

# Incremental AI

# Incremental AI

- **Backlash to AI vs. Incremental AI**

- ▶ potential bias, reduced autonomy, status quo, transparency, erosion of learning, adversarial attack

- In Stage 0, assess judges vs. a bootstrapped judge (predicted decision-maker)
- In Stage 1, people use AI as a support tool, speeding up existing processes (for example, by prefilling forms)
- Once they're used to this, they can more easily accept an added functionality (Stage 2) in which AI becomes a choice monitor, pointing out choice inconsistencies (pay more attention / be less indifferent)
- Stage 3 elevates the AI to the role of a more general coach, providing outcome feedback on choices and highlighting decision patterns.
  - ▶ Transparent + explainable | explain why deviate
- Then, in Stage 4, the AI brings in other people's decision histories and patterns, serving as a platform for a community of experts.
  - ▶ (1) self-image, (2) self-improvement, (3) self-understanding, (4) ego
- Only in Stage 5, recommend the 'optimal decision'

# Incremental AI

- Backlash to AI vs. Incremental AI

- ▶ potential bias, reduced autonomy, status quo, transparency, erosion of learning, adversarial attack

- In Stage 0, assess judges vs. a bootstrapped judge (predicted decision-maker)
- In Stage 1, people use AI as a support tool, speeding up existing processes (for example, by prefilling forms)
- Once they're used to this, they can more easily accept an added functionality (Stage 2) in which AI becomes a choice monitor, pointing out choice inconsistencies (pay more attention / be less indifferent)
- Stage 3 elevates the AI to the role of a more general coach, providing outcome feedback on choices and highlighting decision patterns.
  - ▶ Transparent + explainable | explain why deviate
- Then, in Stage 4, the AI brings in other people's decision histories and patterns, serving as a platform for a community of experts.
  - ▶ (1) self-image, (2) self-improvement, (3) self-understanding, (4) ego
- Only in Stage 5, recommend the 'optimal decision'

# Incremental AI

- Backlash to AI vs. Incremental AI
  - ▶ potential bias, reduced autonomy, status quo, transparency, erosion of learning, adversarial attack
- In Stage 0, assess judges vs. a bootstrapped judge (predicted decision-maker)
- In Stage 1, people use AI as a support tool, speeding up existing processes (for example, by prefilling forms)
- Once they're used to this, they can more easily accept an added functionality (Stage 2) in which AI becomes a choice monitor, pointing out choice inconsistencies (pay more attention / be less indifferent)
- Stage 3 elevates the AI to the role of a more general coach, providing outcome feedback on choices and highlighting decision patterns.
  - ▶ Transparent + explainable | explain why deviate
- Then, in Stage 4, the AI brings in other people's decision histories and patterns, serving as a platform for a community of experts.
  - ▶ (1) self-image, (2) self-improvement, (3) self-understanding, (4) ego
- Only in Stage 5, recommend the 'optimal decision'

# Incremental AI

- Backlash to AI vs. Incremental AI
  - ▶ potential bias, reduced autonomy, status quo, transparency, erosion of learning, adversarial attack
- In Stage 0, assess judges vs. a bootstrapped judge (predicted decision-maker)
- In Stage 1, people use AI as a support tool, speeding up existing processes (for example, by prefilling forms)
- Once they're used to this, they can more easily accept an added functionality (Stage 2) in which AI becomes a choice monitor, pointing out choice inconsistencies (pay more attention / be less indifferent)
- Stage 3 elevates the AI to the role of a more general coach, providing outcome feedback on choices and highlighting decision patterns.
  - ▶ Transparent + explainable | explain why deviate
- Then, in Stage 4, the AI brings in other people's decision histories and patterns, serving as a platform for a community of experts.
  - ▶ (1) self-image, (2) self-improvement, (3) self-understanding, (4) ego
- Only in Stage 5, recommend the 'optimal decision'

# Incremental AI

- Backlash to AI vs. Incremental AI
  - ▶ potential bias, reduced autonomy, status quo, transparency, erosion of learning, adversarial attack
- In Stage 0, assess judges vs. a bootstrapped judge (predicted decision-maker)
- In Stage 1, people use AI as a support tool, speeding up existing processes (for example, by prefilling forms)
- Once they're used to this, they can more easily accept an added functionality (Stage 2) in which AI becomes a choice monitor, pointing out choice inconsistencies (pay more attention / be less indifferent)
- Stage 3 elevates the AI to the role of a more general coach, providing outcome feedback on choices and highlighting decision patterns.
  - ▶ Transparent + explainable | explain why deviate
- Then, in Stage 4, the AI brings in other people's decision histories and patterns, serving as a platform for a community of experts.
  - ▶ (1) self-image, (2) self-improvement, (3) self-understanding, (4) ego
- Only in Stage 5, recommend the 'optimal decision'

