%

White American: 83.2%

African American: 68.3%

### Model Number: 2

#### Accuracy

Over all: 60.0%

White American: 60.0%

African American: 60.0%

#### Disparity

Disparity in Accuracy: 0%

Disparity in FPR: 6.4%

Disparity in FNR: 9.2%

#### False Positive Rate

Over all: 45.5%

White American: 42.1%

African American: 49.5%

#### False Negative Rate

Over all: 29.8%

White American: 35.1%

African American: 25.9%

### Model Number: 3

#### Accuracy

Over all: 70.4%

White American: 72.9%

African American: 68.0%

#### Disparity

Disparity in Accuracy: 4.9%

Disparity in FPR: 7%

Disparity in FNR: 9%

#### False Positive Rate

Over all: 10.1%

White American: 5.7%

African American: 15.1%

#### False Negative Rate

Over all: 65.2%

White American: 77.1%

African American: 56.5%

### Model Number: 4

#### Accuracy

Over all: 64.1%

White American: 65.2%

African American: 63.1%

#### Disparity

Disparity in Accuracy: 2.1%

Disparity in FPR: 6.7%

Disparity in FNR: 9.8%

#### False Positive Rate

Over all: 31.9%

White American: 28.9%

African American: 35.6%

#### False Negative Rate

Over all: 42.8%

White American: 48.5%

African American: 38.7%

### Model Number: 5

#### Accuracy

Over all: 66.7%

White American: 70.6%

African American: 62.7%

#### Disparity

Disparity in Accuracy: 7.9%

Disparity in FPR: 2.6%

Disparity in FNR: 9.8%

#### False Positive Rate

Over all: 1.9%

White American: 0.8%

African American: 3.4%

#### False Negative Rate

Over all: 90.5%

White American: 96.2%

African American: 86.4%

### Model Number: 6

#### Accuracy

Over all: 67.6%

White American: 69.9%

African American: 65.2%

#### Disparity

Disparity in Accuracy: 4.7%

Disparity in FPR: 18.0%

Disparity in FNR: 23.2%

#### False Positive Rate

Over all: 30.6%

White American: 21.9%

African American: 40.8%

#### False Negative Rate

Over all: 35.6%

White American: 49.0%

African American: 25.8%

### Model Number: 7

#### Accuracy

Over all: 69.1%

White American: 72.6%

African American: 65.4%

#### Disparity

Disparity in Accuracy: 7.2%

Disparity in FPR: 6.9%

Disparity in FNR: 15.1%

#### False Positive Rate

Over all: 4.7%

White American: 1.5%

African American: 8.4%

#### False Negative Rate

Over all: 78.7%

White American: 87.4%

African American: 72.3%

### Model Number: 8

#### Accuracy

Over all: 60%

White American: 60.8%

African American: 57.3%

#### Disparity

Disparity in Accuracy: 3.5%

Disparity in FPR: 19.4%

Disparity in FNR: 14.6%

#### False Positive Rate

Over all: 53.5%

White American: 44.6%

African American: 64%

#### False Negative Rate

Over all: 18%

White American: 26.6%

African American: 12%

Attribute Model	Accuracy	FPR	FNR	Dis/Acc	Dis/FPR	Dis/FNR
1	★					
2	✗	✗	★	★		
3	★					
4				★		
5		★	✗		★	
6					✗	✗
7	★	★	✗			
8	✗	✗	★	★	✗	

**Accuracy Range:**  
43% - 71%

**FNR Range:**  
3.4% - 94%

**FPR Range:**  
1% - 87%

✗ performs relatively badly

★ performs relatively well

**Disclaimer:** The performances are evaluated among the models, and only serve as a subjective reference. Please **always** refer to the model cards for model details.

## Persona: Judge

As a judge, you care most about making the right decision when sentencing a defendant. You probably want the model to have a **higher accuracy**.

However, your teammates, who play other roles in the system, might prioritize other metrics. Throughout the deliberation, please express your value, respect their concerns, and negotiate with them.



## Persona: Defendants

As a defendant, you are most worried about being falsely predicted as “will offend again.” In this case, you probably want the model to have a **lower false positive rate**.

However, your teammates, who play other roles in the system, might prioritize other metrics. Throughout the deliberation, please express your value, respect their concerns, and negotiate with them.



## Persona: Fairness Advocate

As a fairness advocate, you want to prevent the unfair treatment on some demographic. You don't want the recidivism algorithm to be biased against one demographic. In this case, you want the model to have a **lower disparity**.

However, your teammates, who play other roles in the system, might prioritize other metrics. Throughout the deliberation, please express your value, respect their concerns, and negotiate with them.

## Persona: Community Member

As a community member, you want your neighborhood to be safe. You are concerned about re-offending behavior in your community. In this case, you may care mostly about **false negative rate**.

However, your teammates, who play other roles in the system, might prioritize other metrics. Throughout the deliberation, please express your value, respect their concerns, and negotiate with them.

## Checklists

### 1. Understanding societal values in AI

- There is no single definition of societal values that will apply equally well to different applications of AI.
- Prioritizing one value in AI systems often means making tradeoffs on competing values. It is therefore important to be explicit and transparent about priorities and assumptions.
- There are seldom clear-cut answers. It is therefore important to document your processes and considerations (including priorities and tradeoffs).

## Checklists

### 2. Identifying stakeholders

- **Who is at risk of experiencing impacts:** considering both the people who will use the system and the people who will be directly or indirectly affected by the system, either by choice or not.
- **People often belong to overlapping groups**—different combinations of race, gender, and age, for example—and specific intersectional groups may be at greatest risk of experiencing different types of harm.
- **What are the tradeoffs** between expected benefits and potential harms for identified stakeholder groups?

## Checklists

### 3. Analyzing impacts

- What are the **types of impact** (e.g., allocation, quality of service, stereotyping, denigration, over- or underrepresentation) on different stakeholders
- Rate the **degree** of impact [no discernable - minor - moderate - major]
- Estimate the **scale** of impact [small - medium - large]
- Estimate the **overall direction** of impact [positive - mostly positive - mostly negative - negative]