

AI in Consequential Decisions: The Need for Transparency

Sonia Gipson Rankin, University of New Mexico School of Law

AI and Consequential Decisions

AI is being used in both the public and private sector to make decisions that have long-term effects on people's lives:

Employment (automated hiring)

Health care, education, social services, fraud detection

Housing: credit, lending, tenant screening, public housing waiting lists

Criminal justice: pretrial, sentencing, parole, predictive policing

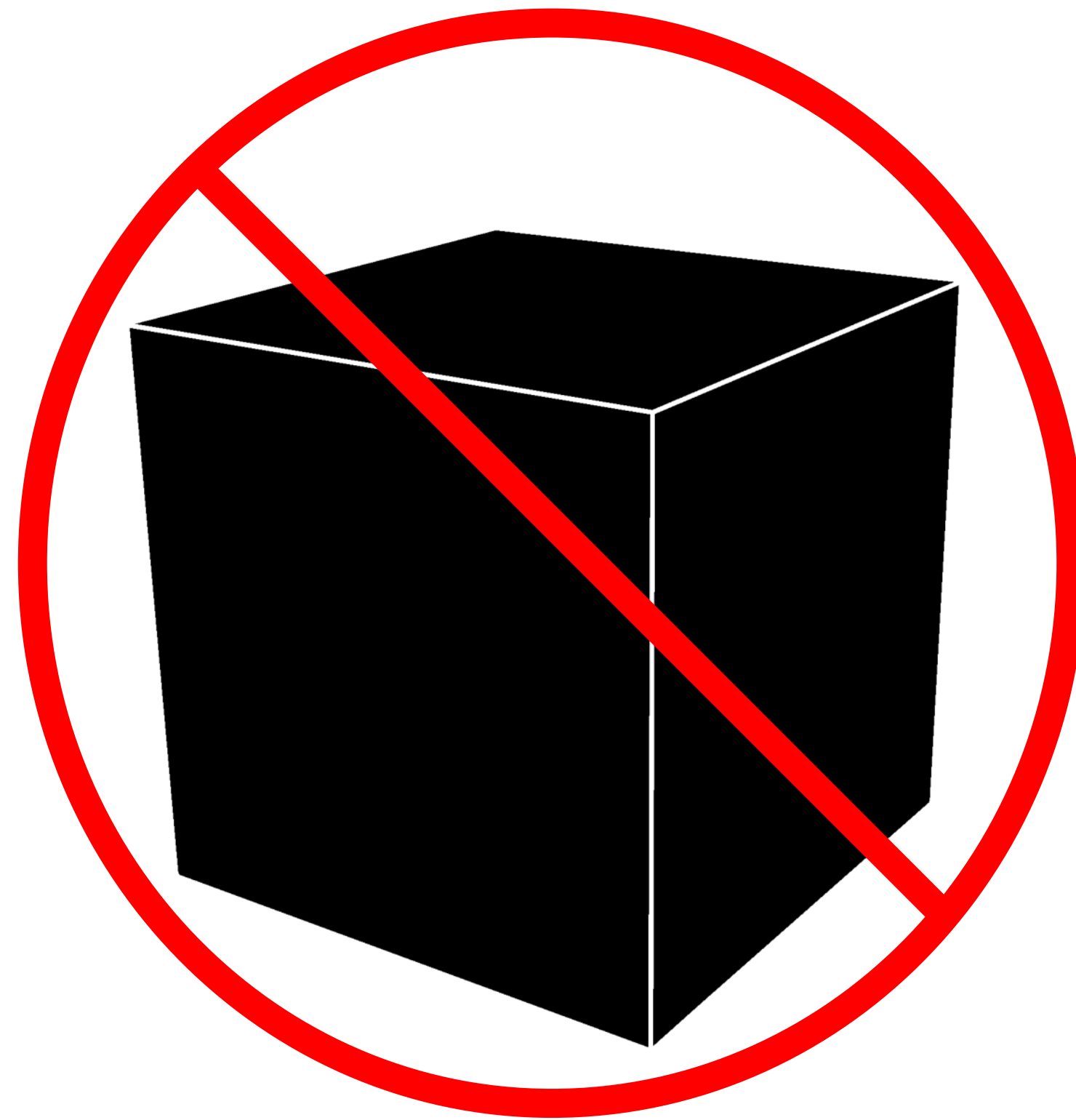
Pros: evidence-based, objective, accurate, avoids stereotypes

Cons: based on historical data, treats people as statistics, black boxes

What do citizens and governments need to know about these systems?