# Incremental AI

- Backlash to AI vs. Incremental AI
  - ▶ potential bias, reduced autonomy, status quo, transparency, erosion of learning, adversarial attack
- In Stage 0, assess judges vs. a bootstrapped judge (predicted decision-maker)
- In Stage 1, people use AI as a support tool, speeding up existing processes (for example, by prefilling forms)
- Once they're used to this, they can more easily accept an added functionality (Stage 2) in which AI becomes a choice monitor, pointing out choice inconsistencies (pay more attention / be less indifferent)
- Stage 3 elevates the AI to the role of a more general coach, providing outcome feedback on choices and highlighting decision patterns.
  - ▶ Transparent + explainable | explain why deviate
- Then, in Stage 4, the AI brings in other people's decision histories and patterns, serving as a platform for a community of experts.
  - ▶ (1) self-image, (2) self-improvement, (3) self-understanding, (4) ego
- Only in Stage 5, recommend the 'optimal decision'

# Incremental AI

- Backlash to AI vs. Incremental AI
  - ▶ potential bias, reduced autonomy, status quo, transparency, erosion of learning, adversarial attack
- In Stage 0, assess judges vs. a bootstrapped judge (predicted decision-maker)
- In Stage 1, people use AI as a support tool, speeding up existing processes (for example, by prefilling forms)
- Once they're used to this, they can more easily accept an added functionality (Stage 2) in which AI becomes a choice monitor, pointing out choice inconsistencies (pay more attention / be less indifferent)
- Stage 3 elevates the AI to the role of a more general coach, providing outcome feedback on choices and highlighting decision patterns.
  - ▶ Transparent + explainable | explain why deviate
- Then, in Stage 4, the AI brings in other people's decision histories and patterns, serving as a platform for a community of experts.
  - ▶ (1) self-image, (2) self-improvement, (3) self-understanding, (4) ego
- Only in Stage 5, recommend the 'optimal decision'

# Incremental AI

- Backlash to AI vs. Incremental AI
  - ▶ potential bias, reduced autonomy, status quo, transparency, erosion of learning, adversarial attack
- In Stage 0, assess judges vs. a bootstrapped judge (predicted decision-maker)
- In Stage 1, people use AI as a support tool, speeding up existing processes (for example, by prefilling forms)
- Once they're used to this, they can more easily accept an added functionality (Stage 2) in which AI becomes a choice monitor, pointing out choice inconsistencies (pay more attention / be less indifferent)
- Stage 3 elevates the AI to the role of a more general coach, providing outcome feedback on choices and highlighting decision patterns.
  - ▶ Transparent + explainable | explain why deviate
- Then, in Stage 4, the AI brings in other people's decision histories and patterns, serving as a platform for a community of experts.
  - ▶ (1) self-image, (2) self-improvement, (3) self-understanding, (4) ego
- Only in Stage 5, recommend the 'optimal decision'

# Incremental AI

- Backlash to AI vs. Incremental AI
  - ▶ potential bias, reduced autonomy, status quo, transparency, erosion of learning, adversarial attack
- In Stage 0, assess judges vs. a bootstrapped judge (predicted decision-maker)
- In Stage 1, people use AI as a support tool, speeding up existing processes (for example, by prefilling forms)
- Once they're used to this, they can more easily accept an added functionality (Stage 2) in which AI becomes a choice monitor, pointing out choice inconsistencies (pay more attention / be less indifferent)
- Stage 3 elevates the AI to the role of a more general coach, providing outcome feedback on choices and highlighting decision patterns.
  - ▶ Transparent + explainable | explain why deviate
- Then, in Stage 4, the AI brings in other people's decision histories and patterns, serving as a platform for a community of experts.
  - ▶ (1) self-image, (2) self-improvement, (3) self-understanding, (4) ego
- Only in Stage 5, recommend the 'optimal decision'