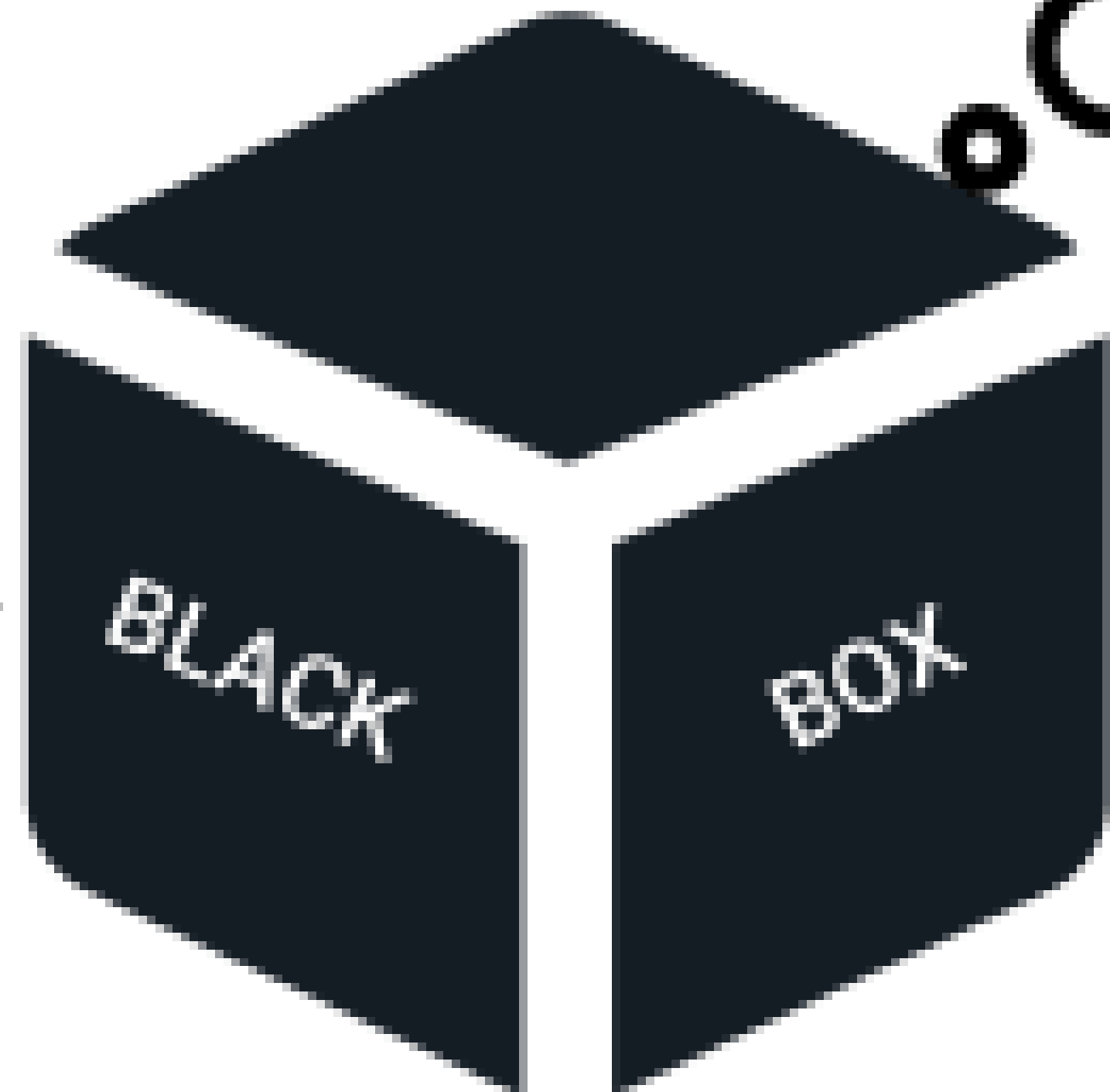
Transparency vs. Black Boxes

- What data does an AI use about a defendant or applicant?
- Where does this data come from?
- What does the AI do with this data to make a decision, a score, or a recommendation?