# Incremental AI

- Backlash to AI vs. Incremental AI
  - ▶ potential bias, reduced autonomy, status quo, transparency, erosion of learning, adversarial attack
- In Stage 0, assess judges vs. a bootstrapped judge (predicted decision-maker)
- In Stage 1, people use AI as a support tool, speeding up existing processes (for example, by prefilling forms)
- Once they're used to this, they can more easily accept an added functionality (Stage 2) in which AI becomes a choice monitor, pointing out choice inconsistencies (pay more attention / be less indifferent)
- Stage 3 elevates the AI to the role of a more general coach, providing outcome feedback on choices and highlighting decision patterns.
  - ▶ Transparent + explainable | explain why deviate
- Then, in Stage 4, the AI brings in other people's decision histories and patterns, serving as a platform for a community of experts.
  - ▶ (1) self-image, (2) self-improvement, (3) self-understanding, (4) ego
- Only in Stage 5, recommend the 'optimal decision'

“Ferguson and the Violence of Indifference” (Cuenca 2017)

The New York Times

---

# *Before an Arrest, Officers Tossed a (Virtual) Coin*

July 14, 2018

Justice: equal treatment before the law ( $y = f(X) + \varepsilon, a \rightarrow X$ )  
equality based on recognition of difference  
( $y \perp W, \text{var}(\varepsilon) \perp W, a \rightarrow W$ )

control principle and merit principle: individuals liable only for events that are under their control  
W: race, gender, masculinity, name, football, weather, judge's lunchtime, preceding case, ...

“Ferguson and the Violence of Indifference” (Cuenca 2017)

The New York Times

---

# *Before an Arrest, Officers Tossed a (Virtual) Coin*

July 14, 2018

**Justice:** equal treatment before the law ( $y = f(X) + \varepsilon, a \rightarrow X$ )  
equality based on recognition of difference  
( $y \perp W, \text{var}(\varepsilon) \perp W, a \rightarrow W$ )

control principle and merit principle: individuals liable only for events that are under their control  
W: race, gender, masculinity, name, football, weather, judge's lunchtime, preceding case, ...

The New York Times

---

# *Before an Arrest, Officers Tossed a (Virtual) Coin*

July 14, 2018

**Justice:** equal treatment before the law ( $y = f(X) + \varepsilon, a \rightarrow X$ )  
equality based on recognition of difference  
( $y \perp W, \text{var}(\varepsilon) \perp W, a \rightarrow W$ )

control principle and merit principle: individuals liable only for events that are under their control

W: race, gender, masculinity, name, football, weather, judge's lunchtime, preceding case, ...

The New York Times

---

# *Before an Arrest, Officers Tossed a (Virtual) Coin*

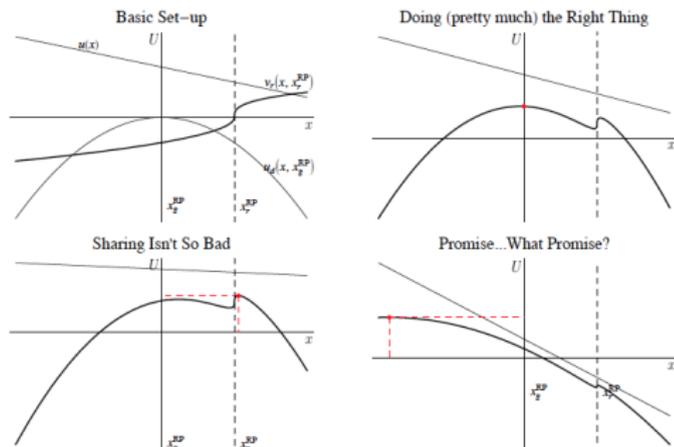
July 14, 2018

**Justice:** equal treatment before the law ( $y = f(X) + \varepsilon, a \rightarrow X$ )  
equality based on recognition of difference  
( $y \perp W, \text{var}(\varepsilon) \perp W, a \not\rightarrow W$ )

control principle and merit principle: individuals liable only for events that are under their control  
W: race, gender, masculinity, name, football, weather, judge's lunchtime, preceding case, ...

# Judicial Inattention

- Behavioral anomalies offer intuitive understanding of feature relevance
- “settings where people are closer to *indifference* among options are more likely to lead to detectable effects [of behavioral biases] outside of it.” (Simonsohn, JPSP 2011)



A model of recognition-respect and  
revealed preference indifference

# Using ML to Diagnose Judicial Inattention

- 1 Early predictability
- 2 Behavioral anomalies
- 3 If systematic indifference, judge identity might predict appeal
- 4 Inattentiveness to appellate reversals
- 5 Implicit risk rankings of asylees closer to random
- 6 Is indifference greater for some refugees (e.g., from Global South)?
- 7 Can we use judicial analytics to increase recognition & dignity?