- Do the people affected by an AI, and the decision makers advised by it, understand the logic behind its decisions?
- Do they know what its limitations are, and what kinds of errors it can make?
- Can we independently assess AIs for accuracy and fairness, or do we just have to take the vendor's word for it?

INPUT



I know I did it. I just
don't know how I
did it :{



OUTPUT

Transparency, Due Process, and Equal Protection

Procedural Due Process

U.S. CONST. amend. XIV, § 1

N.M. CONST. art. II, § 14

Substantive Due Process

U.S. CONST. amend. IX, § 1

U.S. CONST. amend. XIV, § 1

N.M. CONST. art. II, § 18

Equal Protection

U.S. CONST. amend. XIV, § 1

N.M. CONST. art. II, § 18

Example #1: Pretrial Supervision

- **Public Safety Assessment:** Simple point system, publicly known weights
- **Based on criminal record:** Past convictions, past failures to appear
- Uses age, but not *race*, gender, employment, education, or environment

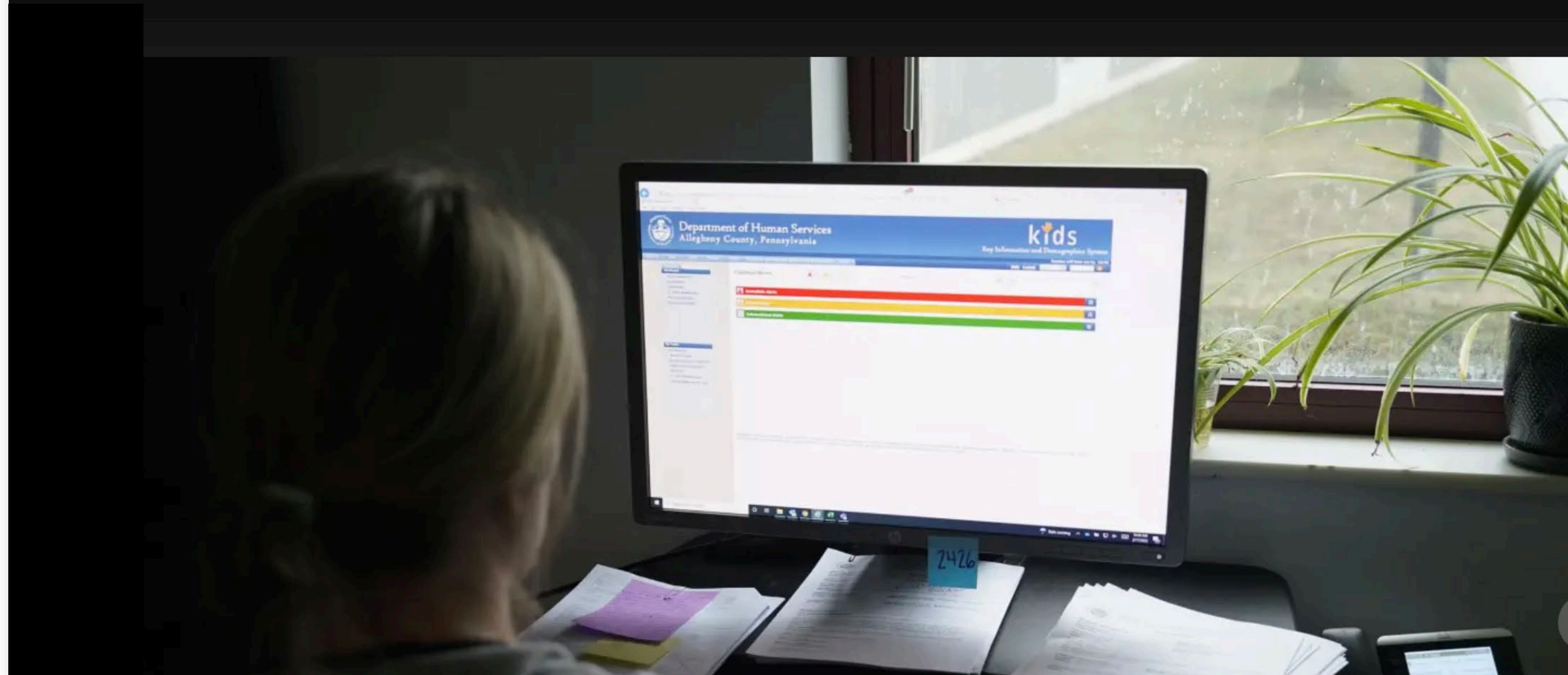
PUBLIC SAFETY ASSESSMENT RISK FACTORS

RISK FACTOR	WEIGHTS
FAILURE TO APPEAR maximum total weight = 7 points	
Pending charge at the time of the offense	No = 0 Yes = 1
Prior conviction	No = 0 Yes = 1
Prior failure to appear pretrial in past 2 years	0 = 0 1 = 2 2 or more = 4
Prior failure to appear pretrial older than 2 years	No = 0 Yes = 1
NEW CRIMINAL ACTIVITY maximum total weight = 13 points	