# Using ML to Diagnose Judicial Inattention

- 1 Early predictability
- 2 Behavioral anomalies
- 3 If systematic indifference, judge identity might predict appeal
- 4 Inattentiveness to appellate reversals
- 5 Implicit risk rankings of asylees closer to random
- 6 Is indifference greater for some refugees (e.g., from Global South)?
- 7 Can we use judicial analytics to increase recognition & dignity?

# Using ML to Diagnose Judicial Inattention

- 1 Early predictability
- 2 Behavioral anomalies
- 3 If systematic indifference, judge identity might predict appeal
- 4 Inattentiveness to appellate reversals
- 5 Implicit risk rankings of asylees closer to random
- 6 Is indifference greater for some refugees (e.g., from Global South)?
- 7 Can we use judicial analytics to increase recognition & dignity?

# Using ML to Diagnose Judicial Inattention

- 1 Early predictability
- 2 Behavioral anomalies
- 3 If systematic indifference, judge identity might predict appeal
- 4 Inattentiveness to appellate reversals
- 5 Implicit risk rankings of asylees closer to random
- 6 Is indifference greater for some refugees (e.g., from Global South)?
- 7 Can we use judicial analytics to increase recognition & dignity?

# Using ML to Diagnose Judicial Inattention

- 1 Early predictability
- 2 Behavioral anomalies
- 3 If systematic indifference, judge identity might predict appeal
- 4 Inattentiveness to appellate reversals
- 5 Implicit risk rankings of asylees closer to random
- 6 Is indifference greater for some refugees (e.g., from Global South)?
- 7 Can we use judicial analytics to increase recognition & dignity?

# Using ML to Diagnose Judicial Inattention

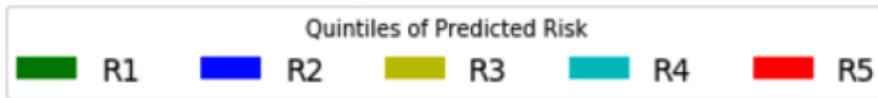
- 1 Early predictability
- 2 Behavioral anomalies
- 3 If systematic indifference, judge identity might predict appeal
- 4 Inattentiveness to appellate reversals
- 5 Implicit risk rankings of asylees closer to random
- 6 Is indifference greater for some refugees (e.g., from Global South)?
- 7 Can we use judicial analytics to increase recognition & dignity?

# Using ML to Diagnose Judicial Inattention

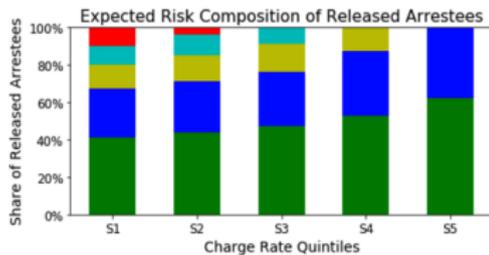
- 1 Early predictability
- 2 Behavioral anomalies
- 3 If systematic indifference, judge identity might predict appeal
- 4 Inattentiveness to appellate reversals
- 5 Implicit risk rankings of asylees closer to random
- 6 Is indifference greater for some refugees (e.g., from Global South)?
- 7 Can we use judicial analytics to increase recognition & dignity?

# Using ML to Diagnose Judicial Inattention

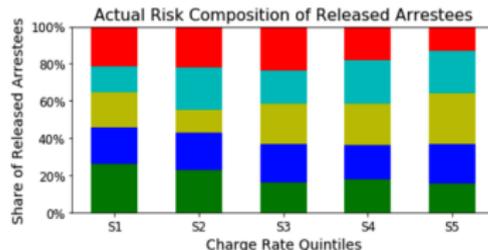
- 1 Early predictability
- 2 Behavioral anomalies
- 3 If systematic indifference, judge identity might predict appeal
- 4 Inattentiveness to appellate reversals
- 5 Implicit risk rankings of asylees closer to random
- 6 Is indifference greater for some refugees (e.g., from Global South)?
- 7 Can we use judicial analytics to increase recognition & dignity?