Age at current arrest	23 or older = 0 22 or younger = 2
Pending charge at the time of the offense	No = 0 Yes = 3
Prior misdemeanor conviction	No = 0 Yes = 1
Prior felony conviction	No = 0 Yes = 1
Prior violent conviction	0 = 0 1 or 2 = 1 3 or more = 2
Prior failure to appear pretrial in past 2 years	0 = 0 1 = 1 2 or more = 2
Prior sentence to incarceration	No = 0 Yes = 2
NEW VIOLENT CRIMINAL ACTIVITY maximum total weight = 7 points	
Current violent offense	No = 0 Yes = 2
Current violent offense & 20 years old or younger	No = 0 Yes = 1
Pending charge at the time of the offense	No = 0 Yes = 1
Prior conviction	No = 0 Yes = 1
Prior violent conviction	0 = 0 1 or 2 = 1 3 or more = 2

Source: Laura and John Arnold Foundation

Example #2: Child Welfare and Protective Services

Child welfare algorithm faces Justice Department scrutiny



Allegheny County, PA
(Pittsburgh)

- Uses prior allegations, publicly funded mental health and drug/alcohol services, jail bookings
- Predicts removal from home within 2 years, re-referral after initially being screened out, or injury

Oregon Department of Human Services to End Its Use of Child Abuse Risk Algorithm

Example #3: Fraud Detection

Government's Use of Algorithm Serves Up False Fraud Charges

Using a flawed automated system, Michigan falsely charged thousands with unemployment fraud and took millions from them.

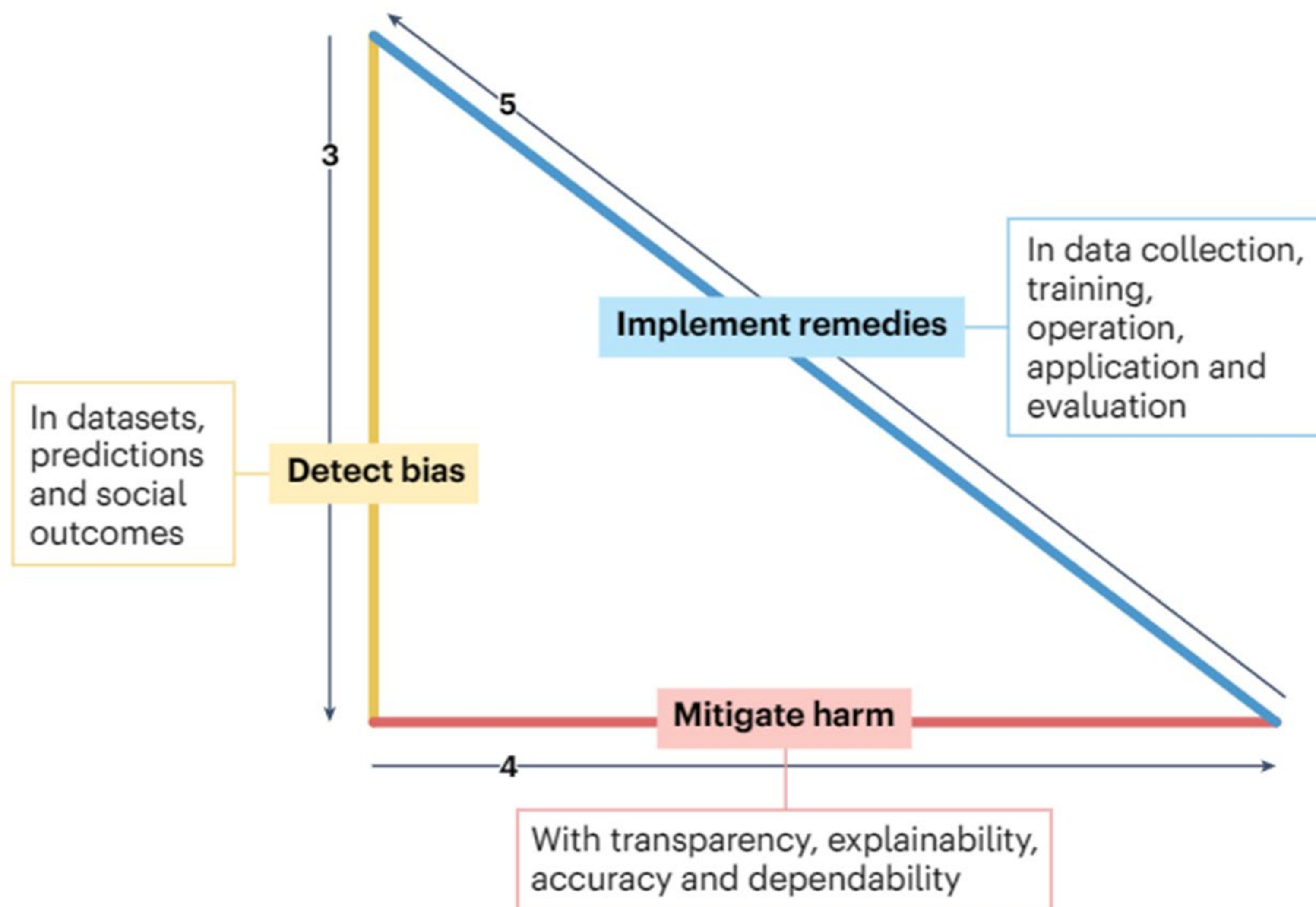
- Agency charged 40,000 people, billing them at five times the original benefits at a rate of 400 percent plus interest
- 93% of the charges were erroneous
- Agency failed to repay millions of dollars for years

AI can help inform consequential decisions if...

- People affected by them understand what data about them is used and what the AI does with this data
- Decision makers advised by them understand what they mean and what mistakes they can make
- Policymakers understand their strengths and weaknesses
- They are regularly and independently assessed for accuracy and fairness, rather than relying on vendor's claims

All this requires transparency!

Transparency is step one.



Types of Transparency

“Where constitutional rights are involved, transparency is paramount.”

— Computing Community Consortium

Simple notice: Alert consumers or applicants that an AI is being used

Applicant Challenges: Allow applicants to see their data and correct it (e.g. FCRA)

Self-assessment: Require AI developers to assess their own product for bias, and perform due diligence to avoid it (like an impact statement)

Local studies: Require AI deployers to periodically test the AI for accuracy and bias on local data to make sure it works well for local populations

Independent assessments: Independent third parties (e.g. ISR at UNM)

Full transparency: Public disclosure of design and methods, sources of data, and how the AI uses that data to produce its output