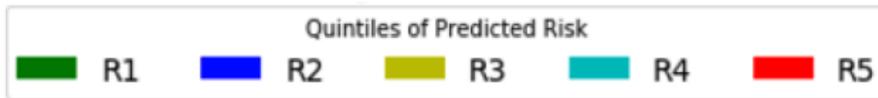
## Robot Prosecutors



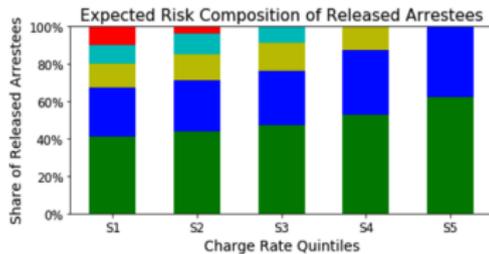
## Human Prosecutors



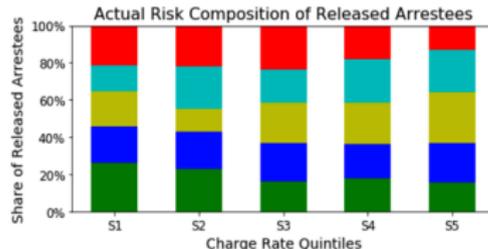
- If defendants released based only on risk score, the harshest prosecutors would only be releasing low-risk defendants.
- Distribution of risk scores for released defendants is similar for most lenient and least lenient prosecutors.
- Are the lenient asylum judges, only denying the 'riskiest' applicants
  - ▶ i.e., seeing the lowest reversal rates (of their asylum denials)?



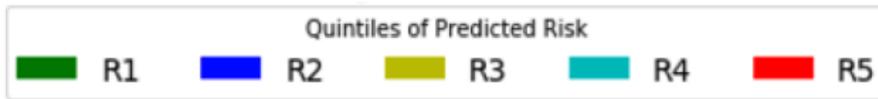
## Robot Prosecutors



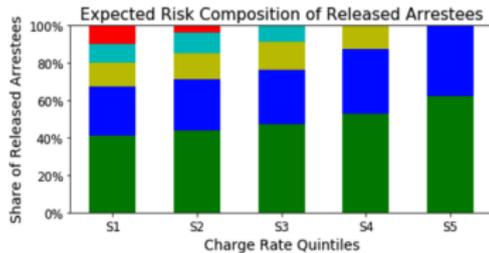
## Human Prosecutors



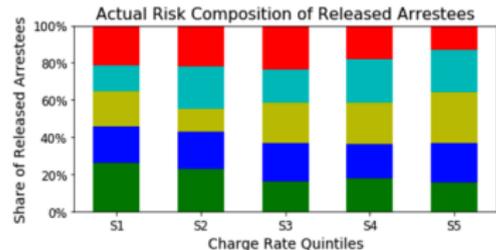
- If defendants released based only on risk score, the harshest prosecutors would only be releasing low-risk defendants.
- Distribution of risk scores for released defendants is similar for most lenient and least lenient prosecutors.
- Are the lenient asylum judges, only denying the 'riskiest' applicants
  - ▶ i.e., seeing the lowest reversal rates (of their asylum denials)?



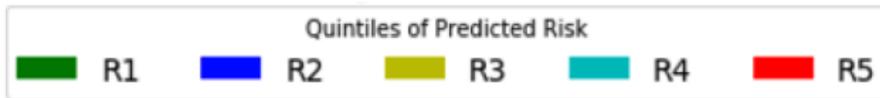
## Robot Prosecutors



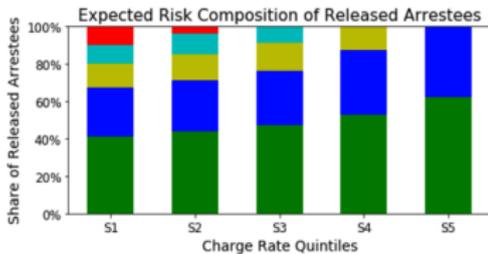
## Human Prosecutors



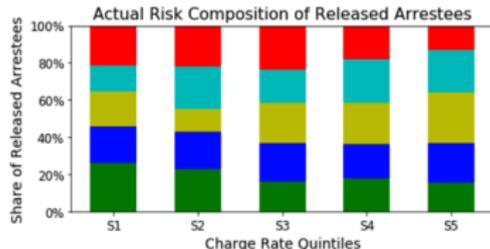
- If defendants released based only on risk score, the harshest prosecutors would only be releasing low-risk defendants.
- Distribution of risk scores for released defendants is similar for most lenient and least lenient prosecutors.
- Are the lenient asylum judges, only denying the 'riskiest' applicants
  - ▶ i.e., seeing the lowest reversal rates (of their asylum denials)?



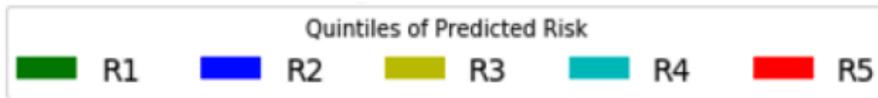
## Robot Prosecutors



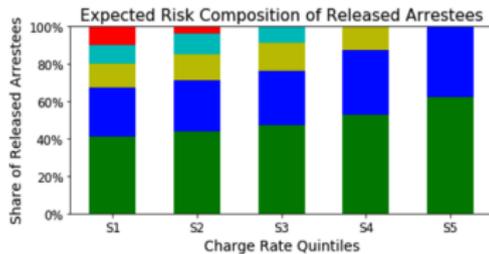
## Human Prosecutors



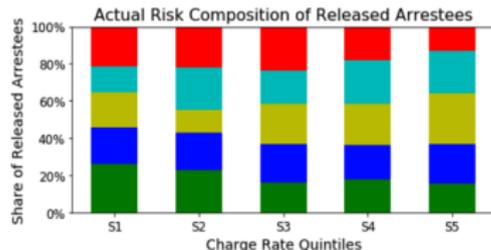
- If defendants released based only on risk score, the harshest prosecutors would only be releasing low-risk defendants.
- Distribution of risk scores for released defendants is similar for most lenient and least lenient prosecutors.
- Are the lenient asylum judges, only denying the 'riskiest' applicants
  - ▶ i.e., seeing the lowest reversal rates (of their asylum denials)?



## Robot Prosecutors



## Human Prosecutors



- If defendants released based only on risk score, the harshest prosecutors would only be releasing low-risk defendants.
- Distribution of risk scores for released defendants is similar for most lenient and least lenient prosecutors.
- Are the lenient asylum judges, only denying the 'riskiest' applicants
  - ▶ i.e., seeing the lowest reversal rates (of their asylum denials)?

Prediction App (Beta): <https://floating-lake-11821.herokuapp.com/>

Schedule Type  
Detained Master Reset

Judge  
John Milo Bryant

Hearing City  
ADELANTO

Asylum type  
Affirmative

Hearing Language  
ABRON

Attorney present?  
Yes

Case Type  
ASYLUM ONLY CASE

Hearing Location  
DHS-LITIGATION UNIT/OAKDALE (ADC)

Nationality  
ANTIGUA AND BARBUDA

Adjudication Medium  
N

Base City  
ADELANTO

[View Prediction](#)

- Program evaluate its effects on trust and perceived indifference of lawmakers
- and applications, decisions, reversals, speed, disparities, etc.

Prediction App (Beta): <https://floating-lake-11821.herokuapp.com/>

Schedule Type  
Detained Master Reset

Judge  
John Milo Bryant

Hearing City  
ADELANTO

Asylum type  
Affirmative

Hearing Language  
ABRON

Attorney present?  
Yes

Case Type  
ASYLUM ONLY CASE

Hearing Location  
DHS-LITIGATION UNIT/OAKDALE (ADC)

Nationality  
ANTIGUA AND BARBUDA

Adjudication Medium  
N

Base City  
ADELANTO

[View Prediction](#)

- **Program evaluate** its effects on trust and perceived indifference of lawmakers
- and applications, decisions, reversals, speed, disparities, etc.

Prediction App (Beta): <https://floating-lake-11821.herokuapp.com/>

Schedule Type  
Detained Master Reset

Judge  
John Milo Bryant

Hearing City  
ADELANTO

Asylum type  
Affirmative

Hearing Language  
ABRON

Attorney present?  
Yes

Case Type  
ASYLUM ONLY CASE

Hearing Location  
DHS-LITIGATION UNIT/OAKDALE (ADC)

Nationality  
ANTIGUA AND BARBUDA

Adjudication Medium  
N

Base City  
ADELANTO

[View Prediction](#)

- **Program evaluate** its effects on trust and perceived indifference of lawmakers
- and applications, decisions, reversals, speed, disparities, etc.

# Quantum App (pilot) to test role of self-image motives

Another possibility for an algorithm is to choose the **three worst metrics** as oppose to just one.



- **Program evaluate** its effects on trust and perceived indifference of lawmakers
- and applications, decisions, reversals, speed, disparities, etc.